RESEARCH CENTER

FIELD
Perception, Cognition and Interaction

Activity Report 2019

Section New Results

Edition: 2020-03-21

CEDAR Project-Team

7. New Results

7.1. Quotient summaries of RDF graphs

We have continued and finalized our work on the question of efficiently computing informative summaries of large, heterogeneous RDF graphs. Such summaries simplify the users' efforts to understand and grasp the content of an RDF graph with which they are not familiar. For instance, Figure 1 shows the summary constructed fully automatically out of a benchmark graph of a bit more than 100 million triples.

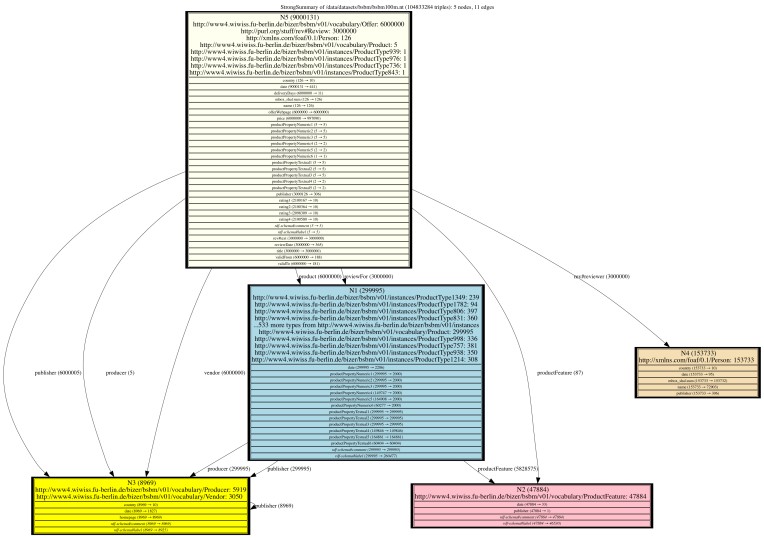


Figure 1. RDFQuotient summary of a 100 million triples graph.

We have presented, together with co-authors, a tutorial on the problem of summarizing RDF graphs, at the EDBT 2019 conference [21].

We have demonstrated new algorithms for efficiently building RDF quotient summaries out of large RDF graphs, in an incremental fashion, in [19].

Last but not least, a VLDB Journal submitted article systematizing most of our contributions in this area has been accepted (pending a minor, strictly cosmetic revision which will be sent out in January 2020).

7.2. Efficient query answering over semantic graphs

Query answering in RDF knowledge bases has traditionally been performed either through graph saturation, that is, adding all implicit triples to the graph, or through query reformulation, i.e. modifying the query to look for the explicit triples entailing precisely what the original query asks for. The most expressive fragment of RDF for which reformulation-based quey answering exists is the so-called database fragment of RDF (Goasdoué et al., EDBT 2013), in which implicit triples are restricted to those entailed using an RDFS ontology. Within this fragment, query answering was so far limited to the interrogation of data triples (non-RDFS ones); however, a powerful feature specific to RDF is the ability to query data and schema

triples together. In [12], we address the general query answering problem by reducing it, through a pre-query reformulation step, to that solved by the query reformulation technique mentioned above (EDBR 2013). Our experiments also demonstrate the very modest cost (performance overhead) of this more powerful (more expressive) reformulation algorithm.

7.3. Scalable storage for polystores

Big data applications routinely involve diverse datasets: relations flat or nested, complex-structure graphs, documents, poorly structured logs, or even text data. To handle the data, application designers usually rely on several data stores used side-by-side, each capable of handling one or a few data models (e.g., many relational stores can also handle JSON data), and each very efficient for some, but not all, kinds of processing on the data.

A current limitation is that applications are written taking into account which part of the data is stored in which store and how. This fails to take advantage of (i) possible redundancy, when the same data may be accessible (with different performance) from distinct data stores; (ii) partial query results (in the style of materialized views) which may be available in the stores. If data migrates to another store, to take advantage of its performance for a specific task, applications must be re-written; this is tedious and error-prone.

In [11], we present ESTOCADA, a novel approach connecting applications to the potentially heterogeneous systems where their input data resides. ESTOCADA can be used in a polystore setting to transparently enable each query to benefit from the best combination of stored data and available processing capabilities. ESTOCADA leverages recent advances in the area of view-based query rewriting under constraints, which we use to describe the various data models and stored data. Our experiments illustrate the significant performance gains achieved by ESTOCADA.

7.4. Novel fact-checking architectures and algorithms

A frequent journalistic fact-checking scenario is concerned with the **analysis of statements** made by individuals, whether in public or in private contexts, and the propagation of information and hearsay ("who said/knew what when"), mostly in the public sphere (e.g., in discourses, statements to the media, or on public social networks such as Twitter), but also in private contexts (these become accessible to journalists through their sources). Inspired by our collaboration with fact-checking journalists from Le Monde, France's leading newspaper, we have described in [17] a Linked Data (RDF) model, endowed with formal foundations and semantics, for describing *facts, statements*, and *beliefs*. Our model combines temporal and belief dimensions to trace propagation of knowledge between agents along time, and can answer a large variety of interesting questions through RDF query evaluation. A preliminary feasibility study of our model incarnated in a corpus of tweets demonstrates its practical interest.

Based on the above model, we implemented and demonstrated BELINK [13], a prototype capable of storing such interconnected corpora, and answer powerful queries over them relying on SPARQL 1.1. The demo showcased the exploration of a rich real-data corpus built from Twitter and mainstream media, and interconnected through extraction of statements with their sources, time, and topics.

Statistic (numerical) data, e.g., on unemployment rates or immigrant populations, are hot fact-checking topics. In prior work, we have transformed a corpus of high-quality statistics from INSEE, the French national statistics institute, into an RDF dataset (Cao et al., Semantic Big Data Workshop, 2017, https://hal.inria.fr/hal-01583975), and shown how to locate inside the information most relevant to (thus, most likely to be useful to fact-check) a given keyword query (Cao et al., Web and Databases Workshop, 2018, https://hal.inria.fr/hal-01745768). Following on the above work, in [16], we present a novel approach to extract from text documents, e.g., online media articles, mentions of statistic entities from a reference source. A claim states that an entity has certain value, at a certain time. This completes a fact-checking pipeline from text, to the reference data closest to the claim. Using it, fact-checking journalists only have to interpret the difference between the claimed and the reference value. We evaluated our method on the INSEE reference dataset and show that it is efficient and effective. Further, this algorithm was adapted also to the (more challenging) context of content

published on Twitter. This has lead to a semi-automatic interface for detecting statistic claims made in tweets and starting a semi-automatic fact-check of those claims, based on INSEE data. Figure 2 depicts the interface of this Twitter fact-checking system, which was shared with our Le Monde journalist partners.

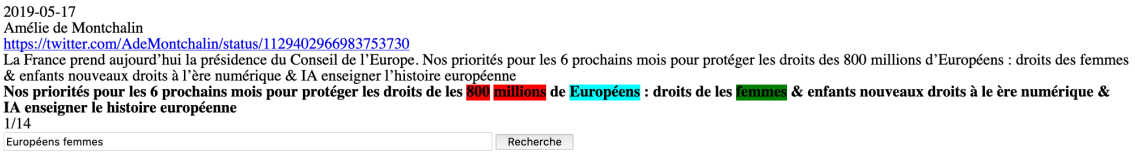


Figure 2. Screen capture of our Twitter fact-checking module.

7.5. Semantic graph exploration through interesting aggregates

RDF graphs can be large and complex; finding out interesting information within them is challenging. One easy method for users to discover such graphs is to be shown *interesting aggregates* (under the form of two-dimensional graphs, i.e., bar charts), where interestingness is evaluated through statistics criteria. While well understood for relational data, such exploration raises multiple challenges for RDF: facts, dimensions and measures have to be *identified* (as opposed to known beforehand); as there are more candidate aggregates, assessing their interestingness can be very costly; finally, *ontologies* bring novel specific challenges through the presence of *implicit* data, but also novel opportunities, enabling *ontology-driven exploration* from an aggregate initially proposed by the system.

The system DAGGER we had previously proposed (2017) pioneered this approach, however its is quite inefficient, in particular due to the need to evaluate numerous, expensive aggregation queries.

In 2019, we have built upon DAGGER to develop more efficient and more expressive versions thereof. Thus:

- In [22], we describe DAGGER$^+$, which builds upon DAGGER and leverages *sampling* to speed up the evaluation of potentially interesting agregates. We show that DAGGER$^+$ achieves very significant execution time reductions, while reaching results very close to those of the original, less efficient system.
- Going beyond the expressive power of (candidate aggregates enumerated by) DAGGER, we have developed and demonstrated [15] SPADE, a *generic, extensible framework*, which we instantiated with: (i) novel methods for enumerating candidate measures and dimensions in the vast space of possibilities provided by an RDF graph; (ii) a set of aggregate interestingness functions; (iii) ontology-based interactive exploration, and (iv) efficient early-stop techniques for estimating the interestingness of an aggregate query. A multi-dimensional aggregate automatically identified by SPADE appears in Figure 3 .

7.6. A Next-Generation Unified Data Analytics Optimizer

Big data analytics systems today still lack the ability to take user performance goals and budgetary constraints, collectively referred to as "objectives", and automatically configure an analytic job to achieve the objectives.

In [10], we present a unified data analytics optimizer that can automatically determine the parameters of the runtime system, collectively called a job configuration, for general dataflow programs based on user objectives. UDAO embodies key techniques including in-situ modeling, which learns a model for each user objective in the same computing environment as the job is run, and multi-objective optimization, which computes a Pareto optimal set of job configurations to reveal tradeoffs between different objectives.

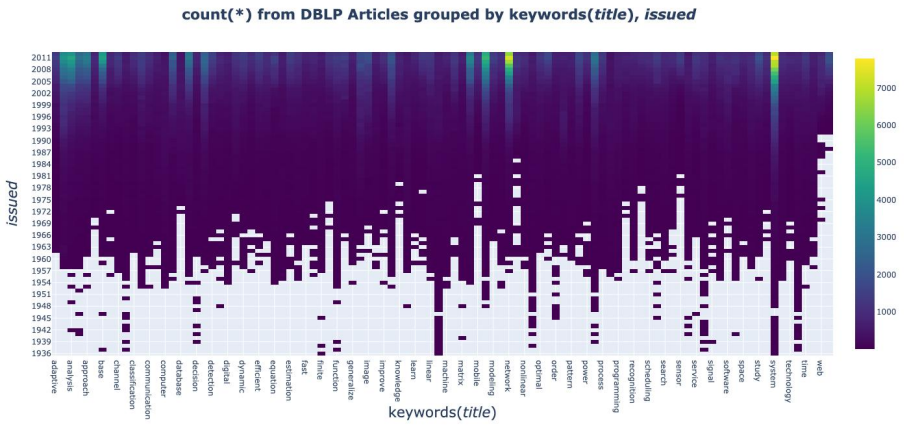


Figure 3. Interesting multi-dimensional aggregate automatically identified by DAGGER.

Using benchmarks developed based on industry needs, our demonstration will allow the user to explore (1) learned models to gain insights into how various parameters affect user objectives; (2) Pareto frontiers to understand interesting tradeoffs between different objectives and how a configuration recommended by the optimizer explores these tradeoffs; (3) end- to-end benefits that UDAO can provide over default configurations or those manually tuned by engineers.

We demonstrated this work at the VLDB 2019 conference.

7.7. A factorized version space algorithm for interactive database exploration

One challenge in building an interactive database exploration system is that existing active learning (AL) techniques experience slow convergence when learning the user interest on large datasets. To address this slow convergence problem, we augmented version space-based AL algorithms, which have strong theoretical results on convergence but are very costly to run, with additional insights obtained in the user labeling process. These insights lead to a novel algorithm that factorizes the version space to perform active learning in a set of subspaces, with provable results on optimality, as well as optimizations for better performance. Evaluation results using real world datasets show that our algorithm significantly outperforms state-of-the-art version space algorithms, as well as our previous data exploration algorithm DSM (Huang et al., PVLDB 2018), for large database exploration.

The above work was accepted as a conference paper at ICDM 2019 [14]. In addition, we have presented a demonstration of our software at NeurIPS 2019 [26], where people could interact with our system over two real-world datasets, and also observe how our system compares against traditional AL algorithms.

GRAPHIK Project-Team

7. New Results

7.1. Ontology-Mediated Query Answering

Participants: Jean-François Baget, Meghyn Bienvenu, Efstathios Delivorias, Michel Leclère, Marie-Laure Mugnier, Olivier Rodriguez, Federico Ulliana.

Ontolology-mediated query answering (OMQA) is the issue of querying data while taking into account inferences enabled by ontological knowledge. From an abstract viewpoint, this gives rise to *knowledge bases*, composed of an ontology and a factbase (in database terms: a database instance under incomplete data assumption). Answers to queries are logically entailed from the knowledge base.

This year, we obtained two kinds of results: *theoretical results* on fundamental issues raised by OMQA, and *practical algorithms* for OMQA on key-value stores and RDF integration systems.

7.1.1. Fundamental issues on OMQA with existential rules

Existential rules (a.k.a. datalog+, as this framework generalizes the deductive database language datalog) have emerged as a new ontological language in the OMQA context. Techniques for query answering under existential rules mostly rely on the two classical ways of processing rules, namely forward chaining and backward chaining. In forward chaining, known as the *chase* in database theory, the rules are applied to enrich the factbase and query answering can then be solved by evaluating the query against the *saturated* factbase (as in a classical database system, i.e., with forgetting the ontological knowledge). The backward chaining process is divided into two steps: first, the query is *rewritten* using the rules into a first-order query (typically a union of conjunctive queries, but it can be a more compact form) or into a datalog query; then the rewritten query is evaluated against the factbase (again, as in a classical database system). Depending on the considered class of existential rules, the chase and/or query rewriting may terminate or not.

7.1.1.1. Decidability of chase termination for linear existential rules.

Several chase variants have long been studied in database theory. These chase variants yield logically equivalent results, but differ in their ability to detect redundancies possibly caused by the introduction of unknown individuals (nulls, blank nodes). Given a chase variant, the chase termination problem takes as input a set of existential rules and asks if this set of rules ensures the termination of the chase for any factbase. It is well-known that this problem is undecidable for all known chase variants. Hence, a crucial issue is whether chase termination becomes decidable for some known subclasses of existential rules. We considered linear existential rules, a simple yet important subclass of existential rules that generalizes database inclusion dependencies. We showed the decidability of the chase termination problem on linear rules for three main chase variants, namely skolem (a.k.a. semi-oblivious), restricted (a.k.a. standard) and core chase. The restricted chase is the most used in practice, however its study is notoriously tricky because the order in which rule applications are performed matters. Indeed, for the same factbase, some restricted chase sequences may terminate, while others may not. To obtain our results, we introduced a novel approach based on so-called derivation trees and a single notion of forbidden pattern. The simplicity of these structures make them subject to implementation. Besides the theoretical interest of a unified approach and new proofs, we provided the first positive decidability results (and complexity upper bounds) concerning the termination of the restricted chase, proving that chase termination on linear existential rules is decidable for both versions of the problem: Does every chase sequence terminate? Does some chase sequence terminate?

- *ICDT 2019 [29]. In collaboration with Michael Thomazo (Inria VALDA).*

7.1.1.2. Boundedness: Enforcing both chase termination and first-order rewritability.

We carried out the first studies on the boundedness problem for existential rules. This problem asks whether a given set of existential rules is bounded, i.e., whether there is a predefined bound on the "depth" of the chase independently from any factbase (for breadth-first chase versions, the depth corresponds to the number of breadth-first steps). It has been deeply studied in the context of datalog, where it is key to query optimization, although boundedness is undecidable in general. For datalog rules, boundedness is equivalent to a desirable property, namely first-order rewritability: a set of rules is called first-order rewritable if any conjunctive query can be rewritten into a union of conjunctive queries, whose evaluation on any factbase yields the expected answers (i.e., the relevant part of the ontology can be compiled into the rewritten query, which allows to reduce query answering to a simple query evaluation task). This equivalence does not hold for existential rules. Moreover, the notion of boundedness has to be parametrized by the chase variant, as they all behave differently with respect to termination. Beside potential practical use, the notion of boundedness is closely related to an interesting theoretical question on existential rules: what are the relationships between chase termination and first-order query rewritability? With respect to this question, we obtained the following salient result: for the oblivious and skolem (semi-oblivious) chase variants, a set of existential rules is bounded if and only if it ensures both chase termination for any factbase and first-order rewritability for any conjunctive query.

- *IJCAI 2019 [22]. In collaboration with Pierre Bourhis (Inria SPIRALS) and Sophie Tison (Inria LINKS).*

7.1.2. Practical Algorithms for OMQA on key-value stores and RDF integration systems

7.1.2.1. Ontology-mediated query answering on top of key-value stores.

Ontology-mediated query answering was mainly investigated so far based on the assumption that data conforms to relational structures (we include here RDF) and that the paradigm can be deployed on top of relational databases with conjunctive queries at the core (e.g., in SQL or SPARQL). However, this is not the prominent way on which data is today stored and exchanged, especially in the Web. Whether OMQA can be developed for non-relational structures, like those shared by increasingly popular NOSQL languages sustaining Big-Data analytics, has just begun to be investigated. Since 2016, we have been studying OMQA for key-values stores, which are systems providing fast and scalable access to JSON records. We proposed a rule language to express domain knowledge, with rules being directly applicable to key-value stores, without any translation of JSON into another data model (results published at AAAI 2016 and IJCAI 2017). In 2018-2019, we implemented a prototype for MongoDB, with a restricted part of this rule language (featuring key inclusions and mandatory keys) and tree-pattern queries, and devised optimization techniques based on parallellizing query rewriting and query answering. This work is pursued within a starting PhD thesis (Olivier Rodriguez).

- *Rule-ML 2019 [31]. In collaboration with Reza Akbarinia (Inria ZENITH).*

7.1.2.2. Ontology-mediated query answering in RDF integration systems

Within the iCODA project devoted to data journalism and the co-supervision of Maxime Buron's PhD thesis, we are considering the so-called Ontology-Based Data Access framework, which is composed of three components: the data level, the ontological level and mappings that relate data to facts described in the vocabulary of the ontology. Our framework more precisely considers heterogeneous data sources integrated through mappings into a (possibly virtual) RDF graph, provided with an RDFS ontology and RDFS entailment rules. The innovative aspects with respect to the state of the art are (i) SPARQL queries that extend classical conjunctive queries by the ability of querying data and ontological triples together, and (ii) Global-Local-As-View (GLAV) mappings, which can be seen as source-to-target existential rules. GLAV mappings enable the creation of unknown entities (blank nodes), which increases the amount of information accessible through the integration system. In particular, they allow one to palliate missing data values, by stating the existence of data whose values are not known in the sources. We devised, implemented and experimentally compared several query answering techniques in this setting.

- *ESWC 2019 [23], technical report [36] basis of a paper accepted to EDBT 2020. In collaboration with Maxime Buron and Ioana Manolescu (Inria CEDAR), and François Goasdoué (IRISA).*

7.2. Reasoning with conflicts and decision support

Participants: Pierre Bisquert, Patrice Buche, Michel Chein, Madalina Croitoru, Jérôme Fortin, Alain Gutierrez, Abdelraouf Hecham, Martin Jedwabny, Michel Leclère, Rallou Thomopoulos, Bruno Yun.

The work carried out during this year can be structured into two main research directions: *structured logic-based argumentation* and *collective decision making*.

7.2.1. Structured argumentation

To solve real-world problems we sometimes need to consider features that cannot be expressed purely (or naturally) in classical logic. Indeed, real world information is often "imperfect": it can be partially contradictory, vague or uncertain, etc. During the evaluation period, we mostly considered reasoning in presence of conflicts. To handle this issue, as a reasoning method robust to contradiction, we have used structured argumentation, where arguments have an internal logical structure representing an inference step (i.e. some premises inducing a conclusion). In this context, arguments and their interaction are typically generated from an inconsistent knowledge base. Such arguments are in contrast to those employed in abstract argumentation where they are considered a black box (usually provided as input to a problem and not computed).

More precisely, this year, we mainly worked on two issues: the first one concerns the question of scrutinizing a structured argument, i.e. checking both the validity ("is the conclusion induced by the premisses?") and its soundness ("is the argument valid and are its premisses true?"). This is interesting in the context of collective decision making, where participants utter arguments that can be assessed. The second one relates to the computational complexity of generating arguments from a knowledge base. Indeed, it can potentially produce a huge number of arguments, which impedes the usability of argumentation for big knowledge bases.

7.2.1.1. Formalizing argument schemes and fallacies

More precisely, we have presented a logical framework allowing us to express assessment of facts and arguments together with a proof system to answer these questions. Our motivation was to clarify the notion of validity in the context of logic-based arguments along different aspects (such as the formulas used and the inference scheme). Originality lies in the possibility for the user to design their own argument schemes, i.e. specific inference patterns (e.g. expert argument, analogy argument). We showed that classical inference obtains when arguments are based on classical schemes (e.g. Hilbert axioms). We went beyond classical logic by distinguishing "proven" formulas from "uncontroversial" ones (whose negation is not proven) and provided a definition of a fallacious argument in this context.

- *LPNMR 2019 [20]. In collaboration with Florence Dupin de Saint-Cyr and Philippe Besnard (IRIT).*

7.2.1.2. Optimising argumentation frameworks

Another problem addressed was the large number of logical arguments that can be potentially constructed from a knowledge base. To address this problem we have proposed a compact representation of the structured argumentation system under the form of hypergraphs and implemented it in the NAKED prototype. The tool allows to import a knowledge base (expressed in the existential rule framework), generate, visualise and export the corresponding argumentation hypergraph. These functions, paired with the aim of improving the extension computation efficiency, make this software an interesting tool for non-computer science experts, such as people working in the agronomy domain.

- *AAMAS 2019 [33]. In collaboration with Srdjan Vesic (CRIL).*

7.2.2. Collective decision making

In this setting we have focused towards the deliberation and voting techniques. We have investigated how deliberation can help generate or impact the structure of preferences underlying the voting process. We have implemented the PAPOW prototype [27] that allows for filtering of voters depending on their individual characteristics.

7.2.2.1. Argumentation as a tool to generate new preferences

We have investigated how argumentation can solve the Condorcet paradox by using the notion of extension (maxi-consistent sets of arguments) in order to compute new preferences. Our research hypothesis is that a decision made by a group of participants understanding the qualitative rationale (expressed by arguments) behind each other's preferences has better chances to be accepted and used in practice. Accordingly, we proposed a novel qualitative procedure which combines argumentation with computational social choice for modeling the collective decision-making problem. We showed that this qualitative approach produces structured preferences that can overcome major deficiencies that were exhibited in the social choice literature and affect most of the major voting rules. More precisely, we have dealt with the Condorcet Paradox and the properties of monotonicity and homogeneity, which are unsatisfiable by many voting rules.

- *PRAI 2019 [14]. In collaboration with Christos Kaklamanis and Nikos Karanikolas (CTI, Greece).*

7.2.2.2. Argumentation as a tool to modify individual preferences

The previous approach implies that voters are replaced by the extensions which, while it allows to circumvent the Condorcet Paradox, might prove difficult to implement as it disregards the notion of (voters') majority. Hence, we proposed a decision-making procedure based on argumentation and preference aggregation which permits us to explore the effect of reasoning and deliberation along with voting for the decision process. We represented the deliberation phase by defining a new voting argumentation framework, that uses vote and generic arguments, and its acceptability semantics based on the notion of pairwise comparisons between alternatives. We proved for these semantics some theoretical results regarding well-known properties from argumentation and social choice theory.

Moreover, we also studied the notion of unshared features (i.e., alternatives' criteria that constitute justifications of preferences for some agents but not for others) and showed under which conditions it is possible to reach a Condorcet consensus. We provided a deliberation protocol that ensures that, after its completion, the number of unshared features of the decision problem can only be reduced, which would tend to show that deliberation allows to lower the risk of Condorcet Paradox.

- *ICAART 2019 [28]. In collaboration with Christos Kaklamanis and Nikos Karanikolas (CTI, Greece). PRIMA 2019 [21].*

7.2.3. Discovering and qualifying authority links

We finalized this year the description of the engine SudoQual, devoted to the evaluation of link quality in document bases, developed in collaboration with ABES, the French National Agency for Academic Libraries (http://www.abes.fr), in the context of ANR Qualinca research project (2012-2016) (https://www.lirmm.fr/qualinca/). We presented the methodology and general algorithms used to discover and qualify so-called authority links (which are coreference links between entities mentioned in descriptions of documents and entities described in referential bases). Moreover, ABES has put in production this year a professional tool for documentalists, called Paprika (https://paprika.idref.fr/), whose kernel is the SudoQual engine.

- *KCAP 2019 [25].*

7.3. Miscellaneous: Automated design of biological devices

Participants: Michel Leclère, Guillaume Perution Kihli, Federico Ulliana.

We mention here results obtained in a collaboration with a team of biologists from the Center for Structural Biochemistry (CBS, Montpellier) on the logical computing capabilities of living organisms. More precisely, this joint work focuses on the development of a framework dedicated to the design of so-called Recombinase-based devices, whose behavior is specified as Boolean functions. We looked at the case of single-cell devices, whose expressivity limits, that is, the Boolean functions they can implement without distributing the Boolean function in several parts, are still unknown. While it is easy to determine which Boolean function is implemented by a device, the converse problem of automatically designing a device implementing a given Boolean function is a difficult task for which no automatic method exists. To tackle this problem, we experimented in the past years a combinatorial approach consisting in exhaustively generating all devices up to

a given size, then determining the Boolean function they implement. A generating program and a database for these devices were developed. This year, we achieved the first formal study of this problem, which we believe can serve as foundations for the development of new biological design solutions. A set of minimality properties naturally emerged from our study, which led us to define the notion of canonical and representative devices, by which infinitely large classes of design solutions can be finitely expressed. These results strengthen the reliability of the approach and show that our program generates all representative canonical devices. Finally, our results also indicate some interesting expressivity limits for single-cell devices. Indeed, the generation process showed that 8% among all 4-input Boolean functions cannot be implemented. We also formally proved that single-cell devices cannot implement some n-input Boolean functions, for every $n \geq 7$.

- *TPNC 2019 [30]. In collaboration with Jérôme Bonnet and Sarah Guiziou (CBS).*

LACODAM Project-Team

7. New Results

7.1. Introduction

In this section, we organize the bulk of our contributions this year along two of our research axes, namely Pattern Mining and Decision Support. Some other contributions lie within the domain of machine learning.

7.1.1. Pattern Mining

In the domain of pattern mining we can categorize our contributions along the following lines:

- *Efficient Pattern Mining (Sections 7.2-7.4).* In [9], we propose a method to accelerate itemset sampling on FPGAs, whereas [18] proposes SSDPS, an efficient algorithm to mine discriminant patterns in two-class datasets, common in genetic data. Finally [11] presents a succinct data structure that represents concisely a cube of skypatterns.
- *Semantics of Pattern Mining (Sections 7.5-7.6).*[14] discusses the ambiguity of the semantics of pattern mining with absent events (negated statements). Likewise [8] shows formal properties of admissible generalizations in pattern mining and machine learning.

7.1.2. Decision Support

In regards to the axis of decision support, our contributions can be organized in two categories: forecasting & prediction, and modelisation.

- *Forecasting & Prediction (Sections 7.7-7.9).* In [10], we propose solutions to automate the task of capacity planning in the context of a large data network as the one available at Orange. [17] applies machine learning techniques for estrus detection in diary farms. [21] proposes a machine learning architecture in multi-sensor environments for earthquake early warning.
- *Modelling (Section 7.10).* In [5] we present a modeling approach for the nutritional requirements of lactating sows.
- *Data Exploration (Section 7.11).*[6] proposes a formal framework for the exploration of care trajectories in medical databases.

7.1.3. Others

- *Machine Learning (Section 7.12-7.14).*[7], [16] proposes novel methods to optimize the F-measure in ML, and to improve the task of domain adaptation by source selection. [19] proposes the use of GANs to make time series classification more interpretable.

7.2. Accelerating Itemset Sampling using Satisfiability Constraints on FPGA

Finding recurrent patterns within a data stream is important for fields as diverse as cybersecurity or e-commerce. This requires to use pattern mining techniques. However, pattern mining suffers from two issues. The first one, known as “pattern explosion”, comes from the large combinatorial space explored and is the result of too many patterns output to be analyzed. Recent techniques called output space sampling solve this problem by outputting only a sampled set of all the results, with a target size provided by the user. The second issue is that most algorithms are designed to operate on static datasets or low throughput streams. In [9], we propose a contribution to tackle both issues, by designing an FPGA accelerator for pattern mining with output space sampling. We show that our accelerator can outperform a state-of-the-art implementation on a server class CPU using a modest FPGA product.

7.3. Statistically Significant Discriminative Patterns Searching

In [18], we propose a novel algorithm, named SSDPS, to discover patterns in two-class datasets. The SSDPS algorithm owes its efficiency to an original enumeration strategy of the patterns, which allows to exploit some degrees of anti-monotonicity on the measures of discriminance and statistical significance. Experimental results demonstrate that the performance of the SSDPS algorithm is better than others. In addition, the number of generated patterns is much less than the number of the other algorithms. Experiment on real data also shows that SSDPS effciently detects multiple SNPs combinations in genetic data.

7.4. Compressing and Querying Skypattern Cubes

Skypatterns are important since they enable to take into account user preference through Pareto-dominance. Given a set of measures, a skypattern query finds the patterns that are not dominated by others. In practice, different users may be interested in different measures, and issue queries on any subset of measures (a.k.a subspace). This issue was recently addressed by introducing the concept of skypattern cubes. However, such a structure presents high redundancy and is not well adapted for updating operations like adding or removing measures, due to the high costs of subspace computations in retrieving skypatterns. In [11], we propose a new structure called Compressed Skypattern Cube (abbreviated CSKYC), which concisely represents a skypattern cube, and gives an efficient algorithm to compute it. We thoroughly explore its properties and provide an efficient query processing algorithm. Experimental results show that our proposal allows to construct and to query a CSKYC very efficiently.

7.5. Semantics of Negative Sequential Patterns

In the field of pattern mining, a negative sequential pattern expresses behavior by a sequence of present and absent events. In [14], we shed light on the ambiguity of this notation and identify eight possible semantics with the relation of inclusion of a motif in a sequence. These semantics are illustrated and we are studying them formally. We thus propose dominance and equivalence relationships between these semantics, and we highlight new properties of anti-monotony. These results could be used to develop new efficient algorithms for mining frequent negative sequential patterns.

7.6. Admissible Generalizations of Examples as Rules

Rule learning is a data analysis task that consists in extracting rules that generalize examples. This is achieved by a plethora of algorithms. Some generalizations make more sense for the data scientists, called here admissible generalizations. The purpose of our work in [8] is to show formal properties of admissible generalizations. A formalization for generalization of examples is proposed allowing the expression of rule admissibility. Some admissible generalizations are captured by preclosure and capping operators. Also, we are interested in selecting supersets of examples that induce such operators. We then define classes of selection functions. This formalization is more particularly developed for examples with numerical attributes. Classes of such functions are associated with notions of generalization and they are used to comment some results of the CN2 algorithm [22].

7.7. Towards a Framework for Seasonal Time Series Forecasting Using Clustering

Seasonal behaviours are widely encountered in various applications. For instance, requests on web servers are highly influenced by our daily activities. Seasonal forecasting consists in forecasting the whole next season for a given seasonal time series. It may help a service provider to provision correctly the potentially required resources, avoiding critical situations of over or under provision. In [10], we propose a generic framework to make seasonal time series forecasting. The framework combines machine learning techniques 1) to identify the typical seasons and 2) to forecast the likelihood of having a season type in one season ahead. We study this framework by comparing the mean squared errors of forecasts for various settings and various datasets. The best setting is then compared to state-of-the-art time series forecasting methods. We show that it is competitive with then.

7.8. Towards Sustainable Dairy Management - A Machine Learning Enhanced Method for Estrus Detection

Our research tackles the challenge of milk production resource use efficiency in dairy farms with machine learning methods. Reproduction is a key factor for dairy farm performance since cows milk production begin with the birth of a calf. Therefore, detecting estrus, the only period when the cow is susceptible to pregnancy, is crucial for farm efficiency. Our goal is to enhance estrus detection (performance, interpretability), especially on the currently undetected silent estrus (35% of total estrus), and allow farmers to rely on automatic estrus detection solutions based on affordable data (activity, temperature). In [17] we first propose a novel approach with real-world data analysis to address both behavioral and silent estrus detection through machine learning methods. Second, we present LCE, a local cascade based algorithm that significantly outperforms a typical commercial solution for estrus detection, driven by its ability to detect silent estrus. Then, our study reveals the pivotal role of activity sensors deployment in estrus detection. Finally, we propose an approach relying on global and local (behavioral versus silent) algorithm interpretability (SHAP) to reduce the mistrust in estrus detection solutions.

7.9. A Distributed Multi-Sensor Machine Learning Approach to Earthquake Early Warning

Our research [21] aims to improve the accuracy of Earthquake Early Warning (EEW) systems by means of machine learning. EEW systems are designed to detect and characterize medium and large earthquakes before their damaging effects reach a certain location. Traditional EEW methods based on seismometers fail to accurately identify large earthquakes due to their sensitivity to the ground motion velocity. The recently introduced high-precision GPS stations, on the other hand, are ineffective to identify medium earthquakes due to its propensity to produce noisy data. In addition, GPS stations and seismometers may be deployed in large numbers across different locations and may produce a significant volume of data consequently, affecting the response time and the robust-ness of EEW systems. In practice, EEW can be seen as a typical classification problem in the machine learning field: multi-sensor data are given in input, and earthquake severity is the classification result. In this paper, we introduce the Distributed Multi-Sensor Earthquake Early Warning (DMSEEW) system, a novel machine learning-based approach that combines data from both types of sensors (GPS stations and seismometers) to detect medium and large earthquakes. DMSEEW is based on a new stacking ensemble method which has been evaluated on a real-world dataset validated with geoscientists. The system builds on a geographically distributed infrastructure, ensuring an efficient computation in terms of response time and robustness to partial infrastructure failures. Our experiments show that DMSEEW is more accurate than the traditional seismometer-only approach and the combined-sensors (GPS and seis-mometers) approach that adopts the rule of relative strength.

7.10. Dynamic Modeling of Nutrient Use and Individual Requirements of Lactating Sows

Nutrient requirements of sows during lactation are related mainly to their milk yield and feed intake, and vary greatly among individuals. In practice, nutrient requirements are generally determined at the population level based on average performance. The objective of the present modeling approach was to explore the variability in nutrient requirements among sows by combining current knowledge about nutrient use with on-farm data available on sows at farrowing [parity, BW, backfat thickness (BT)] and their individual performance (litter size, litter average daily gain, daily sow feed intake) to estimate nutrient requirements. The approach was tested on a database of 1,450 lactations from 2 farms. The effects of farm (A, B), week of lactation (W1: week 1, W2: week 2, W3+: week 3 and beyond), and parity (P1: 1, P2: 2, P3+: 3 and beyond) on sow performance and their nutrient requirements were evaluated. The mean daily ME requirement was strongly correlated with litter growth ($R2 = 0.95$; $P < 0.001$) and varied slightly according to sow BW, which influenced the maintenance cost. The mean daily standardized ileal digestible (SID) lysine requirement was influenced by farm, week of lactation, and parity. Variability in SID lysine requirement per kg feed was related mainly to feed intake

(R2 = 0.51; P < 0.001) and, to a smaller extent, litter growth (R2 = 0.27; P < 0.001). It was lowest in W1 (7.0 g/kg), greatest in W2 (7.9 g/kg), and intermediate in W3+ (7.5 g/kg; P < 0.001) because milk production increased faster than feed intake capacity did. It was lower for P3+ (6.7 g/kg) and P2 sows (7.3 g/kg) than P1 sows (8.3 g/kg) due to the greater feed intake of multiparous sows. The SID lysine requirement per kg of feed was met for 80% of sows when supplies were 112 and 120% of the mean population requirement on farm A and B, respectively, indicating higher variability in requirements on farm B. Other amino acid and mineral requirements were influenced in the same way as SID lysine. In [5], we present a modeling approach that allows us to capture individual variability in the performance of sows and litters according to farm, stage of lactation, and parity. It is an initial step in the development of new types of models able to process historical farm data (e.g., for ex post assessment of nutrient requirements) and real-time data (e.g., to control precision feeding).

7.11. Temporal Models of Care Sequences for the Exploration of Medico-administrative Data

Pharmaco-epidemiology with medico-administrative databases enables the study of the impact of health products in real-life settings. These studies require to manipulate the raw data and the care trajectories, in order to identify pieces of data that may witness the medical information that is looked for. The manipulation can be seen as a querying process in which a query is a description of a medical pattern (e.g. occurrence of illness) with the available raw features from care trajectories (e.g. occurrence of medical procedures, drug deliveries, etc.). The more expressive is the querying process, the more accurate is the medical pattern search. The temporal dimension of care trajectories is a potential information that may improve the description of medical patterns. The objective of this work [6] is to propose a formal framework that would design a well-founded tool for querying care trajectories with temporal medical patterns. In this preliminary work, we present the problematic and we introduce a use case that illustrates the comparison of several querying formalisms.

7.12. Improving Domain Adaptation By Source Selection

Domain adaptation consists in learning from a source data distribution a model that will be used on a different target data distribution. The domain adaptation procedure is usually unsuccessful if the source domain is too different from the target one. In [16], we study domain adaptation for image classification with deep learning in the context of multiple available source domains. This work proposes a multi-source domain adaptation method that selects and weights the sources based on inter-domain distances. We provide encouraging results on both classical benchmarks and a new real world application with 21 domains.

7.13. From Cost-Sensitive Classification to Tight F-measure Bounds

The F-measure is a classification performance measure, especially suited when dealing with imbalanced datasets, which provides a compromise between the precision and the recall of a classifier. As this measure is non-convex and non-linear, it is often indirectly optimized using cost-sensitive learning (that affects different costs to false positives and false negatives). In [7], we derive theoretical guarantees that give tight bounds on the best F-measure that can be obtained from cost-sensitive learning. We also give an original geometric interpretation of the bounds that serves as an inspiration for CONE, a new algorithm to optimize for the F-measure. Using 10 datasets exhibiting varied class imbalance, we illustrate that our bounds are much tighter than previous work and show that CONE learns models with either superior F-measures than existing methods or comparable but in fewer iterations.

7.14. Time Series Classification Based on Interpretable Shapelets

[19] proposes a new architecture, called AI$\longleftrightarrow$PR-CNN, composed of generative adversarial neural networks (GANs), which addresses the problem of the lack of interpretability of the existing methods for time series classification. Our network has two components: a classifier and a discriminator. The classifier is a CNN, it serves to classify series. Convolutions are discriminant patterns learned from the data that allow for a more

discriminating representation of time series (similar to a shapelet). To be able to explain the decision of the classifier, we would like to impose that the convolutions used are real "shapelets", that is to say that they are close to real sub-series present in the training set. This constraint is implemented by a GAN whose purpose will determine how much the weight matrices classifier convolutions are close to subset of the training set.

LINKS Project-Team

7. New Results

7.1. Querying Heterogeneous Linked Data

7.1.1. Data Integration and Schema Validation

Data integration requires knowledge about the structure of the various data. Such a structure is usually described by schemas. While for relational databases, schemas are hard-coded, this is not the case for many other formats. In XML for instance, several schema formalisms exists, such as DTD, XML Schema or Schematron. The Links Project-Team investigate the problem of defining schemas and use them to data, in particular for RDF and JSON Formats.

With P. Wieczorek of the University of Wroclaw, Poland, S. Staworko et al. have studied the containment problem of ShEx schemas for RDF documents in *PODS* [10].

Also, J. Dusart develops under the supervision of I. Boneva and S. Staworko the software *ShEx Validator* so as to foster the practical usage of ShEx. It is also worth noting that ShEx is now being adopted by several institutions such as *WikiData*.

7.1.2. Aggregates

Aggregation refers to computations that are alien to mere logical data manipulation (e.g. such as in relational algebra). Typically, aggregation means counting the number of answers, or performing other kinds of statistics. We have a slightly larger understanding as we may also include enumerating all answers with a *small delay*. Aggregation algorithms are generally subtle as they in most cases avoid the explicit generation of the whole set of answers. We study aggregation problems within the ANR project *Aggreg* coordinated by Niehren.

In the same spirit, Capelli et al. (in a joint work with Mengel from the CNRS in Lens) showed at *STACS* [7] a new knowledge compilation procedure which allows a polynomial algorithm to test the satisfiability quantified Boolean formulas with bounded tree width. In *Theory of Computing Systems*, [25], Capelli also gave a taxonomy of results according to various restrictions of tree-width of graphs.

Also, in *Theory of Computing Systems*, [25], Capelli gave a taxonomy of results according to various restrictions of tree-width of graphs.

Finally, in an article in *JCSS* [14], F. Capelli (with Bergougnoux and Kanté from Bordeaux and Clermont-Ferrand) propose an algorithm for counting the number of transversals (i.e. subset of nodes intersecting all hyperedges) in some hypergraphs.

7.1.3. Certain Query Answering

When data is incomplete, logical constraints and knowledge about its intended structure help to infer the answers of queries. This inference problem is known as *certain query answering*.

L. Gallois and S. Tison [6] presented in *IJCAI* - one of the main conferences of Artificial Intelligence. L. Gallois and S. Tison study boundedness of the chase procedure in the context of positive existential rules, providing decidability results for several classes and outlining the complexity of the problem. This work is done in collaboration with P. Bourhis and Graphik team-project. These results also belong to the PhD thesis of L. Gallois [11] supervised by S. Tison and P. Bourhis.

7.2. Managing Dynamic Linked Data

7.2.1. Complex Event Processing

Complex event processing requires to answer queries on streams of complex events, i.e., nested words or equivalently linearizations of data trees, but also to produce dynamically evolving data structures as output.

In an article published in *LATA* [17], I. Boneva, J. Niehren and M. Sakho studied certain query answering for hyperstreams - which are collections of connected streams - with *complex events* (i.e. that correspond to tree patterns). They showed that the problem is EXP-complete in general, and obtained PTIME algorithms when restricted to *linear* tree patterns (possibly with compression) and to deterministic tree automata.

MAGNET Project-Team

7. New Results

7.1. Natural Language Processing

Multi-Lingual Dependency Parsing

In [1], MATHIEU DEHOUCK presents his work on Word Representation and Joint Training for Syntactic Analysis. Syntactic analysis is a key step in working with natural languages. With the advances in supervised machine learning, modern parsers have reached human performances. However, despite the intensive efforts of the dependency parsing community, the number of languages for which data have been annotated is still below the hundred, and only a handful of languages have more than ten thousands annotated sentences. In order to alleviate the lack of training data and to make dependency parsing available for more languages, previous research has proposed methods for sharing syntactic information across languages. By transferring models and/or annotations or by jointly learning to parse several languages at once, one can capitalise on languages grammatical similarities in order to improve their parsing capabilities. However, while words are a key source of information for mono-lingual parsers, they are much harder to use in multi-lingual settings because they vary heavily even between very close languages. Morphological features on the contrary, are much more stable across related languages than word forms and they also directly encode syntactic information. Furthermore, it is arguably easier to annotate data with morphological information than with complete dependency structures. With the increasing availability of morphologically annotated data using the same annotation scheme for many languages, it becomes possible to use morphological information to bridge the gap between languages in multi-lingual dependency parsing.

In his thesis, MATHIEU DEHOUCK has proposed several new approaches for sharing information across languages. These approaches have in common that they rely on morphology as the adequate representation level for sharing information. We therefore also introduce a new method to analyse the role of morphology in dependency parsing relying on a new measure of morpho-syntactic complexity. The first method uses morphological information from several languages to learn delexicalised word representations that can then be used as feature and improve mono-lingual parser performances as a kind of distant supervision. The second method uses morphology as a common representation space for sharing information during the joint training of model parameters for many languages. The training process is guided by the evolutionary tree of the various language families in order to share information between languages historically related that might share common grammatical traits. We empirically compare this new training method to independently trained models using data from the Universal Dependencies project and show that it greatly helps languages with few resources but that it is also beneficial for better resourced languages when their family tree is well populated. We eventually investigate the intrinsic worth of morphological information in dependency parsing. Indeed not all languages use morphology as extensively and while some use morphology to mark syntactic relations (via cases and persons) other mostly encode semantic information (such as tense or gender). To this end, we introduce a new measure of morpho-syntactic complexity that measures the syntactic content of morphology in a given corpus as a function of preferential head attachment. We show through experiments that this new measure can tease morpho-syntactic languages and morpho-semantic languages apart and that it is more predictive of parsing results than more traditional morphological complexity measures.

Modal sense classification with task-specific context embeddings Sense disambiguation of modal constructions is a crucial part of natural language understanding. Framed as a supervised learning task, this problem heavily depends on an adequate feature representation of the modal verb context. Inspired by recent work on general word sense disambiguation, we propose in [8] a simple approach of modal sense classification in which standard shallow features are enhanced with task-specific context embedding features. Comprehensive experiments show that these enriched contextual representations fed into a simple SVM model lead to significant classification gains over shallow feature sets.

Learning Rich Event Representations and Interactions for Temporal Relation Classification Most existing systems for identifying temporal relations between events heavily rely on hand-crafted features derived from event words and explicit temporal markers. Besides, less attention has been given to automatically learning con-textualized event representations or to finding complex interactions between events. In [9], we fill this gap in showing that a combination of rich event representations and interaction learning is essential to more accurate temporal relation classification. Specifically, we propose a method in which i) Recurrent Neural Networks (RNN) extract contextual information ii) character embeddings capture morpho-semantic features (e.g. tense, mood, aspect), and iii) a deep Convolutional Neural Network (CNN) finds out intricate interactions between events. We show that the proposed approach outperforms most existing systems on the commonly used dataset while using fully automatic feature extraction and simple local inference.

Phylogenetic Multi-Lingual Dependency Parsing Languages evolve and diverge over time. Their evolutionary history is often depicted in the shape of a phylogenetic tree. Assuming parsing models are representations of their languages grammars, their evolution should follow a structure similar to that of the phylo-genetic tree. In [7], drawing inspiration from multi-task learning, we make use of the phylogenetic tree to guide the learning of multilingual dependency parsers leverag-ing languages structural similarities. Experiments on data from the Universal Dependency project show that phylogenetic training is beneficial to low resourced languages and to well furnished languages families. As a side product of phylogenetic training, our model is able to perform zero-shot parsing of previously unseen languages.

7.2. Decentralized Learning

Trade-offs in Large-Scale Distributed Tuplewise Estimation and Learning The development of cluster computing frameworks has allowed practitioners to scale out various statistical estimation and machine learning algorithms with minimal programming effort. This is especially true for machine learning problems whose objective function is nicely separable across individual data points, such as classification and regression. In contrast, statistical learning tasks involving pairs (or more generally tuples) of data points-such as metric learning, clustering or ranking-do not lend themselves as easily to data-parallelism and in-memory computing. In [13], we investigate how to balance between statistical performance and computational efficiency in such distributed tuplewise statistical problems. We first propose a simple strategy based on occasionally repartitioning data across workers between parallel computation stages, where the number of repartition-ing steps rules the trade-off between accuracy and runtime. We then present some theoretical results highlighting the benefits brought by the proposed method in terms of variance reduction, and extend our results to design distributed stochastic gradient descent algorithms for tuplewise empirical risk minimization. Our results are supported by numerical experiments in pairwise statistical estimation and learning on synthetic and real-world datasets.

Who started this rumor? Quantifying the natural differential privacy guarantees of gossip protocols Gossip protocols, also called rumor spreading or epidemic protocols, are widely used to disseminate information in massive peer-to-peer networks. These protocols are often claimed to guarantee privacy because of the uncertainty they introduce on the node that started the dissemination. But is that claim really true? Can one indeed start a gossip and safely hide in the crowd? In [14], we study gossip protocols using a rigorous mathematical framework based on differential privacy to determine the extent to which the source of a gossip can be traceable. Considering the case of a complete graph in which a subset of the nodes are curious, we derive matching lower and upper bounds on differential privacy showing that some gossip protocols achieve strong privacy guarantees. Our results further reveal an interesting tension between privacy and dissemination speed: the standard “push” gossip protocol has very weak privacy guarantees, while the optimal guarantees are attained at the cost of a drastic increase in the spreading time. Yet, we show that it is possible to leverage the inherent randomness and partial observability of gossip protocols to achieve both fast dissemination speed and near-optimal privacy.

Fully Decentralized Joint Learning of Personalized Models and Collaboration Graphs In [15], we consider the fully decentralized machine learning scenario where many users with personal datasets collaborate to learn models through local peer-to-peer exchanges, without a central coordinator. We propose to train personalized models that leverage a collaboration graph describing the relationships between the users' personal tasks, which we learn jointly with the models. Our fully decentralized optimization procedure alternates between training nonlinear models given the graph in a greedy boosting manner, and updating the collaboration graph (with controlled sparsity) given the models. Throughout the process, users exchange messages only with a small number of peers (their direct neighbors in the graph and a few random users), ensuring that the procedure naturally scales to large numbers of users. We analyze the convergence rate, memory and communication complexity of our approach, and demonstrate its benefits compared to competing techniques on synthetic and real datasets.

Advances and Open Problems in Federated Learning Federated learning (FL) is a machine learning setting where many clients (e.g. mobile devices or whole organizations) collaboratively train a model under the orchestration of a central server (e.g. service provider), while keeping the training data decentralized. FL embodies the principles of focused data collection and minimization, and can mitigate many of the systemic privacy risks and costs resulting from traditional, centralized machine learning and data science approaches. Motivated by the explosive growth in FL research, we participated in a collaborative paper [18] that discusses recent advances and presents an extensive collection of open problems and challenges.

7.3. Privacy and Machine Learning

Private Protocols for U-Statistics in the Local Model and Beyond In [16], we study the problem of computing U-statistics of degree 2, i.e., quantities that come in the form of averages over pairs of data points, in the local model of differential privacy (LDP). The class of U-statistics covers many statistical estimates of interest, including Gini mean difference, Kendall's tau coefficient and Area under the ROC Curve (AUC), as well as empirical risk measures for machine learning problems such as ranking, clustering and metric learning. We first introduce an LDP protocol based on quantizing the data into bins and applying randomized response, which guarantees an ϵ-LDP estimate with a Mean Squared Error (MSE) of $O(1/\sqrt{n}\epsilon)$ under regularity assumptions on the U-statistic or the data distribution. We then propose a specialized protocol for AUC based on a novel use of hierarchical histograms that achieves MSE of $O(\alpha^3/n\epsilon^2)$ for arbitrary data distribution. We also show that 2-party secure computation allows to design a protocol with MSE of $O(1/n\epsilon^2)$, without any assumption on the kernel function or data distribution and with total communication linear in the number of users n. Finally, we evaluate the performance of our protocols through experiments on synthetic and real datasets.

Privacy-Preserving Adversarial Representation Learning in ASR: Reality or Illusion? In [11], we study Automatic Speech Recognition (ASR), a key technology in many services and applications. This typically requires user devices to send their speech data to the cloud for ASR decoding. As the speech signal carries a lot of information about the speaker, this raises serious privacy concerns. As a solution, an encoder may reside on each user device which performs local computations to anonymize the representation. In this paper, we focus on the protection of speaker identity and study the extent to which users can be recognized based on the encoded representation of their speech as obtained by a deep encoder-decoder architecture trained for ASR. Through speaker identification and verification experiments on the Librispeech corpus with open and closed sets of speakers, we show that the representations obtained from a standard architecture still carry a lot of information about speaker identity. We then propose to use adversarial training to learn representations that perform well in ASR while hiding speaker identity. Our results demonstrate that adversarial training dramatically reduces the closed-set classification accuracy, but this does not translate into increased open-set verification error hence into increased protection of the speaker identity in practice. We suggest several possible reasons behind this negative result.

Evaluating Voice Conversion-based Privacy Protection against Informed Attackers Speech signals are a rich source of speaker-related information including sensitive attributes like identity or accent. With a small amount of found speech data, such attributes can be extracted and modeled for malicious purposes like voice cloning, spoofing, etc. In [19], we investigate speaker anonymization strategies based on voice conversion. In contrast to prior evaluations, we argue that different types of attackers can be defined depending on the extent of their knowledge about the conversion scheme. We compare two frequency warping-based conversion methods and a deep learning based method in three attack scenarios. The utility of the converted speech is measured through the word error rate achieved by automatic speech recognition, while privacy protection is assessed by state-of-the-art speaker verification techniques (i-vectors and x-vectors). Our results show that voice conversion schemes are unable to effectively protect against an attacker that has extensive knowledge of the type of conversion and how it has been applied, but may provide some protection against less knowledgeable attackers.

7.4. Learning in Graphs

Correlation Clustering with Adaptive Similarity Queries In correlation clustering, we are given n objects together with a binary similarity score between each pair of them. The goal is to partition the objects into clusters so to minimise the disagreements with the scores. In [6], we investigate correlation clustering as an active learning problem: each similarity score can be learned by making a query, and the goal is to minimise both the disagreements and the total number of queries. On the one hand, we describe simple active learning algorithms, which provably achieve an almost optimal trade-off while giving cluster recovery guarantees, and we test them on different datasets. On the other hand, we prove information-theoretical bounds on the number of queries necessary to guarantee a prescribed disagreement bound. These results give a rich characterization of the trade-off between queries and clustering error.

Flattening a Hierarchical Clustering through Active Learning In [12], we investigate active learning by pairwise similarity over the leaves of trees originating from hierarchical clustering procedures. In the realizable setting, we provide a full characterization of the number of queries needed to achieve perfect reconstruction of the tree cut. In the non-realizable setting, we rely on known important-sampling procedures to obtain regret and query complexity bounds. Our algorithms come with theoretical guarantees on the statistical error and, more importantly, lend themselves to linear-time implementations in the relevant parameters of the problem. We discuss such implementations, prove running time guarantees for them, and present preliminary experiments on real-world datasets showing the compelling practical performance of our algorithms as compared to both passive learning and simple active learning baselines.

MaxHedge: Maximising a Maximum Online In [10], we introduce a new online learning framework where, at each trial, the learner is required to select a subset of actions from a given known action set. Each action is associated with an energy value, a reward and a cost. The sum of the energies of the actions selected cannot exceed a given energy budget. The goal is to maximise the cumulative profit, where the profit obtained on a single trial is defined as the difference between the maximum reward among the selected actions and the sum of their costs. Action energy values and the budget are known and fixed. All rewards and costs associated with each action change over time and are revealed at each trial only after the learner's selection of actions. Our framework encompasses several online learning problems where the environment changes over time; and the solution trades-off between minimising the costs and maximising the maximum reward of the selected subset of actions, while being constrained to an action energy budget. The algorithm that we propose is efficient, general and may be specialised to multiple natural online combinatorial problems.

Closed-loop cycles of experiment design, execution, and learning accelerate systems biology model development in yeast One of the most challenging tasks in modern science is the development of systems biology models: Existing models are often very complex but generally have low predictive performance. The construction of high-fidelity models will require hundreds/thousands of cycles of model improvement, yet few current systems biology research studies complete even a single cycle. In [2], we combined multiple software tools with integrated laboratory robotics to execute three cycles of model improvement of the prototypical eukaryotic cellular transformation, the yeast (Saccharomyces cerevisiae) diauxic shift. In the first cycle, a model outperforming the best previous diauxic shift model was developed using bioinformatic and systems biology tools. In the second cycle, the model was further improved using automatically planned experiments. In the third cycle, hypothesis-led experiments improved the model to a greater extent than achieved using high-throughput experiments. All of the experiments were formalized and communicated to a cloud laboratory automation system (Eve) for automatic execution, and the results stored on the semantic web for reuse. The final model adds a substantial amount of knowledge about the yeast diauxic shift: 92 genes (+45%), and 1 048 interactions (+147%). This knowledge is also relevant to understanding cancer, the immune system, and aging. We conclude that systems biology software tools can be combined and integrated with laboratory robots in closed-loop cycles.

7.5. Metric Learning

Metric learning is at the core of many algorithms for learning graphs. A new software has been published in the scikit-learn contrib repository (See the Software section).

Escaping the Curse of Dimensionality in Similarity Learning: Efficient Frank-Wolfe Algorithm and Generalization Bounds Similarity and metric learning provides a principled approach to construct a task-specific similarity from weakly supervised data. However, these methods are subject to the curse of dimensionality: as the number of features grows large, poor generalization is to be expected and training becomes intractable due to high computational and memory costs. In [3], we propose a similarity learning method that can efficiently deal with high-dimensional sparse data. This is achieved through a parameterization of similarity functions by convex combinations of sparse rank-one matrices, together with the use of a greedy approximate Frank-Wolfe algorithm which provides an efficient way to control the number of active features. We show that the convergence rate of the algorithm, as well as its time and memory complexity, are independent of the data dimension. We further provide a theoretical justification of our modeling choices through an analysis of the generalization error, which depends logarithmically on the sparsity of the solution rather than on the number of features. Our experiments on datasets with up to one million features demonstrate the ability of our approach to generalize well despite the high dimensionality as well as its superiority compared to several competing methods.

metric-learn: Metric Learning Algorithms in Python In [20], we present metric-learn, an open source Python package implementing supervised and weakly-supervised distance metric learning algorithms. As part of scikit-learn-contrib, it provides a unified interface compatible with scikit-learn which allows to easily perform cross-validation, model selection, and pipelining with other machine learning estimators. metric-learn is thoroughly tested and available on PyPi under the MIT licence.

7.6. Graph Algorithms

We collaborate with the Links project team on graph-based computations and evaluation in databases.

Dependency Weighted Aggregation on Factorized Databases In [17], we study a new class of aggregation problems, called dependency weighted aggregation. The underlying idea is to aggregate the answer tuples of a query while accounting for dependencies between them, where two tuples are considered dependent when they have the same value on some attribute. The main problem we are interested in is to compute the dependency weighted count of a conjunctive query. This aggregate can be seen as a form of weighted counting, where the weights of the answer tuples are computed by solving a linear program. This linear program enforces that dependent tuples are not over represented in the final weighted count. The dependency weighted count can be used to compute the s-measure, a measure that is used in data mining to estimate the frequency of a pattern in a graph database. Computing the dependency weighted count of a conjunctive query is NP-hard in general. In this paper, we show that this problem is actually tractable for a large class of structurally restricted conjunctive queries such as acyclic or bounded hypertree width queries. Our algorithm works on a factorized representation of the answer set, in order to avoid enumerating it exhaustively. Our technique produces a succinct representation of the weighting of the answers. It can be used to solve other dependency weighted aggregation tasks, such as computing the (dependency) weighted average of the value of an attribute in the answers set.

7.7. Learning and Speech Recognition

We have worked on privacy and machine learning for speech recognition (See Section 7.3). Additional results concern kernel method for speech recognition.

Kernel Approximation Methods for Speech Recognition In [4], we study the performance of kernel methods on the acoustic modeling task for automatic speech recognition, and compare their performance to deep neural networks (DNNs). To scale the kernel methods to large data sets, we use the random Fourier feature method of Rahimi and Recht (2007). We propose two novel techniques for improving the performance of kernel acoustic models. First, we propose a simple but effective feature selection method which reduces the number of random features required to attain a fixed level of performance. Second, we present a number of metrics which correlate strongly with speech recognition performance when computed on the heldout set; we attain improved performance by using these metrics to decide when to stop training. Additionally, we show that the linear bottleneck method of Sainath et al. (2013a) improves the performance of our kernel models significantly, in addition to speeding up training and making the models more compact. Leveraging these three methods, the kernel methods attain token error rates between 0.5% better and 0.1% worse than fully-connected DNNs across four speech recognition data sets, including the TIMIT and Broadcast News benchmark tasks.

MOEX Project-Team

6. New Results

6.1. Cultural knowledge evolution

Our cultural knowledge evolution work currently focusses on alignment evolution.

Agents may use ontology alignments to communicate when they represent knowledge with different ontologies: alignments help reclassifying objects from one ontology to the other. Such alignments may be provided by dedicated algorithms [4], but their accuracy is far from satisfying. Yet agents have to proceed. They can take advantage of their experience in order to evolve alignments: upon communication failure, they will adapt the alignments to avoid reproducing the same mistake.

We performed such repair experiments [3] and revealed that, by playing simple interaction games, agents can effectively repair random networks of ontologies or even create new alignments.

6.1.1. Modelling in dynamic epistemic logic

Participants: Manuel Atencia, Jérôme Euzenat, Line Van Den Berg [Correspondent].

We explored how closely these operators resemble logical dynamics. We developed a variant of Dynamic Epistemic Logic to capture the dynamics of the cultural alignment repair game. The ontologies are modelled as knowledge and alignments as beliefs in a variant of plausibility-based dynamic epistemic logic. The dynamics of the game is achieved through (public) announcement of the game issue and the adaptation operators are defined through conservative upgrades, i.e. modalities that transform models by reordering world-plausibility. This allowed us to formally establish some limitations and redundancy of the operators [9]. More precisely, for a complete logical reasoner, the operators are redundant and some may be inconsistent with the agent knowledge.

These results hold for one agent in the game but not necessarily for the other that may not know the classes by which the alignment is repaired, nor the relations between them. The former can be dealt with by declaring that agents are aware of the signature of both ontologies (public signature assumption) but this does not allow ontologies to evolve. We are currently investigating partial semantics as a more dynamic alternative solution to this problem.

This work is part of the PhD thesis of Line van den Berg.

6.1.2. Populations

Participants: Manuel Atencia, Fatima Danash, Jérôme Euzenat [Correspondent].

We started taking the population standpoint on experimental cultural evolution. For that purpose we introduced the concept of population within the experiments. So far, a population is characterised as a set of agents sharing the same ontology. Such agents play the same alignment repair games as before with agents of other populations.

The notion of population enables to experiment with different transmission mechanisms found in cultural evolution: vertical transmission, in which culture spreads, like genes, from parents to sibblings, and horizontal transmission, in which it spreads among all members of a population. We implemented explicit horizontal transmission through a synchronisation procedure in which, at a given interval, agents of the same population exchange their knowledge, i.e. alignments.

6.1.3. Link with interactor-replicator

Participant: Jérôme Euzenat [Correspondent].

Cultural evolution may be studied at a 'macro' level, inspired from population dynamics, or at a 'micro' level, inspired from genetics. The replicator-interactor model generalises the genotype-phenotype distinction of genetic evolution. We considered how it can be applied to cultural knowledge evolution experiments [8]. More specifically, we consider knowledge as the replicator and the behaviour it induces as the interactor. We showed that this requires to address problems concerning transmission. We discussed the introduction of horizontal transmission within the replicator-interactor model and/or differential reproduction within cultural evolution experiments.

6.1.4. Experiment reproducibility

Participants: Jimmy Avae, Robin Couret, Jérôme Euzenat [Correspondent].

Experiments are described and performed in our *Lazy lavender* platform which offers scripts to specify, run, and analyse experiments. This year, we investigated expressing experiment descriptions, i.e. design, results and analysis, in RDF. This facilitates the search of experiments based on structured queries that can be expressed in SPARQL: "which experiments have been performed but not analysed?", "which experiments are derived from another specific experiment?", "which hypotheses have not been confirmed since a precise release?", "which experiments test F-measure increase?". This also suggest a better organisation of our experiment reports.

6.2. Link keys

Link keys (§3.2) are explored following two directions:

- Extracting link keys;
- Reasoning with link keys.

6.2.1. Link key extraction with relational concept analysis

Participants: Manuel Atencia, Jérôme David [Correspondent], Jérôme Euzenat.

We first described our extraction approach [1] in the framework of formal context analysis (FCA, [20]). We recently showed that link keys extracted by formal concept analysis are equivalent to an extension of those which were extracted by our former algorithm [15]. We also used pattern structures, an extension of FCA with ordered structures, to reformulate this problem [6].

Furthermore, we used relational concept analysis (RCA, [22]), an extension of FCA taking relations between concepts into account. We showed that it is possible to encode the link key extraction problem in RCA to extract the optimal link keys even in the presence of cyclic dependencies [5]. Moreover, the proposed process does not require information about the alignments between the ontologies to find out from which pairs of classes to extract link keys.

We implemented these methods and evaluated them by reproducing the experiments made in previous studies. This shows that the method extracts the expected results as well as (also expected) scalability issues.

6.2.2. Combining link keys

Participants: Manuel Atencia, Alice Caporali, Jérôme David [Correspondent], Jérôme Euzenat, Basile Legal.

For certain data sets, it may be necessary to use several link keys, even on the same pair of classes, for retrieving a more complete link set. We introduced operators to combine link keys over the same pair of classes, investigated their relations and extended measures to evaluate their quality.

We specifically proposed strategies to extract disjunctions from RDF data and apply existing quality measures to evaluate them. We experimented with these strategies showing their benefits [7].

6.2.3. Tableau method for $\mathcal{ALC}+$Link key reasoning

Participants: Manuel Atencia [Correspondent], Jérôme Euzenat, Khadija Jradeh.

Link keys can also be thought of as axioms in a description logic. As such, they can contribute to infer ABox axioms, such as links, terminological axioms, or other link keys. This has important practical applications, such as link key inference, link key consistency and link key redundancy checking. Yet, no reasoning support existed for link keys.

We previously extended the tableau method designed for the $\mathcal{ALC}$ description logic to support reasoning with link keys in $\mathcal{ALC}$. We showed how this extension enables combining link keys with classical terminological reasoning with and without ABox and TBox and generating non-trivial link keys. We further extended the method and have proven that this extended method terminates, is sound, complete, and that its complexity is 2EXPTIME [11].

This work is part of the PhD thesis of Khadija Jradeh, co-supervised with Chan Le Duc (LIASD).

ORPAILLEUR Project-Team

7. New Results

7.1. Mining of Complex Data

Participants: Nacira Abbas, Guilherme Alves Da Silva, Alexandre Bazin, Alexandre Blansché, Lydia Boudjeloud-Assala, Quentin Brabant, Brieuc Conan-Guez, Miguel Couceiro, Adrien Coulet, Sébastien Da Silva, Alain Gély, Laurine Huber, Nyoman Juniarta, Florence Le Ber, Tatiana Makhalova, Jean-François Mari, Pierre Monnin, Amedeo Napoli, Laureline Nevin, Abdelkader Ouali, François Pirot, Frédéric Pennerath, Justine Reynaud, Claire Theobald, Yannick Toussaint, Laura Alejandra Zanella Calzada, Georgios Zervakis.

7.1.1. FCA and Variations: RCA, Pattern Structures, and Biclustering

Advances in data and knowledge engineering have emphasized the needs for pattern mining tools working on complex and possibly large data. FCA, which usually applies to binary data-tables, can be adapted to work on more complex data. In this way, we have contributed to some main extensions of FCA, namely Pattern Structures, Relational Concept Analysis and application of the "Minimum Description Length" (MDL) within FCA. Pattern Structures (PS [80], [85]) allow building a concept lattice from complex data, e.g. numbers, sequences, trees and graphs. Relational Concept Analysis (RCA [90]) is able to analyze objects described both by binary and relational attributes and can play an important role in text classification and text mining. Many developments were carried out in pattern mining and FCA for improving data mining algorithms and their applicability, and for solving some specific problems such as information retrieval, discovery of functional dependencies and biclustering.

We got several results in the discovery of approximate functional dependencies [29], the mining of RDF data, the visualization of the discovered patterns, and redescription mining. Moreover, based on Relational Concept Analysis, we worked also on the discovery and representation of n-ary relations in the framework of FCA [3]. In the same way, reusing ideas form subgroup discovery, we have initiated a whole line of research on the covering of the pattern spaces based on the "Minimum Description Length" (MDL) principle and we are working on the adaptation of MDL within the FCA framework [36] [7].

We are also working on designing hybrid mining methods, based on mining methods able to deal with symbolic and numerical data in parallel. In the context of the GEENAGE project, we are interested in the identification, in biomedical data, of biomarkers that are predictive of the development of diseases in the elderly population. Actually, the data are issued from a preceding study on metabolomic data for the detection of diabetes of type 2 [23]. The problem can be viewed as a classification problem where features which are predictive of a class should be identified. This leads us to study the notions of prediction and discrimination in classification problems. Combining numerical machine learning methods such as random forests, neural networks, and SVM, then multicriteria decision making methods (Pareto fronts), and pattern mining methods (including FCA), we developed a hybrid mining approach for selecting the features which are the most predictive and/or discriminant. Then the selected features are organized within a concept lattice to be presented to the analyst together with the reasons for their selection. The concept lattice makes more easy and natural the understanding of the feature selection. As such, this approach can also be seen as an explicable mining method, where the output includes the reasons for which features are selected in terms of prediction and discrimination.

In the framework of the CrossCult European Project about cultural heritage, we worked on the mining of visitor trajectories in a museum or a touristic site. We presented a theoretical and practical research work about the characterization of visitor trajectories and the mining of these trajectories as sequences [83], [84]. The mining process is based on two approaches in the framework of FCA. We focused on different types of sequences and more precisely on subsequences without any constraint and frequent contiguous subsequences. We also introduced a similarity measure allowing us to build a hierarchical classification which is used for interpretation and characterization of the trajectories. A natural extension of this research work

on the characterization of trajectories is related to recommendation, i.e. based on an actual trajectory, how to recommend next items to be visited? Biclustering is a good candidate for designing recommendation methods and we especially worked on this topic this current year. In particular, we worked on several aspects of biclustering in the framework of FCA and we also tried to build a generic and unified framework from which several biclustering methods can be derived [34], [52].

7.1.2. Redescription Mining

Redescription mining is one of the pattern mining methods developed in the team. This method aims at finding distinct common characterizations of the same objects and, reciprocally, at identifying sets of objects having multiple shared descriptions [89]. This is motivated by the idea that in scientific investigations data oftentimes have different nature. For example, they might originate from distinct sources or be cast over separate terminologies.

In order to gain insight into the phenomenon of interest, a natural task is to identify the correspondences existing between these different aspects. A practical example in biology consists in finding geographical areas having two characterizations, one in terms of their climatic profile and one in terms of the occupying species. Discovering such redescriptions can contribute to better understand the influence of climate over species distribution. Besides biology, redescription mining can be applied in many concrete domains.

Following this way, we applied redescription mining for analyzing and mining RDF data in the web of data with the objective of discovering definitions of concepts and as well disjunctions (incompatibilities) of concepts, for completing knowledge bases in a semi-automated way [41] [10]. Redescription mining is well adapted to the task as a definition is naturally based on two sides of an equation, a left-hand side and a right-hand side.

7.1.3. Text Mining

The research work in text mining is mainly based on two ongoing PhD theses. The first research subject is related to the study of discourse and argumentation structures in a text based on tree mining and redescription mining [33], while the second research work is related to the mining of Pubmed abstracts about rare diseases. In the first research line, we investigate the similarities existing between discourse and argumentation structures by aligning subtrees in a corpus where texts are annotated. Contrasting related work, here we focus on the comparison of substructures within the text and not only the matching of relations. Based on data mining techniques such as tree mining and redescription mining, we are able to show that the structures underlying discourse and argumentation can be (partially) aligned. There the annotations related to discourse and argumentation allow us to derive a mapping between the structures. In addition, the approach enables the study of similarities between diverse discourse structures, and as well the differences in terms of expressive power.

In the second research line, the objective is to discover features related to rare diseases, e.g. symptoms, related diseases, treatments, and possible disease evolution or variations. The texts to be analyzed are from Pubmed, i.e. a platform collecting millions of publications in the medical domain. This research project aims at developing new methods and tools for supporting knowledge discovery in textual data by combining methods from Natural Language Processing (NLP) and Knowledge Discovery in Databases (KDD). Here a key idea is to design an interacting and convergent process where NLP methods are used for guiding text mining and KDD methods are used for analyzing textual documents. In this way, NLP is based on extraction of general and temporal information, while KDD methods are especially based on pattern mining, FCA, and graph mining.

7.1.4. Consensus, Aggregation Functions and Multicriteria Decision Aiding Functions

Aggregation and consensus theory study processes dealing with the problem of merging or fusing several objects, e.g., numerical or qualitative data, preferences or other relational structures, into a single or several objects of similar type and that best represents them in some way. Such processes are modeled by so-called aggregation or consensus functions [81], [82]. The need to aggregate objects in a meaningful way appeared naturally in classical topics such as mathematics, statistics, physics and computer science, but it became

increasingly emergent in applied areas such as social and decision sciences, artificial intelligence and machine learning, biology and medicine.

We are working on a theoretical basis of a unified theory of consensus and to set up a general machinery for the choice and use of aggregation functions. This choice depends on properties specified by users or decision makers, the nature of the objects to aggregate as well as computational limitations due to prohibitive algorithmic complexity. This problem demands an exhaustive study of aggregation functions that requires an axiomatic treatment and classification of aggregation procedures as well as a deep understanding of their structural behavior. It also requires a representation formalism for knowledge, in our case decision rules and methods for discovering them. Typical approaches include rough-set and FCA approaches, that we aim to extend in order to increase expressivity, applicability and readability of results. Applications of these efforts already appeared and further are expected in the context of three multidisciplinary projects, namely the "Fight Heart Failure" (research project with the Faculty of Medicine in Nancy), the European H2020 "CrossCult" project, and the "ISIPA" (Interpolation, Sugeno Integral, Proportional Analogy) project.

In the context of the project RHU "Fighting Heart Failure" (that aims to identify and describe relevant bio-profiles of patients suffering from heart failure) we are dealing with biomedical data, highly complex and heterogeneous, that include, among other, sociodemographical aspects, biological and clinical features, drugs taken by the patients, etc. One of our main challenges is to define relevant aggregation operators on this heterogeneous patient data that lead to a clustering of the patients. Each cluster should correspond to a bio-profile, i.e. a subgroup of patients sharing the same form of the disease and thus the same diagnosis and medical care strategy. We are working on ways for comparing and clustering patients, namely, by defining multidimensional similarity measures on this complex and heterogeneous biomedical data. To this end, we recently proposed a novel approach, that we named "unsupervised extremely randomized trees" (UET), that is inspired by the frameworks of unsupervised random forests (URF) and of extremely randomized trees (ET). The empirical study of UET showed that it outperforms existing methods (such as URF) in running time, while giving better clustering. However, UET was implemented for numerical data only, and this is a drawback when dealing with biomedical data.

To overcome this limitation we have recently proposed an adaptation of UET [63] that is agnostic to variable types –numerical, symbolic or both–, that is robust to noise, to correlated variables, and to monotone transformations, thus drastically limiting the need for preprocessing. In addition, this provides similarity measures for clustering purposes that show outperforming results compared to state-of-the-art clustering methodologies.

Also, motivated by current trends in graph clustering for applications in the semantic web, and community identification in computer and social networks, we recently proposed a novel graph clustering method, i.e. GraphTrees [61], that is based on random decision trees to compute pairwise dissimilarities between vertices in vertex-attributed graphs. Unlike existing methodologies, it applies directly to graphs whose vertex-attributes are heterogeneous without preprocessing, and with promising results in benchmark datasets that are competitive with best known methods.

In the context of the project ISIPA, we mainly focused on the utility-based preference model in which preferences are represented as an aggregation of preferences over different attributes, structured or not, both in the numerical and qualitative settings. In the latter case, the Sugeno integral is widely used in multiple criteria decision making and decision under uncertainty, for computing global evaluations of items based on local evaluations (utilities). The combination of a Sugeno integral with local utilities is called a Sugeno utility functional (SUF). A noteworthy property of SUFs is that they represent multi-threshold decision rules. However, not all sets of multi-threshold rules can be represented by a single SUF. We showed how to represent any set of multi-threshold rules as a combination of SUFs. Moreover, we studied their potential advantages as a compact representation of large sets of rules, as well as an intermediary step for extracting rules from empirical datasets [51]. We also proposed a novel method [58] for learning sets of decision rules that optimally fit the training data and that favors short rules over long ones. This is a competitive alternative to other methods for monotonic classification as in [78].

7.2. Knowledge Discovery in Healthcare and Life Sciences

Participants: Alexandre Bazin, Miguel Couceiro, Adrien Coulet, Sébastien Da Silva, Florence Le Ber, Jean-François Mari, Pierre Monnin, Amedeo Napoli, Abdelkader Ouali, Yannick Toussaint.

7.2.1. Ontology-based Clustering of Biological Data

Biomedical objects can be characterized by ontology annotations. For example, Gene Ontology annotations provide information on the functions of genes, while Human Phenotype Ontology (HPO) annotations provide information about phenotypes associated with diseases. It is usual to consider such annotations in the analysis of biomedical data, most of the time annotations from only one single ontology. However, complex objects such as diseases can be annotated at the same time w.r.t. different ontologies, making clear distinct dimensions. We are investigating how annotations from several ontologies may be cooperating in disease classification. In particular, we classified Genetic Intellectual Disabilities, on the basis of their HPO annotations and of Gene Ontology annotations of genes known for being responsible for these diseases [88]. We used clustering algorithms based on semantic similarities that enable us to compare sets of annotations. In particular, this experiment illustrates the fact that considering several ontologies provides better results in clustering, while selecting the best set of ontologies to combine is depending on the dataset and on the classification task. This study is still going on.

7.2.2. Validation of Pharmacogenomic Knowledge

State of the art knowledge in pharmacogenomics is heterogeneous w.r.t. validation. Some units of knowledge are well validated, observed on a large population and already used in clinical practice, while a large majority of this knowledge is lacking validation and reproducibility, mainly because of scarce observation. Accordingly, validating state of the art knowledge in pharmacogenomics by mining Electronic Health Records (EHRs) is one objective of the ANR project "PractiKPharma" initiated in 2016 (http://practikpharma.loria.fr/).

To carry out this validation, we define a minimal data schema for pharmacogenomic knowledge units (PGxO ontology), which is instantiated with data of different provenance (e.g. biomedical databases, literature and EHRs). The output of this instantiation is a (unique) knowledge graph called PGxLOD (https://pgxlod.loria.fr/). We defined and applied a set of so-called "reconciliation rules" that compare and align whenever possible knowledge units of different provenance [9]. The results of these rule applications are of particular interest since they highlight knowledge units defined in various data and knowledge sources. We are continuing this effort by studying how graph convolutional networks enable us to learn and then to compare the representation of n-ary relationships in the form of graph embeddings [39].

In addition, following our participation in the Biohackathon 2018 in Paris (https://2018.biohackathon-europe.org/), we firstly updated PGxLOD and improved its quality, completeness, and interconnection with other resources. Secondly we mined PGxLOD and searched for explanations about molecular mechanisms of adverse drug responses. Preliminary results where presented at the MedInfo Conference [59].

7.2.3. Mining Electronic Health Records

In the context of the Snowball Inria Associate Team, we studied the use of Electronic Health Records (EHRs) to predict at first prescription the need for a patient to be prescribed with a reduced drug dose [6]. We particularly focused on drugs whose dosage is known to be sensitive and variable. We used data from the Stanford Hospital to construct cohorts of patients that either did or did not need a dose change for each considered drug. After feature selection, we trained Random Forest models which successfully predict whether a new patient will or not require a dose change after being prescribed one of 23 drugs among 22 drug classes. Several of these drugs are related to clinical guidelines that recommend dose reduction exclusively in the case of adverse reaction. For these cases, a reduction in dosage may be considered as a surrogate for an adverse reaction, which our system could help to predict and to prevent.

In collaboration with Stanford University, we continued studying the development of predictive models from EHR data, in particular to evaluate the risk of atherosclerotic cardiovascular diseases (ASCVD). The evaluation of ASCVD risk is crucial for deciding upon the prescription of preventive therapies such as statins and others lipid lowering therapies. The prevalence of these diseases is depending on subgroups in a population, such as African-American and Asian people, which are indeed under-represented in cohorts that were used to fit the model currently used in clinics to evaluate the risk of ASCVD [25]. Due to such under-representation, biases are appearing in the evaluation of the risk when considering these different subgroups in the population. Then we proposed a method and a predictive model that controls, to some extent, the variability in the prediction of ASCVD when considering such "foreign" subgroups [40].

7.3. Knowledge Engineering and Web of Data

Participants: Nacira Abbas, Alexandre Bazin, Miguel Couceiro, Adrien Coulet, Florence Le Ber, Pierre Monnin, Amedeo Napoli, Justine Reynaud, Yannick Toussaint.

A first research topic in this axis relies on knowledge discovery in the web of data. This follows the increase of data published in RDF (Resource Description Framework) format and the interest in machine processable data. The quick growth of Linked Open Data (LOD) has led to challenging aspects regarding quality assessment and data exploration of the RDF triples that shape the LOD cloud. In the team, we are particularly interested in the completeness and the quality of data and their potential to provide concept definitions in terms of necessary and sufficient conditions [73], [74]. We have proposed a novel technique based on Formal Concept Analysis which classifies subsets of RDF data into a concept lattice. This allows data exploration as well as the discovery of implication rules which are used to automatically detect possible completions of RDF data and to provide definitions. Experiments on the DBpedia knowledge base show that this kind of approach is well-founded and effective [41] [10]. In addition, it should be noticed that this research work also involves redescription mining, showing the potential complementarity between definition mining and redescription mining.

The second topic in this axis is related to dependencies [77]. In the relational database model, functional dependencies (FDs) indicate a functional relation between sets of attributes: the values of a set of attributes are determined by the values of another set of attributes. FDs can be generalized into relational dependencies, also known as "link keys" in the web of data [76]. For example, link keys may identify the same book or article in different bibliographical data sources, where a link key is a statement of the form: $\{\langle \texttt{auteur}, \texttt{creator}\rangle, \langle \texttt{titre}, \texttt{title}\rangle\}$ $linkkey$ $\langle \texttt{Livre}, \texttt{Book}\rangle$ stating that whenever an instance of the class `Livre` has the same values for properties `auteur` and `titre` as an instance of class `Book` has for properties `creator` and `title`, then they denote the same entity. Such link keys are more complex than FDs in databases in several respects and they raise new problems to solve [2].

One main objective of this research work is to follow the lines initiated in recent papers [29], and to extend to link keys the characterization of FDs and of Similarity Dependencies within FCA and pattern structures. Indeed, this is one of the objective of the ANR ELKER project. Accordingly, one purpose is to extend the initial proposals based on FCA and to provide adapted implementations. This is part of the thesis work of Nacira Abbas initiated at the end of 2018 [26]. Moreover, we are currently investigating possible connections with Relational Concept Analysis and redescription mining. We would like to study the formulation of the discovery of link keys in reusing and extending some construction heuristics that were developed in redescription mining. Actually, redescription mining is a data mining technique which aims at constructing pairs of descriptions, i.e., pairs of logical statements, one for each of two datasets, such that their support sets, i.e., the sets of objects that satisfy each statements of a pair, respectively, are most similar, as measured for example by their Jaccard index.

PETRUS Project-Team

6. New Results

6.1. The Security Properties of a PDMS (Axis 1)

Participants: Nicolas Anciaux [correspondent], Luc Bouganim, Philippe Pucheral, Iulian Sandu Popa, Guillaume Scerri.

Different Personal Data Management Systems (PDMS) solutions are emerging in both academia and industry. In terms of functionality and security properties, PDMS solutions differ significantly from traditional Data Base Management Systems (DBMS). In a journal article published in Information Systems this year [3], we take stock of the functionality and security of PDMS solutions, propose five very specific security properties to be achieved and provide a preliminary architecture to meet them based on secure hardware [3]. We also presented as a tutorial at VLDB'19 [4] and a keynote at APVP'19 a review of the literature on database and security on data management issues for secure hardware and new research directions for privacy preserving management of personal data.

6.2. SEP2P and DISPERS (Axis 2)

Participants: Luc Bouganim [correspondent], Julien Loudet, Iulian Sandu Popa.

Personal Data Management Systems (PDMS) arrive at a rapid pace allowing us to integrate all our personal data in a single place and use it for our benefit and for the benefit of the community. This leads to a significant paradigm shift since personal data become massively distributed and opens an important question: how can users/applications execute queries and computations over this massively distributed data in a secure and efficient way, relying exclusively on peer-to-peer (P2P) interactions despite covert adversaries which could be executing the query? We first proposed a Secure and Efficient Peer-to-Peer protocol (SEP2P) to randomly select the nodes that will execute the query. This protocol leverages properties of distributed hash tables (DHT) to select nodes in a way that is, at the same time, secure, random and efficient. The security and randomness stem from the fact that we know, with a very high probability, that at least one honest node contributed to the creation and attestation of this list of nodes; while the efficiency stems from the fact that very few nodes are involved in this process. Building on top of SEP2P, we designed DISPERS, a protocol that applies three design rules: (D1) imposed randomness, enforced by SEP2P, (D2) knowledge dispersion, and (D3) task compartmentalization: Each user provides profile information to indexing nodes, chosen randomly thanks to the DHT (D1). Shamir secret-sharing techniques are used to avoid that any indexing node has a full knowledge of indexed nodes (D2). Then, for each query, a set of random nodes is selected (SEP2P) to coordinate the research for query targets using the indexing nodes. Each of these random nodes learns a part of the query targets IP address but does not know the query (D2, D3). Another set of random nodes is chosen to compute of the final answer based on partial local results from targets. These nodes learn part of the results but do not know the targets, thanks to proxies, nor the meaning of these results (D2, D3). These results are the core of Julien Loudet's thesis [1]. SEP2P was published at EDBT'19 [9] while a demonstration of DISPERS was published at VLDB'19 [8].Both works were also exposed/demonstrated at BDA'19 [13] [12] and APVP'19 [14] for the French research community in databases and security and privacy.

6.3. Manifest-based Framework for Secure Decentralized Queries (Axis 2)

Participants: Riad Ladjel [correspondent], Nicolas Anciaux, Philippe Pucheral, Guillaume Scerri.

The PDMS context calls for a new decentralized way of handling processing. The challenge is to allow generic treatment of large populations of PDMS, with a double objective: to preserve the mutual trust of individuals in their PDMS, and to guarantee an honest result (calculated on the right data, with the right code). To achieve this goal, our approach introduces a computational 'manifest', stipulating its execution plan and the privacy clauses (e.g., collection rules) to be guaranteed at runtime, based on trusted hardware (e.g., Intel SGX processor). Our contributions consist of (1) a protocol for randomly assigning compute tasks to participants to prevent targeted attacks, (2) a mechanism guaranteeing global compute integrity through local-only checks (without centralized trusted third party) and (3) database countermeasures limiting the impact of hidden channel attacks from corrupted participants. These contributions resulted in articles in TrustCom'19 [7] and ISD'19 [6]. Our approach guarantees confidentiality and processing integrity, it is generic and scalable, and goes far beyond existing approaches (e.g., secure multiparty computing or differential privacy).

6.4. Mobile Participatory Sensing with Strong Privacy Guarantees (Axis 2)

Participant: Iulian Sandu Popa [correspondent].

Mobile participatory sensing (MPS) could benefit many application domains. A major domain is smart transportation, with applications such as vehicular traffic monitoring, vehicle routing, or driving behavior analysis. However, MPS's success depends on finding a solution for querying large numbers of smart phones or vehicular systems, which protects user location privacy and works in real-time. This work proposes PAMPAS, a privacy-aware mobile distributed system for efficient data aggregation in MPS. In PAMPAS, mobile devices enhanced with secure hardware, called secure probes (SPs), perform distributed query processing, while preventing users from accessing other users' data. A supporting server infrastructure (SSI) coordinates the inter-SP communication and the computation tasks executed on SPs. PAMPAS ensures that SSI cannot link the location reported by SPs to the user identities even if SSI has additional background information. Moreover, an enhanced version of the protocol, named PAMPAS$^+$, makes the system robust even against advanced hardware attacks on the SPs. Hence, the risk of user location privacy leakage remains very low even for an attacker controlling the SSI and a few corrupted SPs. Our experimental results demonstrate that these protocols work efficiently on resource constrained SPs being able to collect the data, aggregate them, and share statistics or derive models in real-time. This work has been accomplished in collaboration with NJIT and DePaul University and has been recently accepted as a journal paper (an 'Online first' version is available at https://link.springer.com/article/10.1007/s10707-019-00389-4).

6.5. Empowerment and Big Data on Personal Data: from Portability to Agency (Axis 3)

Participants: Nicolas Anciaux [correspondent], Riad Ladjel, Philippe Pucheral, Guillaume Scerri.

The current highly centralised model of personal data management is based on established business practices that have led to widespread adoption, in contrast to user-centric and privacy-oriented systems such as PDMS, which therefore need to be studied in terms of technical, economic and legal feasibility and adoptability with researchers from other disciplines. In the context of the DATAIA GDP-ERE project, we are analyzing the technical and legal conditions under which individuals can exercise their right to data portability. Over the period, we have jointly studied a new notion that characterizes the true power of the individual over his or her personal data: agency. In particular, we have shown how the notion of agency, which comes from the social sciences, can be transposed and used to our context to measure the empowerment of individuals in Big Data applications. This study led to two joint publications with law researchers over the period, in particular in the journal Dalloz IP/IT [5], as well as several international panels (see in Section Popularization, e.g., panel at BDVA forum organized in Helsinky with the European Commission, at the Annual Forum of Trans Europ Expert, etc.)

6.6. OwnCare Inria Inovation Lab

Participants: Philippe Pucheral [correspondent], Nicolas Anciaux, Luc Bouganim, Laurent Schneider.

The OwnCare IILab was created in January 2018 (see section: Bilateral Contracts with Industry) and involves the Hippocad SME and the PETRUS team around the management of medical-social data at patient's home. The objective is to build a fully decentralized and highly secure personal medical-social folder based on PlugDB, and deploy it at large scale. Besides this industrial objective, the goal is also to leverage and validate the PETRUS research contributions related to secured Personal Cloud architectures. Before the creation of the OwnCare IILab initiative, PlugDB was an advanced research prototype. It is now evolving towards a transferable product. To reach this state, a considerable effort has been made in terms of development, testing platform, validation procedures and documentation. PlugDB engine is regularly registered at APP (Agence de Protection des Programmes), for both the PlugDB hardware datasheets and the code of the PlugDB-engine. The next PlugDB code registration will cover all functionalities added since the beginning of the IILab, notably: dynamic upgrade of the embedded code, TPM-based secure boot, ad-hoc embedded stored procedures, RBAC-style access control model, aggregate computation, SSL certificate management, event/error logging mechanism. Some of these developments are highly challenging considering the embedded context and the energy consumption constraints we have to face (the current device hosting PlugDB is based on two microcontrollers – MCU – powered by small batteries). Typically, we had to implement the first coupling between a TPM and a STM MCU, a lightweight version of SSL that accommodates MCU resources and energy-saving synchronization protocols between 2 MCU.

TYREX Project-Team

6. New Results

6.1. On the Optimization of Recursive Relational Queries: Application to Graph Queries

Graph databases have received a lot of attention as they are particularly useful in many applications such as social networks, life sciences and the semantic web. Various languages have emerged to query graph databases, many of which embed forms of recursion which reveal essential for navigating in graphs. The relational model has benefited from a huge body of research in the last half century and that is why many graph databases rely on techniques of relational query engines. Since its introduction, the relational model has seen various attempts to extend it with recursion and it is now possible to use recursion in several SQL or Datalog based database systems. The optimization of recursive queries remains, however, a challenge. We propose μ-RA, a variation of the Relational Algebra equipped with a fixpoint operator for expressing recursive relational queries. μ-RA can notably express unions of conjunctive regular path queries. Leveraging the fact that this fixpoint operator makes recursive terms more amenable to algebraic transformations, we propose new rewrite rules. These rules make it possible to generate new query execution plans, that cannot be obtained with previous approaches. We have defined the syntax and semantics of μ-RA, together with the rewriting rules that we specifically devised to tackle the optimization of recursive queries. We have also conducted practical experiments that show that the newly generated plans can provide significant performance improvements for evaluating recursive queries over graphs.

These results will be presented at the SIGMOD 2020 conference [9].

6.2. An Algebra with a Fixpoint Operator for Distributed Data Collections

We propose an algebra with a fixpoint operator which is suitable for modeling recursive computations with distributed data collections. We show that under reasonable conditions this fixpoint can be evaluated by parallel loops with one final merge rather than by a global loop requiring network overhead after each iteration. We also propose rewrite rules, showing when and how filters can be pushed through recursive terms, and how to filter inside a fixpoint before a join. Experiments with the Spark platform illustrate performance gains brought by these systematic optimizations [10].

6.3. Backward Type Inference for XML Queries

Although XQuery is a statically typed, functional query language for XML data, some of its features such as upward and horizontal XPath axes are typed imprecisely. The main reason is that while the XQuery data model allows to navigate upwards and between siblings from a given XML node, the type model, e.g., regular tree types, can only describe the subtree structure of the given node. In 2015, Giuseppe Castagna and our team independently proposed a precise forward type inference system for XQuery using an extended type language that can describe not only a given XML node but also its context. Recently, as a complementary method to such forward type inference systems, we propose an enhanced backward type inference system for XQuery, based on an extended type language. Results include an exact type system for XPath axes and a sound type system for XQuery expressions.

6.4. Scalable and Interpretable Predictive Models for Electronic Health Records

Early identification of patients at risk of developing complications during their hospital stay is currently one of the most challenging issues in healthcare. Complications include hospital-acquired infections, admissions to intensive care units, and in-hospital mortality. Being able to accurately predict the patients' outcomes is a crucial prerequisite for tailoring the care that certain patients receive, if it is believed that they will do poorly without additional intervention. We consider the problem of complication risk prediction, such as patient mortality, from the electronic health records of the patients. We study the question of making predictions on the first day at the hospital, and of making updated mortality predictions day after day during the patient's stay. We are developping distributed models that are scalable and interpretable. Key insights include analyzing diagnoses known at admission and drugs served, which evolve during the hospital stay. We leverage a distributed architecture to learn interpretable models from training datasets of gigantic size. We test our analyses with more than one million of patients from hundreds of hospitals, and report on the lessons learned from these experiments.

Preliminary results were presented at the 2018 International Conference on Data Science and Applications, and extended results have been submitted for publication consideration.

6.5. What can millions of laboratory test results tell us about the temporal aspect of data quality? Study of data spanning 17 years in a clinical data warehouse.

In this work, our objective is to identify common temporal evolution profiles in biological data and to propose a semi-automated method to these patterns in a clinical data warehouse (CDW). We leveraged the CDW of the European Hospital Georges Pompidou and tracked the evolution of 192 biological parameters over a period of 17 years (for 445,0 0 0 + patients, and 131 million laboratory test results). We have identified three common profiles of evolution: discretization, breakpoints, and trends. We developed computational and statistical methods to identify these profiles in the CDW. Overall, of the 192 observed biological parameters (87,814,136 values), 135 presented at least one evolution. We identified breakpoints in 30 distinct parameters, discretizations in 32, and trends in 79. As a conclusion, we can say that our method allows the identification of several temporal events in the data. Considering the distribution over time of these events, we identified probable causes for the observed profiles: instruments or software upgrades and changes in computation formulas. We evaluated the potential impact for data reuse. Finally, we formulated recommendations to enable safe use and sharing of biological data collection to limit the impact of data evolution in retrospective and federated studies (e.g. the annotation of laboratory parameters presenting breakpoints or trends) [4].

6.6. Interactive Mapping Specification with Exemplar Tuples

While schema mapping specification is a cumbersome task for data curation specialists, it becomes unfeasible for non-expert users, who are unacquainted with the semantics and languages of the involved transformations.

In this work, we propose an interactive framework for schema mapping specification suited for non-expert users. The underlying key intuition is to leverage a few exemplar tuples to infer the underlying mappings and iterate the inference process via simple user interactions under the form of Boolean queries on the validity of the initial exemplar tuples. The approaches available so far are mainly assuming pairs of complete universal data examples, which can be solely provided by data curation experts, or are limited to poorly expressive mappings.

We present a quasi-lattice-based exploration of the space of all possible mappings that satisfy arbitrary user exemplar tuples. Along the exploration, we challenge the user to retain the mappings that fit the user's requirements at best and to dynamically prune the exploration space, thus reducing the number of user interactions. We prove that after the refinement process, the obtained mappings are correct and complete. We present an extensive experimental analysis devoted to measure the feasibility of our interactive mapping strategies and the inherent quality of the obtained mappings [2].

6.7. Schema Validation and Evolution for Graph Databases

Despite the maturity of commercial graph databases, little consensus has been reached so far on the standardization of data definition languages (DDLs) for property graphs (PG). Discussion on the characteristics of PG schemas is ongoing in many standardization and community groups. Although some basic aspects of a schema are already present in most commercial graph databases, full support is missing allowing to constraint property graphs with more or less flexibility. In this work, we show how schema validation can be enforced through homomorphisms between PG schemas and PG instances by leveraging a concise schema DDL inspired by Cypher syntax. We also briefly discuss PG schema evolution that relies on graph rewriting operations allowing to consider both prescriptive and descriptive schemas [6].

6.8. MapRepair: Mapping and Repairing under Policy Views

Mapping design is overwhelming for end users, who have to check at par the correctness of the mappings and the possible information disclosure over the exported source instance. In our tool MapRepair, we focus on the latter problem by proposing a novel practical solution to ensure that a mapping faithfully complies with a set of privacy restrictions specified as source policy views. We showcase MapRepair, that guides the user through the tasks of visualizing the results of the data exchange process with and without the privacy restrictions. MapRepair leverages formal privacy guarantees and is inherently data-independent, i.e. if a set of criteria are satisfied by the mapping statement, then it guarantees that both the mapping and the underlying instances do not leak sensitive information. Furthermore, MapRepair also allows to automatically repair an input mapping w.r.t. a set of policy views in case of information leakage. We build on various demonstration scenarios, including synthetic and real-world instances and mappings [5].

6.9. Approximate Querying on Property Graphs

Property graphs are becoming widespread when modeling data with complex structural characteristics and enhancing edges and nodes with a list of properties. We worked on the approximate evaluation of counting queries involving recursive paths on property graphs. As such queries are already difficult to evaluate over pure RDF graphs, they require an ad-hoc graph summary for their approximate evaluation on property graphs. We prove the intractability of the optimal graph summarization problem, under our algorithm's conditions. We design and implement a novel property graph summary suitable for the above queries, along with an approximate query evaluation module. Finally, we show the compactness of the obtained summaries as well as the accuracy of answering counting recursive queries on them [8].

6.10. RDF Graph Anonymization Robust to Data Linkage

Privacy is a major concern when publishing new datasets in the context of Linked Open Data (LOD). A new dataset published in the LOD is indeed exposed to privacy breaches due to the linkage to objects already present in the other datasets of the LOD. In this work, we focus on the problem of building safe anonymizations of an RDF graph to guarantee that linking the anonymized graph with any external RDF graph will not cause privacy breaches. Given a set of privacy queries as input, we study the data-independent safety problem and the sequence of anonymization operations necessary to enforce it. We provide sufficient conditions under which an anonymization instance is safe given a set of privacy queries. Additionally, we show that our algorithms for RDF data anonymization are robust in the presence of sameAs links that can be explicit or inferred by additional knowledge.

6.11. Navigating the Maze of Wikidata Query Logs

We propose an in-depth and diversified analysis of the Wikidata query logs, recently made publicly available. Although the usage of Wikidata queries has been the object of recent studies, our analysis of the query traffic reveals interesting and unforeseen findings concerning the usage, types of recursion, and the shape classification of complex recursive queries. Wikidata specific features combined with recursion let us identify a significant subset of the entire corpus that can be used by the community for further assessment. We

consider and analyze the queries across many different dimensions, such as the robotic and organic queries, the presence/absence of constants along with the correctly executed and timed out queries. A further investigation that we pursue is to find, given a query, a number of queries structurally similar to the given query. We provide a thorough characterization of the queries in terms of their expressive power, their topological structure and shape, along with a deeper understanding of the usage of recursion in these logs. We make the code for the analysis available as open source [7].

6.12. Graph Generators: State of the Art and Open Challenges

The abundance of interconnected data has fueled the design and implementation of graph generators reproducing real-world linking properties, or gauging the effectiveness of graph algorithms, techniques and applications manipulating these data. We consider graph generation across multiple subfields, such as Semantic Web, graph databases, social networks, and community detection, along with general graphs. Despite the disparate requirements of modern graph generators throughout these communities, we analyze them under a common umbrella, reaching out the functionalities, the practical usage, and their supported operations. We argue that this classification is serving the need of providing scientists, researchers and practitioners with the right data generator at hand for their work. This survey provides a comprehensive overview of the state-of-the-art graph generators by focusing on those that are pertinent and suitable for several data-intensive tasks. Finally, we discuss open challenges and missing requirements of current graph generators along with their future extensions to new emerging fields [3].

6.13. A trichotomy for regular simple path queries on graphs

We focus on the computational complexity of regular simple path queries (RSPQs). We consider the following problem RSPQ(L) for a regular language L: given an edge-labeled digraph Gand two nodes xand y, is there a simple path from x to y that forms a word belonging to L? We fully characterize the frontier between tractability and intractability for RSPQ(L). More precisely, we prove RSPQ(L)is either AC0, NL-complete or NP-complete depending on the language L. We also provide a simple characterization of the tractable fragment in terms of regular expressions. Finally, we also discuss the complexity of deciding whether a language L belongs to the fragment above. We consider several alternative representations of L: DFAs, NFAs or regular expressions, and prove that this problem is NL-complete for the first representation and PSpace-complete for the other two [1].

VALDA Project-Team

7. New Results

7.1. Foundations of data management

We obtained a number of results on the foundations of data management, i.e., in database theory.

We worked on **knowledge bases**. In our work a knowledge base consists of an incomplete database together with a set of existential rules. We investigated the problem of query answering: computing the answers that are logically entailed from the knowledge base. This brings to light the fundamental chase tool, and its different variants that have been proposed in the literature. We studied the problem of chase termination, which has applications beyond query answering, and studied its complexity for restricted but useful classes of existential rules [27].

We worked on **data integration**. In our scenario a user can access data sitting in multiple sources by means of queries over a global schema, related to the sources via mappings. Data sources often contain sensitive information, and thus an analysis is needed to verify that a schema satisfies a privacy policy, given as a set of queries whose answers should not be accessible to users. We show that source constraints can have a dramatic impact on disclosure analysis [22]. Another work related to data integration is [16], where we connect the problem of answering queries under limited accesses (e.g., using Web forms) to two foundational issues: containment of Monadic datalog (MDL) programs, and containment problems involving regular tree languages. In particular, we establish a 2EXPTIME lower bound on the problem of containment of a MDL program into a conjunctive query, resolving an open problem from the early 1990s.

We also considered some other foundational topics, further from core database topics. In [18], we establish bounds on the height of maximal finite towers (a *tower* is a sequence of words alternating between two languages in such a way that every word is a subsequence of the following word) between two regular languages. In [17], we present an online $O(\sigma|y|)$-time algorithm for finding approximate occurrences of a word x within a word y, where σ is the alphabet size.

Note that two other works in this theme will be described in the 2020 activity report, as they are published in 2020 conferences [25], [26].

7.2. Uncertainty and provenance of data

We have a strong focus on the uncertainty and provenance in databases. See [20] for a high-level introduction to the area.

In [15], we investigate the use of knowledge compilation, i.e., obtaining compact circuit-based representations of functions, for (Boolean) provenance. Some width parameters of the circuit, such as bounded treewidth or pathwidth, can be leveraged to convert the circuit to structured classes, e.g., deterministic structured NNFs (d-SDNNFs) or OBDDs. In [14], we investigate parameterizations of both database instances and queries that make query evaluation fixed-parameter tractable in combined complexity. We show that clique-frontier-guarded Datalog with stratified negation (CFG-Datalog) enjoys bilinear-time evaluation on structures of bounded treewidth for programs of bounded rule size. Such programs capture in particular conjunctive queries with simplicial decompositions of bounded width, guarded negation fragment queries of bounded CQ-rank, or two-way regular path queries. Our result is shown by translating to alternating two-way automata, whose semantics is defined via cyclic provenance circuits (cycluits) that can be tractably evaluated.

In previous work [39], [40]. we have shown that the only restrictions to database instances that make probabilistic query evaluation tractable for a large class of queries is that of having a small treewidth. In [28], [32], we provide the first large-scale experimental study of treewidth and tree decompositions of real-world database instances (25 datasets from 8 different domains, with sizes ranging from a few thousand to a few million vertices). The goal is to determine which data, if any, has reasonably low treewidth. We also show that, even when treewidth is high, using partial tree decompositions can result in data structures that can assist algorithms.

To conclude on provenance management, in [23], [24], after investigating the complexity of satisfiability and query answering for attributed DL-LiteR ontologies, we propose a new semantics, based on provenance semirings, for integrating provenance information with query answering. Finally, we establish complexity results for satisfiability and query answering under this semantics.

We also consider **other notions of incompleteness**, such as in [13], where we study the complexity of query evaluation for databases whose relations are partially ordered; the problem commonly arises when combining or transforming ordered data from multiple sources. We focus on queries in a useful fragment of SQL, namely positive relational algebra with aggregates, whose bag semantics we extend to the partially ordered setting. Our semantics leads to the study of two main computational problems: the possibility and certainty of query answers. We show that these problems are respectively NP-complete and coNP-complete, but identify tractable cases depending on the query operators or input partial orders.

Finally, we also consider uncertainty through another angle, that of learning in a dynamic environment, using techniques from **reinforcement learning** and the **multi-armed bandit** field.

In [19], we tackle the problem of *influence maximization*: finding influential users, or nodes, in a graph so as to maximize the spread of information. We study a highly generic version of influence maximization, one of optimizing influence campaigns by sequentially selecting "spread seeds" from a set of influencers, a small subset of the node population, under the hypothesis that, in a given campaign, previously activated nodes remain persistently active. We introduce an estimator on the influencers' remaining potential – the expected number of nodes that can still be reached from a given influencer – and justify its strength to rapidly estimate the desired value, relying on real data gathered from Twitter. We then describe a novel algorithm, GT-UCB, relying on probabilistic upper confidence bounds on the remaining potential.

In [21], we propose a Bayesian information-geometric approach to the exploration-exploitation trade-off in stochastic multi-armed bandits. The uncertainty on reward generation and belief is represented using the manifold of joint distributions of rewards and beliefs. Accumulated information is summarised by the barycentre of joint distributions, the pseudobelief-reward. While the pseudobelief-reward facilitates information accumulation through exploration, another mechanism is needed to increase exploitation by gradually focusing on higher rewards, the pseudobelief-focal-reward. Our resulting algorithm, BelMan, alternates between projection of the pseudobelief-focal-reward onto belief-reward distributions to choose the arm to play, and projection of the updated belief-reward distributions onto the pseudobelief-focal-reward.

In [29], we consider another form of bandits, *linear bandits*, in which the available actions correspond to arbitrary context vectors whose associated rewards follow a non-stationary linear regression model. In this setting, the unknown regression parameter is al- lowed to vary in time. To address this problem, we propose D-LinUCB, a novel optimistic algorithm based on discounted linear regression, where exponential weights are used to smoothly forget the past.

7.3. Web data management

We finally describe research more oriented towards applications.

The PhD of Karima Rafes [11] dealt with **semantic knowledge bases** and their applications to the management of scientific data, through the development of the LinkedWiki platform. Another practical work on semantic knowledge bases is [30], where we show how the edit history of of a knowledge base can help correct constraint violations.

Finally, we investigate **transparency and bias** in data management and artificial intelligence. [12] presents to the data management community the challenges raised by new regulatory frameworks in this area. In [31], we discuss the possibility for artificial intelligence systems to be used in the practice of law.

WIMMICS Project-Team

7. New Results

7.1. Users Modeling and Designing Interaction

7.1.1. Design of a User-Centered Evaluation Method for Exploratory Search Systems: Consolidation of the CheXplore plugin

Participants: Alain Giboin, Jean-Marie Dormoy, Emilie Palagi, Fabien Gandon.

Designed and implemented in the context of the PhD of Emilie Palagi [64], CheXplore is a Chrome plugin that supports the user-centered evaluation of exploratory search systems. This year, CheXplore has been consolidated, i.e., in particular, refactoring of the source code – from jQuery to JavaScript; addition of some new functionalities mentioned in Emilie Palagi's PhD thesis.

7.1.2. User Evaluation of the WASABI demonstrators

Participants: Alain Giboin, Michel Buffa, Elmahdi Korfed.

In the context of the ANR project WASABI, and in collaboration with Guillaume Pellerin (IRCAM), we specified a generic methodological framework for evaluating the WASABI musical demonstrators through their use. The demonstrators are targeted to six kinds of users: composers, musicologists, journalists, content providers, music school students and teachers, and sound-engineers.

7.1.3. Territoriality-theory-based Rules and Method for Designing Multi-device Games

Participant: Alain Giboin.

A research action performed in the context of a collaboration with Anne-Marie Dery-Pinna, Philippe Renevier (I3S, Sparks team) and Sophie Lepreux (UVHC, LAMIH Lab). Observing that "territorial behavior" occurs during human interaction at a table – i.e. that humans engaged in a collaborative task partition the table workspace into different zones (so-called personal territory, group territory and storage territory), in order to get collaborative benefits –, Scott and Carpendale [65] proposed to rely on a tabletop territoriality (or workspace partitioning) theory to support the design of collaborative digital tabletop applications. Concerned by competitive game applications involving multiple devices (e.g., tabletop, tablet, smartphone), we adapted Scott and Carpendale's theory, and, based on this adapted theory, we developed a set of rules and a method for designing the user interfaces of these multi-device applications [57]. This year, we refined this set of rules and this method after having tested them [58].

7.1.4. Linked Data Visualization

Participants: Yun Tian, Olivier Corby.

We started a collaboration with M. Winckler from I3S, UNS, on Linked Data visualization with Yun Tian, a Polytech'Nice Master internship. During this internship, we have connected the HAL open data server [0] with the MGExplorer graphic library. The result is a graphic browser for copublications. This work resulted in a server prototype [0].

7.1.5. Linked Data Path Finder

Participants: Marie Destandeau, Olivier Corby, Alain Giboin.

We started a collaboration with the ILDA Inria team from Saclay where we developped an algorithm to explore the content of remote Semantic Web triple stores.

[0] http://sparql.archives-ouvertes.fr/sparql
[0] http://sparks-vm9.i3s.unice.fr:8080/index.html

7.2. Communities and Social Interactions Analysis

7.2.1. Fake News Detection

Participants: Elena Cabrio, Serena Villata, Jérôme Delobelle.

This work is part of the DGA project RAPID CONFIRMA (COntre argumentation contre les Fausses InfoRMAtion) aiming to automatically detect fake news and limit their diffusion. In this purpose, a framework is developed to detect fake news, to reduce their propagation and to propose the best response strategies. Thus, in addition to identifying the communities propagating these fake news, our goal is to propose a method to convince a person that the information is actually false is a key element in fighting the spread of such a kind of dangerous information. To achieve this goal, we orientate our research towards the generation of counter-argumentation. Counter-argumentation is a process aiming to put forward counter-arguments in order to provide evidences against a certain argument previously proposed. In the case of fake news, in order to convince a person that the (fake) information is true, the author of the fake news will use different methods of persuasion via arguments. Thus, identifying these arguments and attacking them by using carefully constructed arguments from safe sources is a way to fight this phenomenon and its spread along the social network. More precisely, we have identified four steps to address the counter-argumentation process: (1) Identifying the arguments used in the fake news (Argument mining); (2) Determining, for each of the arguments, whether it is for or against the topic of the fake news (Stance detection); (3) Identifying the key arguments that our system must attack (Classification task); and (4) Providing a set of arguments from safe sources to attack the targeted fake arguments (Counter-Argumentation).

We are also interested in studying, from a formal point of view, how to cast the notion of interpretability (i.e. the degree to which an observer can understand the cause(s) of a result) in abstract argumentation so that the reasons leading to the acceptability of one or a set of arguments in a framework (returned by a particular semantics) may be explicitly assessed [13]. More precisely, this research question breaks down into the following sub-questions: *(i)* how to formally define and characterise the notion of *impact* of an argument with respect to the acceptability of the other arguments in the framework? and *(ii)* how does this impact play a role in the interpretation process of the acceptability of arguments in the framework?

7.2.2. Hate Speech Detection

Participants: Elena Cabrio, Alain Giboin, Sara Tonelli, Michele Corazza, Pinar Arslan, Stefano Menini.

On the topic of cyberbullying event detection and hate speech detection, we proposed a message-level cyberbullying annotation on an Instagram dataset. Moreover, we used the correlations on the Instagram dataset annotated with emotion, sentiment and bullying labels. Finally, we built a message-level emotion classifier automatically predicting emotion labels for each comment in the Vine bullying dataset. We built a session-based bullying classifier with the use of n-grams, emotion, sentiment and concept-level features. For both emotion and bullying classifiers, we used Linear Support Vector Classification. Our results showed that "anger" and "negative" labels have a positive correlation with the presence of bullying. Concept-level features, emotion and sentiment features in different levels contribute to the bullying classifier, especially to the bullying class. Our best performing bullying classifier with n-grams and concept-level features (e.g., polarity, averaged polarity intensity, moodtags and semantics features) reached to an F1-score of 0.65 for bullying class and a macro average F1-score of 0.7520. The results of this research have been published at SAC 2019 [7].

Together with some colleagues at FBK Trento, we performed a comparative evaluation on datasets for hate speech detection in Italian, extracted from four different social media platforms, i.e. Facebook, Twitter, Instagram and WhatsApp. We showed that combining such platform-dependent datasets to take advantage of training data developed for other platforms is beneficial, although their impact varies depending on the social network under consideration. The results of this research have been published at SAC 2019 [11].

7.3. Vocabularies, Semantic Web and Linked Data Based Knowledge Representation and Artificial Intelligence Formalisms on the Web

7.3.1. Semantic Web for Biodiversity

Participants: Franck Michel, Catherine Faron Zucker, Antonia Ettorre.

The development of an activity related to biodiversity data sharing and integration is going on through the sustained collaboration with the "Muséum National d'Histoire Naturelle" of Paris (MNHN).

First, at the very end of 2018, we published a journal paper about the SPARQL Micro-Services architecture and how this can be useful in the biodiversty domain [62]. Then, through the internship of a Ubinet master student, we explored how SPARQL Micro-Services can help biologists in editing taxonomic information by confronting multiple, heterogeneous biodiversity-related data sources. We presented some results of this work at the Biodiversty_Next conference 2019 [28].

Within the same internship we continued the work meant to publish biodiversity data as linked data (TAXREF-LD [0]). The goal is to extend the dataset from simple taxonomic data to new types of data: species interactions, multi-lingual names, conservation and legal statuses. This work should lead to a publication in 2020.

During the last two years, we have lead the biodiversity task within the Bioschemas.org W3C community group that seeks the definition and adoption of common biology-related markup terms. We proposed the creation of the Taxon term [0] whose adoption in Schema.org is under discussion. The work now starts bearing fruits as 180.000+ webpages of the MNHN are now annotated with the Taxon term, paving the way to more biodiversity resources being published as structured data that search engines can process to provide more accurate search results.

7.3.2. Semantic Web for eEducation

Participants: Catherine Faron Zucker, Géraud Fokou Pelap.

In the framework of the EduMICS project we developed and populated an ontology to represent the students' activity on the Educlever learning platform.

7.3.3. Semantic Web for B2B applications

Participants: Molka Dhouib, Catherine Faron Zucker, Andrea Tettamanzi.

In the framework of the collaborative project with Silex France company aiming to model the social network of service providers and companies, as a preliminary step, we developed an ontology alignment approach combining word embedding and the radius measure to detect matching concepts and determining equivalence or hierarchical relations between them. We report and discuss the results of the evaluation of our approach on the OAEI complex alignment benchmark and on the SILEX use case: aligning reference vocabularies to annotate B2B services (ESCO to Cigref, ESCO to ROME, NAF to kompass and NAF to Silex activity domains) [35].

7.3.4. Integration of Heterogeneous Data Sources

Participants: Franck Michel, Catherine Faron Zucker, Fabien Gandon.

With the incentive of fostering the integration of Linked Data and non RDF data sources, we continued the work initiated around the SPARQL Micro-Service architecture that harnesses the Semantic Web standards to enable automatic combination of Linked Data and data residing in Web APIs. We published a paper at the LDOW workshop of the Web Conference that explores how we can leverage Schema.org to enable web-scale discovery and querying of Web APIs using SPARQL micro-services [27].

7.3.5. Uncertainty in the Semantic Web

Participants: Ahmed El Amine Djebri, Fabien Gandon, Andrea Tettamanzi.

In the framework of Ahmed El Amine Djebri's thesis, we proposed an approach to publishing uncertainty on the Semantic Web [15] and to link and negote uncertainty theories [14].

7.3.6. Uncertainty in Human Geography

Participant: Andrea Tettamanzi.

[0] http://agroportal.lirmm.fr/ontologies/TAXREF-LD
[0] http://bioschemas.org/devSpecs/Taxon/

In the framework of the Incertimmo collaborative project between Université Côte d'Azur and Kinaxia, we applied machine learning and urban morphology theory to the investigation of the influence of the urban environment on the value of residential real estate [6].

7.3.7. Ontology Design Rule

Participants: Olivier Corby, Catherine Faron Zucker, Philippe Martin.

We worked on the topic of Ontology Design Rules with Philippe Martin, from université de la Réunion, during his visit to the Wimmics team. This work resulted in a publication at Semantics [25].

7.3.8. Suggestion of Data Sources for SPARQL Queries over Linked Open Data

Participants: Hai Huang, Fabien Gandon.

For querying processing over Linked Open Data, suggestion of relevant data sources with respect to a SPARQL query is crucial since it highly affects the performance of querying. In this work, we focus on the problem of suggesting k relevant data sources with respect to a SPARQL query. We propose a summarization method which models the RDF graph of linked data sources and query graphs as sets of feature paths (star, sink and chain paths) and an effective algorithm to extract these feature paths for data sources and query graphs. To obtain candidate data sources we propose a time and space efficient search algorithm based on locality sensitive hashing. We perform a large-scale experiment on real world linked datasets which shows that our algorithm outperforms existing baselines.

7.4. Analyzing and Reasoning on Heterogeneous Semantic Graphs

7.4.1. SPARQL Function

Participant: Olivier Corby.

We wrote a SHACL interpreter with the LDScript language. Within the SPARQL Function LDScript [56] language we intoduced new datatypes for JSON and XML DOM. We have written a technical documentation for the whole language: http://ns.inria.fr/sparql-extension.

7.4.2. Ontology alignement approach based on Embedded Space

Participants: Molka Dhouib, Catherine Faron Zucker, Andrea Tettamanzi.

In the framework of a collaborative project with Silex France company aiming to model the social network of service providers and companies, as a preliminary step, we developed last year a dedicated vocabulary of competences and fields of activities to semantically annotate B2B service offers. This year, we proposed a new ontology alignment approach based on a set of rules exploiting the embedded space and measuring clusters of labels to discover the relationship between concepts. We tested our system on the OAEI conference complex alignment benchmark track and then applied it to aligning ontologies in a real-world case study of Silex company. The experimental results show that the combination of word embedding and the radius measure make it possible to determine, with good accuracy, not only equivalence relations, but also hierarchical relations between concepts. This work has been presented at the 15th International Conference, SEMANTiCS 2019 [35].

7.4.3. Argument Mining and Argumentation Theory

Participants: Elena Cabrio, Shohreh Haddadan, Tobias Mayer, Milagro Teruel, Laura Alonso Alemany, Johanna Frau.

We have proposed an Argument Mining approach to political debates [23]. We have addressed this task in an empirical manner by annotating 39 political debates from the last 50 years of US presidential campaigns, creating a new corpus of 29k argument components, labeled as premises and claims. We then proposed two tasks: (1) identifying the argumentative components in such debates, and (2) classifying them as premises and claims. We showed that feature-rich SVM learners and Neural Network architectures outperform standard baselines in Argument Mining over such complex data. We released the new corpus USElecDeb60To16 and the accompanying software under free licenses to the research community. As a result of these findings, we have also realized the DISPUTool system [22]. The results of this research have been published at ACL 2019 and IJCAI 2019.

We have contributed to the definition of the ACTA tool, aiming at applying argument mining to clinical text, given the importance of argument-based decision making in medicine [26]. ACTA is a tool for automating the argumentative analysis of clinical trials. The tool is designed to support doctors and clinicians in identifying the document(s) of interest about a certain disease, and in analyzing the main argumentative content and PICO elements. The results of this research have been published at IJCAI 2019.

Finally, together with Laura Alonso Alemany (Univ. Cordoba), Johanna Frau (Univ. Cordoba) and Milagro Teruel (Univ. Cordoba), we evaluated different attention mechanisms applied over a state-of-the-art architecture for sequence labeling [18]. Argument mining is a rising area of Natural Language Processing (NLP) concerned with the automatic recognition and interpretation of argument components and their relations. Neural models are by now mature technologies to be exploited for automating the argument mining tasks, despite the issue of data sparseness. This could ease much of the manual effort involved in these tasks, taking into account heterogeneous types of texts and topics. They assessed the impact of different flavors of attention in the task of argument component detection over two datasets: essays and legal domain. They showed that attention not models the problem better but also supports interpretability. The results of this research have been published at FLAIRS 2019.

7.4.4. Mining and Reasoning on Legal Documents

Participants: Serena Villata, Cristian Cardellino, Milagro Teruel, Laura Alonso Alemany, Guido Governatori, Leendert Van Der Torre, Beishui Liao, Nir Oren.

Together with Cristian Cardellino (Univ. Cordoba), Santiago Marro (Univ. Cordoba), Milagro Teruel (Univ. Cordoba) and Laura Alonso Alemany (Univ. Cordoba), we have adapted the semi-supervised deep learning architecture known as Convolutional Ladder Networks, from the domain of computer vision, and explored how well it works for a semi-supervised Named Entity Recognition and Classification task with legal data. The idea of exploring a semi-supervised technique is to assess the impact of large amounts of unsupervised data (cheap to obtain) in specific tasks that have little annotated data, in order to develop robust models that are less prone to overfitting. In order to achieve this, first we checked the impact on a task that is easier to measure. We presented some preliminary results, however, the experiments carried out showed some interesting insights that foster further research in the topic. The results of this research have been published at FLAIRS 2019 [9].

Together with some colleagues from Data61 Queensland (Australia) and Antonino Rotolo (University of Bologna), Serena Villata proposed a framework for modelling legislative deliberation in the form of dialogues. Roughly, in legislative dialogues coalitions can dynamically change and propose rule-based theories associated with different utility functions, depending on the legislative theory the coalitions are trying to determine. The results of this research have been published at ICAIL 2019 [21].

Finally, together with Nir Oren (Univ. Aberdeen), Leendert van der Torre (Univ. Luxembourg) and Beishui Liao (Univ. Zhejiang), we defined, using hierarchical abstract normative systems (HANS), three kinds of prioritized normative reasoning approaches called Greedy, Reduction and Optimization. Then, after formulating an argumentation theory for a HANS, we showed that for a totally ordered HANS, Greedy and Reduction can be represented in argumentation by applying the weakest link and the last link principles, respectively, and Optimization can be represented by introducing additional defeats capturing the idea that for each argument that contains a norm not belonging to the maximal obeyable set then this argument should be rejected. The results of this research have been published on the Journal of Logic and Computation [3].

7.4.5. Natural Language Processing of Song Lyrics

Participants: Michael Fell, Elena Cabrio, Fabien Gandon, Alain Giboin.

We progressed our work in the WASABI ANR project in two directions. First, we tackled the problem of summarizing song lyrics. Given the peculiar structure of songs, applying generic text summarization methods to lyrics can lead to the generation of highly redundant and incoherent text. We thus proposed to enhance state-of-the-art text summarization approaches with a method inspired by audio thumbnailing. We showed how these summaries that take into account the audio nature of the lyrics outperform the generic methods according to both an automatic evaluation and human judgments. The work resulted in an RANLP publication

[17]. Second, we investigated the task of detecting swear words and other potential harmful content in lyrics. The Parental Advisory Label (PAL) is a warning label that is placed on audio recordings in recognition of profanity or inappropriate references, with the intention of alerting parents of material potentially unsuitable for children.

Since 2015, digital providers such as iTunes, Spotify, Amazon Music and Deezer also follow PAL guidelines and tag such tracks as explicit.

Nowadays, such labelling is carried out mainly manually on voluntary basis, with the drawbacks of being time consuming and therefore costly, error prone and partly a subjective task. Therefore, we compared automated methods ranging from dictionary-based lookup to state-of-the-art deep neural networks to automatically detect explicit contents in English lyrics. We showed that more complex models perform only slightly better on this task, and relying on a qualitative analysis of the data, we discussed the inherent hardness and subjectivity of the task. The work was published at the RANLP conference [16]. We are currently modelling emotion in song lyrics, with the focus on the hierarchical and sequential structure of these texts, in which lines make up segments which make up the full lyric. And later parts may be perceived differently in light of the emotion previous parts have caused.

7.4.6. RDF Mining

Participants: Thu Huong Nguyen, Andrea Tettamanzi.

In collaboration with our former PhD student Tran Duc Minh, Claudia d'Amato of the University of Bari, and Nguyen Thanh Binh of the Danang University, we made a comparison of rule evaluation metrics for EDMAR, our evolutionary approach to discover multi-relational rules from ontological knowledge bases exploiting the services of an OWL reasoner [36].

In the framework of Nguyen Thu Huong's thesis, we have proposed a grammar-based evolutionary method to mine RDF datasets for OWL class disjointness axioms [31], [30].

7.4.7. Machine Learning for Operations Research

Participant: Andrea Tettamanzi.

Together with Alberto Ceselli and Saverio Basso of the University of Milan we used machine learning techniques to understand good decompositions of linear programming problems [1].

7.4.8. Image recognition with Semantic Data

Participants: Anna Bobasheva, Fabien Gandon, François Raygagne, Frédéric Precioso.

The objective of the MonaLIA 2.0 project is to exploit the crossover between the Deep Learning methods of image analysis and knowledge-based representation and reasoning and its application to the semantic indexing of annotated works and images in JocondeLab dataset. The goal is to identify automated or semi-automated tasks to improve the annotation and information retrieval. This project was an 11-month contract with Ministry of Culture plus 6-month internship.

- Training dataset preparation

  Developed SPARQL query to extract the subsets of images to train the multi-label Deep Learning classifiers for a given set of categories

  Developed Python scripts to filter and balance training images and Joconde specific data loader

  Identified categories that are not linked by Garnier Thesaurus but visually related and extended the Joconde metadata with the new RDF triples (e.g. category "Rider" is linked to categories "Horse" and "Human being")

  Researched effects of various image transformations on the object detection performance (resizing, cropping, padding, scaling)

For the underrepresented categories (bicycle, airplane, cat, etc.) downloaded the images from the external sources such as Kaggles' "Painter by Number", the Behance Artistic Media Set, and Cleveland Museum of Art. This has been done with the internship of François Raygagne.

- Building Deep Learning model

  Adapted the pre-trained VGG16 and Inception v3 PyTorch implementations for multi-label classification of the artwork images

  Tuned models hyperparameters

  Experimented with scaling the multi-labeled for 10, 20, 40 classes

  Experimented with binary classifiers for a single category

- Classification results consumption

  Studied the possible dependencies between knowledge graph metrics and classification performance (average precision of object detection)

  Extended the Joconde metadata with prediction scores produced by the classifiers

  Included the scores into category search queries to filter and order the results to produce more relevant results

Results were presented at atelier Culture - Inria, on december 2nd, Institut national d'histoire de l'art in Paris.

7.4.9. Hospitalization Prediction

Participants: Raphaël Gazzotti, Catherine Faron Zucker, Fabien Gandon.

HealthPredict is a project conducted in collaboration with the Département d'Enseignement de Recherche en Médecine Générale (DERMG) at Université Côte d'Azur and the SynchroNext company. It aims at providing a digital health solution for the early management of patients through consultation with their general practitioner and health care circuit. Concretely, it is a predictive Artificial Intelligence interface that allows us to cross the data of symptoms, diagnosis and medical treatments of the population in real time to predict the hospitalization of a patient. We propose and evaluate different ways to enrich the features extracted from electronic medical records with ontological resources before turning them into vectors used by Machine Learning algorithms to predict hospitalization. We reported and discussed the results of our first experiments on the database PRIMEGE PACA at EGC 2019 [38] and ESWC 2019 [19]. We propose a semi-supervised approach based on DBpedia to extract medical subjects from EMRs and evaluate the impact of augmenting the features used to represent EMRs with these subjects in the task of predicting hospitalization. Our results will be presented at SAC 2020 [61]. We designed an interface to assist in the decision-making process of general practitioners that allows them to identify in patients the first signs that lead to hospitalization and medical problems to be treated as a priority. It has been presented at [55].

7.4.10. Learning Analytics and Adaptive learning

Participants: Oscar Rodríguez Rocha, Catherine Faron Zucker.

We developed semantic queries to analyse the student activity data available in the Educlever knowledge graph and the SIDES knowledge graph, showing the added value of Semantic Web modelling enabling ontology-based reasoning. The results of our analysis of the SIDES knowledge graph have been presented at the 2019 French workshop on AI and Health [39].

The faculties of medicine, all grouped together under the auspices of the *Conférence des doyens*, are collectively proposing to upgrade the SIDES solution to an innovative solution called Intelligent Health Education System 3.0 (SIDES 3.0). As part of this community-based approach, the coordination of the project will be carried out by the *Université Numérique Thématique (UNT) en Santé et Sport*, the *GIP UNESS.fr*. This structure offers an ideal national positioning for support and coordination of training centers (UFR) and also offers long-term financial sustainability.

In particular, Inria through the Wimmics research team focuses on the recommendation of existing questions to the students according to their profile. For this, research activities are performed to classify the questions present in the platform by difficulty levels according to the Bloom's revised taxonomy, considering the information contained in text of the question. Also, research activities have focused to predict the probability of the outcomes of the students to questions considering previous answers stored in the SIDES graph.

With the ultimate goal of recommending resources adapted to the student's profile and context, we developed an approach to predict the success of students when answering training or test questions by learning a student model from the SIDES knowledge graph. To learn a user model from the SIDES knowledge graph, we combine state-of-the-art features with node embeddings. Our first results will be presented at SAC 2020.

The level of complexity and specificity of the learning objective associated with a question may be a key criterion to integrate in the recommendation process. For this purpose, we developed an approach to classify the questions of the SIDES platform according to the reference Bloom's taxonomy, by extracting the level of complexity and specificity of their learning objectives from their textual descriptions with semantic rules.

ZENITH Project-Team

7. New Results

7.1. Scientific Workflows

7.1.1. User Steering in Dynamic Workflows

Participants: Renan Souza, Patrick Valduriez.

In long-lasting scientific workflow executions in HPC machines, computational scientists (users) often need to fine-tune several workflow parameters. These tunings are done through user steering actions that may significantly improve performance or improve the overall results. However, in executions that last for weeks, users can lose track of what has been adapted if the tunings are not properly registered. In [18], we address the problem of tracking online parameter fine-tuning in dynamic workflows steered by users. We propose a lightweight solution to capture and manage provenance of the steering actions online with negligible overhead. The resulting provenance database relates tuning data with data for domain, dataflow provenance, execution, and performance, and is available for analysis at runtime. We show how users may get a detailed view of execution, providing insights to determine when and how to tune. We discuss the applicability of our solution in different domains and validate it with a real workflow in Oil and Gas extraction. In this experiment, the user could determine which tuned parameters influence simulation accuracy and performance. The observed overhead for keeping track of user steering actions at runtime is negligible.

7.1.2. ProvLake: Efficient Runtime Capture of Multiworkflow Data

Participants: Renan Souza, Patrick Valduriez.

Computational Science and Engineering (CSE) projects are typically developed by multidisciplinary teams. Despite being part of the same project, each team manages its own workflows, using specific execution environments and data processing tools. Analyzing the data processed by all workflows globally is critical in a CSE project. However, this is hard because the data generated by these workflows are not integrated. In addition, since these workflows may take a long time to execute, data analysis needs to be done at runtime to reduce cost and time of the CSE project. A typical solution in scientific data analysis is to capture and relate workflow runtime data in a provenance database, thus allowing for runtime data analysis. However, such data capture competes with the running workflows, adding significant overhead to their execution. To solve this problem, we introduce a system called ProvLake [39]. While capturing the data, ProvLake logically integrates and ingests them into a provenance database ready for runtime analysis. We validate ProvLake in a real use case in Oil and Gas extraction with four workflows that process 5 TB datasets for a deep learning classifier. Compared with Komadu, the closest competing solution, our approach has much smaller overhead.

7.1.3. Adaptive Caching of Scientific Workflows in the Cloud

Participants: Gaetan Heidsieck, Christophe Pradal, Esther Pacitti, Patrick Valduriez.

We consider the efficient execution of data-intensive scientific workflows in the cloud. Since it is common for workflow users to reuse other workflows or data generated by other workflows, a promising approach for efficient workflow execution is to cache intermediate data and exploit it to avoid task re-execution. In [27], we propose an adaptive caching solution for data-intensive workflows in the cloud. Our solution is based on a new scientific workflow management architecture that automatically manages the storage and reuse of intermediate data and adapts to the variations in task execution times and output data size. We evaluated our solution by implementing it in the OpenAlea system and performing extensive experiments on real data with a data-intensive application in plant phenotyping. The results show that adaptive caching can yield major performance gains.

7.2. Query Processing

7.2.1. Top-k Query Processing Over Encrypted Data in the Cloud

Participants: Sakina Mahboubi, Reza Akbarinia, Patrick Valduriez.

Cloud computing provides users and companies with powerful capabilities to store and process their data in third-party data centers. However, the privacy of the outsourced data is not guaranteed by the cloud providers. One solution for protecting the user data against security attacks is to encrypt the data before being sent to the cloud servers. Then, the main problem is to evaluate user queries over the encrypted data.

In [12], we propose a system, called SD-TOPK (Secure Distributed TOPK), that encrypts and stores user data in a cloud across a set of nodes, and is able to evaluate top-k queries over the encrypted data. SD-TOPK comes with a novel top-k query processing algorithm that finds a set of encrypted data that is proven to contain the top-k data items. This is done without having to decrypt the data in the nodes where they are stored. In addition, we propose a powerful filtering algorithm that removes the false positives as much as possible without data decryption. We implemented and evaluated the performance of our system over synthetic and real databases. The results show excellent performance for SD-TOPK compared to TA-based approaches.

7.2.2. Parallel Query Rewriting in Key-Value Stores under Single-Key Constraints

Participant: Reza Akbarinia.

Semantic constraints bring important knowledge about the structure and the domain of data. They allow users to better exploit their data thanks to the possibility of formulating high-level queries, which use a vocabulary richer than that of the single sources. However, the constraint-based rewriting of a query may lead to a huge set of new queries, which has a consequent impact on the query answering time.

In [37], we propose a novel technique for parallelizing both the generation and the evaluation of the rewriting set of a query serving as the basis for distributed query evaluation under constraints. Our solution relies on a schema for encoding the possible rewritings of a query on an integer interval. This allows us to generate equi-size partitions of rewritings, and thus to balance the load of the parallel working units that are in charge of generating and evaluating the queries. The experimental evaluation of our technique shows a significant reduction of query rewriting and execution time by means of parallelization.

7.3. Data Analytics

7.3.1. SAVIME: Simulation Data Analysis and Visualization

Participant: Patrick Valduriez.

Limitations in current DBMSs prevent their wide adoption in scientific applications. In order to make scientific applications benefit from DBMS support, enabling declarative data analysis and visualization over scientific data, we present an in-memory array DBMS system called SAVIME. In [34], we describe the system SAVIME, along with its data model. Our preliminary evaluation show how SAVIME, by using a simple storage definition language (SDL) can outperform the state-of-the-art array database system, SciDB, during the process of data ingestion. We also show that it is possible to use SAVIME as a storage alternative for a numerical solver without affecting its scalability.

7.3.2. Massively Distributed Indexing of Time Series

Participants: Djamel Edine Yagoubi, Reza Akbarinia, Boyan Kolev, Oleksandra Levchenko, Florent Masseglia, Patrick Valduriez, Dennis Shasha.

Indexing is crucial for many data mining tasks that rely on efficient and effective similarity query processing. Consequently, indexing large volumes of time series, along with high performance similarity query processing, have became topics of high interest. For many applications across diverse domains though, the amount of data to be processed might be intractable for a single machine, making existing centralized indexing solutions inefficient.

In [20], we propose a parallel solution to construct the state of the art iSAX-based index over billions of time series by making the most of the parallel environment by carefully distributing the work load. Our solution takes advantage of frameworks such as MapReduce or Spark. We provide dedicated strategies and algorithms for a deep combination of parallelism and indexing techniques. We also propose a parallel query processing algorithm that, given a query, exploits the available processing nodes to answer the query in parallel using the constructed parallel index. We implemented our index construction and query processing algorithms, and evaluated their performance over large volumes of data (up to 4 billion time series of length 256, for a total volume of 6 TB). Our experiments demonstrate high performance of our algorithm with an indexing time of less than 2 hours for more than 1 billion time series, while the state of the art centralized algorithm needs more than 5 days. They also illustrate that our approach is able to process 10M queries in less than 140 seconds, while the state of the art centralized algorithm need almost 2300 seconds.

We have implemented our solutions in the Imitates software. The demonstration of Imitates [32] is available at http://imitates.gforge.inria.fr/. The demo visitors are able to choose query time series, see how each algorithm approximates nearest neighbors and compare times in a parallel environment.

7.3.3. Online Correlation Discovery in Sliding Windows of Time Series

Participants: Djamel Edine Yagoubi, Reza Akbarinia, Boyan Kolev, Oleksandra Levchenko, Florent Masseglia, Patrick Valduriez, Dennis Shasha.

In some important applications (such as finance, retail, etc.), we need to find correlated time series in a time window, and then continuously slide this window. Doing this efficiently in parallel could help gather important insights from the data in real time. In [30], we address the problem of continuously finding highly correlated pairs of time series over the most recent time window. Our solution, called ParCorr, uses the sketch principle for representing the time series. We implemented ParCorr on top of UPM-CEP, a Complex Event Processing streaming engine developed by our partner Universitat Politecnica de Madrid. Our solution improves the parallel processing of UPM-CEP, allowing higher throughput using less resources. An interesting aspect of our solution is the discovery of time series that are correlated to a certain subset of time series. The discovered correlations can be used to select features for training a regression model for prediction.

7.3.4. Time Series Clustering via Dirichlet Mixture Models

Participants: Khadidja Meguelati, Florent Masseglia.

Dirichlet Process Mixture (DPM) is a model used for clustering with the advantage of discovering the number of clusters automatically and offering nice properties like, *e.g.*, the potential convergence to the actual clusters in the data. These advantages come at the price of prohibitive response times, which impairs its adoption and makes centralized DPM approaches inefficient. In [35], we propose DC-DPM (Distributed Computing DPM), a parallel clustering solution that gracefully scales to millions of data points while remaining DPM compliant, which is the challenge of distributing this process. In [36], we propose HD4C (High Dimensional Data Distributed Dirichlet Clustering), a parallel clustering solution that addresses the curse of dimensionality by distributed computing and performs clustering of high dimensional data such as time series (as a function of time), hyperspectral data (as a function of wavelength) etc. For both methods, our experiments on synthetic and real world data show high performance.

7.4. Machine Learning for Biodiversity Informatics

7.4.1. Phenological Stage Annotation with Deep Convolutional Neural Networks

Participants: Titouan Lorieul, Herve Goeau, Alexis Joly.

Herbarium based phenological research offers the potential to provide novel insights into plant diversity and ecosystem processes under future climate change. The goal of this study [11], conducted in collaboration with US and French ecologists, is to automate the scoring of reproductive phenological stages within a huge amount of digitized herbaria and provide significant resources for the ecological and organismal scientific communities. Specifically, we address three questions: 1) Can fertility, i.e., the presence of reproductive structures, be automatically detected from digitized specimens using deep learning? 2) Are the detection models generalizable to different herbarium data sets? and 3) Is it possible to automatically record stages (i.e., phenophases) within longer phenological events on herbarium specimens? This is the first time that such an analysis has been conducted at this scale, on such a large number of herbarium specimens and species. The results obtained for 7782 species of plants representing angiosperms, gymnosperms, and ferns suggest that it is possible to consider large-scale phenological annotation across broad phylogenetic groups.

7.4.2. Deep Species Distribution Modelling

Participants: Benjamin Deneu, Christophe Botella, Alexis Joly.

Species distribution models (SDM) are widely used for ecological research and conservation purposes. Given a set of species occurrences and environmental data (such as climatic rasters, soil occupation, altitude, etc.), the aim is to infer the spatial distribution of the species over a given territory. In a previous work, we showed that using deep convolutional networks significantly improved predictive performance compared to conventional punctual approaches. We have deepened this methodology with two main contributions. The first one is to extend the model to explicitly take into account species co-occurrences [22]. This is achieved through a new multimodal architecture that allows the joint learning of biotic and abiotic patterns in a common representation space. The second contribution is to experiment deep SDMs at the scale of several tens of thousands of species and tens of millions of occurrences. These contributions were made possible thanks to the use of supercomputer supercomputer Jean Zay (more than 1000 GPUs) of the GENCI national infrastructure.

7.4.3. Evaluation of Species Identification and Prediction Algorithms

Participants: Alexis Joly, Herve Goeau, Christophe Botella, Benjamin Deneu, Fabian Robert Stoter.

We run a new edition of the LifeCLEF evaluation campaign [29] with the involvement of 16 research teams worldwide. The main outcomes of the 2019-th edition are:

- **GeoLifeCLEF.** The main result of the second edition of this challenge [24] is that deep convolutional models outperform the most efficient machine learning models used in ecology (such as random forests or boosted trees). In particular, they are able to transfer knowledge from animals distribution to plant distribution, which had never been shown before.
- **PlantCLEF.** The 2019-th edition of the plant identification challenge [26] was designed to evaluate automated identification on the flora of data deficient regions, tropical ones in particular. It is based on a dataset of 10K species mainly focused on the Guiana shield and the Northern Amazon rainforest, an area known to have one of the greatest diversity of plants and animals in the world. The results reveal that the identification performance in this context is considerably lower than the one obtained on temperate plants of Europe and North America. The performance of convolutional neural networks fall due to the very low number of training images for most species and the higher degree of noise that is occurring in such data.
- **Bird sounds identification.** The 2019-the edition of the BirdCLEF challenge [41] focuses on the difficult task of recognizing all birds vocalizing in omni-directional soundscape recordings. Therefore, the dataset of the previous year has been extended with more than 350 hours of manually annotated soundscapes that were recorded using 30 field recorders in Ithaca (NY, USA). The main outcome is that the recognition performance can be significantly improved thanks to sophisticated data augmentation methods adapted to the problem.

In addition to organizing these challenges, we published a synthesis of the LifeCLEF evaluation campaign since its inception in 2011. This synthesis [44] is part of a larger book published on the occasion of the 20th anniversary of the CLEF international research forum. It highlights the rapid progress that automatic identification has made over the past decade, and allows us to take a step back on the future challenges of this discipline.

7.4.4. Optimal Checkpointing for Heterogeneous Chains: How to Train Deep Neural Networks with Limited Memory

Participants: Alena Shilova, Alexis Joly.

In many deep learning tasks for biodiversity, limited GPU memory is a performance limiting factor. The use of larger image sizes, in particular, is often not possible because the back-propagation algorithm requires storing all network activation maps in memory during for the backward stage. A larger image size could improve the performance of many tasks such as the analysis of digitized herbarium beds, range modeling or early detection of crop weeds in precision agriculture.
In this work [47], done in collaboration with the REAL-OPT team, we introduce a new activation checkpointing method which allows to significantly decrease memory usage when training Deep Neural Networks with the back-propagation algorithm. Similarly to checkpointing techniques coming from the literature on Automatic Differentiation, it consists in dynamically selecting the forward activations that are saved during the training phase, and then automatically recomputing missing activations from those previously recorded. We propose an original computation model that combines two types of activation savings: either only storing the layer inputs, or recording the complete history of operations that produced the outputs (this uses more memory, but requires fewer recomputations in the backward phase), and we provide an algorithm to compute the optimal computation sequence for this model, when restricted to memory persistent sequences. We provide a PyTorch implementation that processes the entire chain, dealing with any sequential DNN whose internal layers may be arbitrarily complex and automatically executing it according to the optimal checkpointing strategy computed given a memory limit. Through extensive experiments, we show that our implementation consistently outperforms existing checkpointing approaches for a large class of networks, image sizes and batch sizes.

7.5. Machine Learning for Audio Heritage Data

Audio data is typically exploited through large repositories. For instance, music right holders face the challenge of exploiting back catalogues of significant sizes while ethnologists and ethnomusicologists need to browse daily through archives of heritage audio recordings that have been gathered across decades. The originality of our research on this aspect is to bring together our expertise in large volumes and probabilistic music signal processing to build tools and frameworks that are useful whenever audio data is to be processed in large batches. In particular, we leverage on the most recent advances in probabilistic and deep learning applied to signal processing from both academia (e.g. Telecom Paris, PANAMA & Multispeech Inria project-teams, Kyoto University) and industry (e.g. Mitsubishi, Sony), with a focus towards large scale community services.

7.5.1. Setting the State of the Art in Music Demixing

Participants: Fabian-Robert Söter, Antoine Liutkus.

We have been very active in the topic of music demixing, with a prominent role in defining the state of the art in this domain. This has been achieved through several means.

- In the previous years, we have been organizing the Signal Separation Evaluation Challenge (SiSEC), an international event in the signal processing community that is held since 2007. Its objective is to bring together researchers to evaluate their algorithms on music separation/demixing on the same data and with the same metrics. From 2016 to 2019, A. Liutkus was the lead chair of SiSEC.
- We have developed the *open-unmix* [19] software, which is a reference implementation for music source separation. For the first time, it makes it possible for any researcher to use and improve a state-of-the art implementation (MIT-licensed) in the domain. In terms of performance, open-unmix mathes the best results we observed over the years as the organizers of SiSEC. The open-unmix software won the second place at the Global Pytorch Summer Hackaton 2019 organized by FaceBook.

The *pro* private version of this software is currently under active development for transfer to industry.

- In [6], we present the field to the non-specialist researcher, in a wide-audience scientific magazine. We are also core contributors of the audio section for the position paper on the use of AI for the creation industry [48].

7.5.2. Generative Modelling for Audio

Participants: Antoine Liutkus, Fabian-Robert Söter, Mathieu Fontaine.

Discriminative training for audio signal processing is inherently limited in the sense that it boils down to assuming that the target signals are present in the input, and can be recovered through some kind of filtering, even if this involves sophisticated deep models. We move forward to a new paradigm for signal processing, in which the observed signals and time series are not assumed to comprise the totality of the target, but rather some arbitrarily degraded version of it. The objective then can be understood as *generating new content given this input*. For instance, bandwidth extension may be thought of as audio super-resolution.

Our research on generative modelling concerns both methodological/theoretical aspects and applied research. On the former, we introduce the Sliced Wasserstein Flow in our ICML paper [33], which enables the optimal transport of particles from two probability spaces in a principled way. On the latter, we study the combination of heavy-tailed probabilistic models with generative audio models for source separation in [31], [25].

Our strategy is to go beyond our current expertise on music demixing to address the new and very active topics of audio style transfer and enhancement, with large scale applications for the exploitation and repurposing of large audio corpora.

7.5.3. Robust Probabilistic Models for Time-series

Participants: Mathieu Fontaine, Antoine Liutkus, Fabian-Robert Söter.

Processing large amounts of data for denoising or analysis comes with the need to devise models that are robust to outliers and permit efficient inference. For this purpose, we advocate the use of non-Gaussian models for this purpose, which are less sensitive to data-uncertainty. Our contributions on this topic can be split in two parts. First, we develop new filtering methods that go beyond least-squares estimation. In collaboration with researchers from Telecom Paris, we introduce several methods that generalize least-squares Wiener filtering to the case of α-stable processes [2]. This work is currently also under review as a journal paper. Second, as mentioned in the previous section, we have been working on generative models for audio, with the particular twist that the deep models we consider are trained probabistically under α-stable assumptions. This has the remarkable effect of significantly augmenting robustness [31], [25].

ALICE Team

7. New Results

7.1. Curved slicing

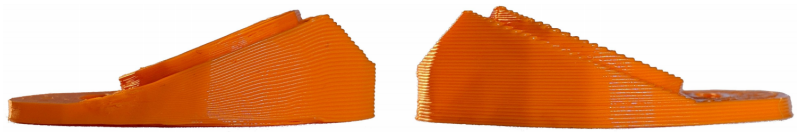


Figure 3. Sides views of curved slicing print (left) and adaptive slicing (right). Curved slicing eliminates all staircasing while closely following the input.

When printing 3D objects with Fused Filament Fabrication technology, the plastic is deposited by following a 2D path for producing the first layer. Each following layer is printed with the same method on the top of the previous layers. For technical reasons, it is convenient to use horizontal layers with constant height, but this generates aliasing errors that are especially visible (Figure 3 , right) when the object's surface is close to horizontal. The objective of this project is to reduce these artefacts by printing curved layers (Figure 3 , left). Printing curved layers is a challenging task because all technical aspects of printing have to be adapted to the curved case. The key idea of our approach is to (virtually) deform the object in such a way that the surface that is close to horizontal becomes exactly horizontal, then define all the printing instructions (tool path, slicing, pressure, etc.) in this deformed space with standard algorithms. The final printing instructions are obtained by coming back to the original space. In collaboration with MFX team, we have worked on the problem of finding the deformation by a global optimization method that tries to make horizontal large portions of the object's surface under constraints of layer thickness, tools collisions, object self-intersections, etc. The results were published at SIGGRAPH this year [7].

7.2. Coarse polycube meshes

This work is done as part of an informal (soon to be formalized) collaboration between our team and CEA. Many simulation codes require block-structured meshes. This requires decomposing the geometric domain into a set of hexahedral blocks, each one being discretized by a regular grid. Our approach to generate such structures is to generate global parameterizations. Those methods give promising results in many cases, but still face many robustness issues. To tackle those issues, we are currently working on a subset of those methods, called Polycube deformation. The idea is to deform our original domain Ω to align its boundary with a regular grid. We start by determining a set of constraints on the boundary of Ω. We then compute a map M that deforms the interior according to those constraints into a polycube. The inverse deformation M^{-1} applied to the polycube produces a structured hexmesh of the domain Ω, refer to Figure 4 . While relying on valid boundary constraints, this method is more robust than global parameterizations methods and gives good results on many models. We focus on obtaining coarse block structures, a very challenging problem with many robustness issues. Now we are able to generate as-coarse-as-possible hexahedral meshes (Figure 4 , right). We are preparing a publication of these results.

7.3. Roof fitting

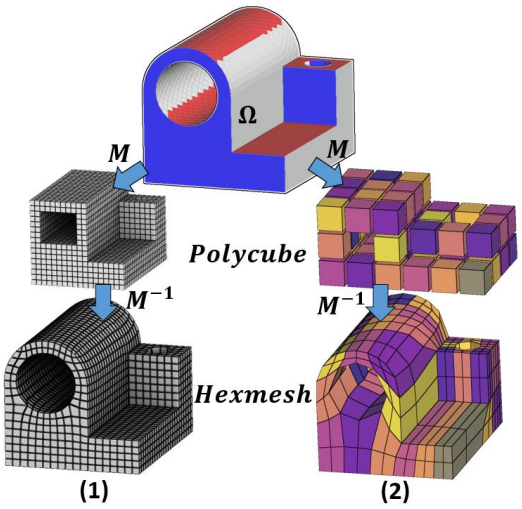


*Figure 4. The state of the art allows us to create fine polycube meshes (**left**), whereas we are trying to create meshes as coarse as possible (**right**).*

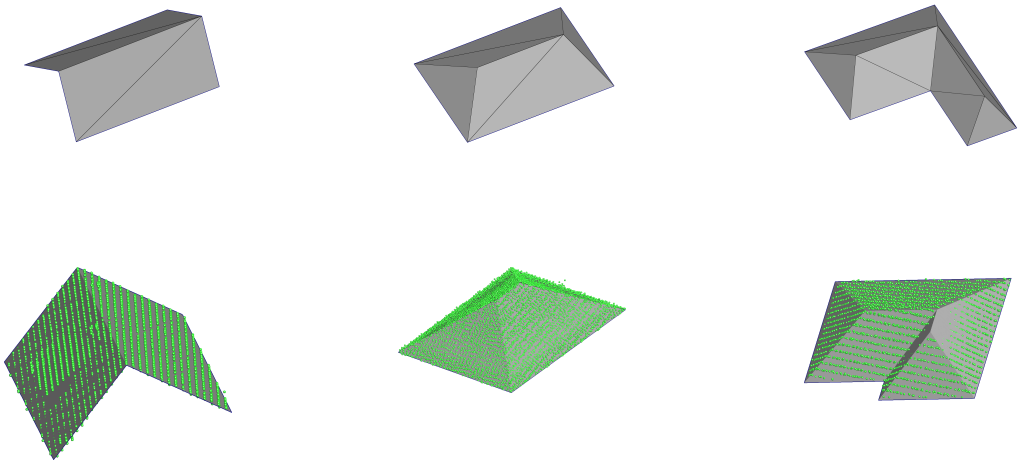


Figure 5. Top row: examples of different roof patterns. Bottom row: fitting of the patterns on LIDAR scans.

This work is done as part of an informal (soon to be formalized) collaboration between our team and RhinoTerrain. We have roof models in the form of surface meshes (Figure 5). Our data are LIDAR point clouds. Based on this data and a roof model chosen by the user, we seek to optimize the position of the model so that it "best" matches the data. This optimization must comply with two constraints:

- It is important to ignore possible outliers in the point cloud, such as parts that do not belong to the roof (trees, electrical wires, *etc.*) or should not be taken into account by the model (chimney, skylight, parabolic antenna, *etc.*);
- The roof geometry is subject to certain constraints, such as the planarity of certain rectangular faces or the alignment of certain axes.

This work is an extension of the VSDM algorithm (*Voronoï Squared Distance Minimization*) developed by the team [31]. The idea is to optimize a well-chosen energy function, the overall minimum of which corresponds to the desired position for the mesh size. The preliminary results are very promising, and we are preparing a publication.

AVIZ Project-Team

7. New Results

7.1. A Model of Spatial Directness in Interactive Visualization

Participants: Stefan Bruckner [University of Bergen], Tobias Isenberg [correspondant], Timo Ropinski [University of Ulm], Alexander Wiebel [Hochschule Worms University of Applied Sciences].

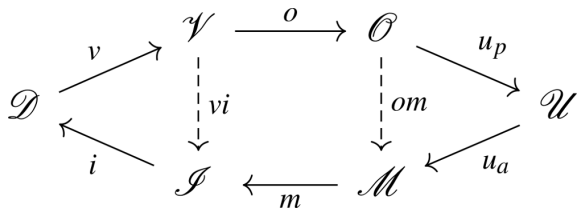


Figure 4. Illustration of the model of interaction directness.

We discuss the concept of directness in the context of spatial interaction with visualization [2] . In particular, we propose a model that allows practitioners to analyze and describe the spatial directness of interaction techniques, ultimately to be able to better understand interaction issues that may affect usability. To reach these goals, we distinguish between different types of directness (Figure 4). Each type of directness depends on a particular mapping between different spaces, for which we consider the data space, the visualization space, the output space, the user space, the manipulation space, and the interaction space. In addition to the introduction of the model itself, we also show how to apply it to several real-world interaction scenarios in visualization, and thus discuss the resulting types of spatial directness, without recommending either more direct or more indirect interaction techniques. In particular, we will demonstrate descriptive and evaluative usage of the proposed model, and also briefly discuss its generative usage.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Bruckner2019MSD.

7.2. Increasing the Transparency of Research Papers with Explorable Multiverse Analyses

Participants: Pierre Dragicevic [correspondant], Yvonne Jansen [CNRS], Abhraneel Sarma [University of Michigan], Matthew Kay [University of Michigan], Fanny Chevalier [University of Toronto].

We presented explorable multiverse analysis reports, a new approach to statistical reporting where readers of research papers can explore alternative analysis options by interacting with the paper itself [34]. This approach draws from two recent ideas: i) multiverse analysis, a philosophy of statistical reporting where paper authors report the outcomes of many different statistical analyses in order to show how fragile or robust their findings are; and ii) explorable explanations, narratives that can be read as normal explanations but where the reader can also become active by dynamically changing some elements of the explanation. Based on five examples and a design space analysis, we showed how combining those two ideas can complement existing reporting approaches and constitute a step towards more transparent research papers. This work received a best paper award at ACM CHI.

More on the project Web page, including interactive demos: https://explorablemultiverse.github.io/.

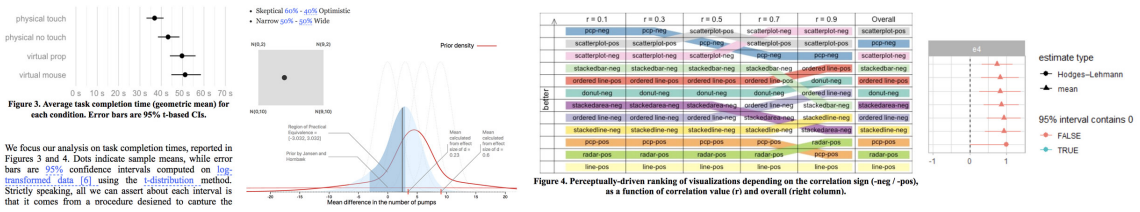


Figure 5. Examples of explorable multiverse analyses.

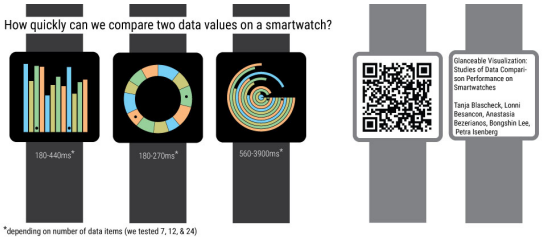


Figure 6. Comparison of bar, donut, and radial bar charts on a smartwatch.

7.3. Glanceable Visualizations for Smartwatches

Participants: Tanja Blascheck [correspondant], Lonni Besançon [Linköpping University], Anastasia Bezerianos, Bongshin Lee [Microsoft Research], Petra Isenberg.

The goal of this project is to study very small data visualizations, micro visualizations, in display contexts that can only dedicate minimal rendering space for data representations. Specifically, we define micro visualizations as small-scale visualizations that lack or have a limited set of reference structures such as labels, data axes, or grid lines and have a small physical footprint of a few square centimeters. Micro visualizations can be as simple as small unit-based visualizations such as a battery indicator but also include multi-dimensional visualizations such as star glyphs, small geographic visualizations or even small network visualizations. Although micro visualizations are essential to mobile visualization contexts, we know surprisingly little about their general visual and interaction design space or people's ability in interpreting micro visualizations. We will address this gap by proposing a common framework, conducting empirical studies to understand people's abilities to interpret these visualizations while in motion, and by developing a software toolkit to aid practitioners in developing micro visualizations for emerging mobile and wearable displays.

In summary, we aim at paving the way for a pervasive use of visualizations and thus a better and broader understanding of the complex world around us.

More information in related publications ([1],[48]) and on the project Web page: https://www.aviz.fr/smartwatchperception.

7.4. Hybrid Touch/Tangible Spatial 3D Data Selection

Participants: Lonni Besançon [Linköpping University], Mickael Sereno [correspondant], Lingyun Yu [Hangzhou Dianzi University], Mehdi Ammi [University of Paris 8], Tobias Isenberg.

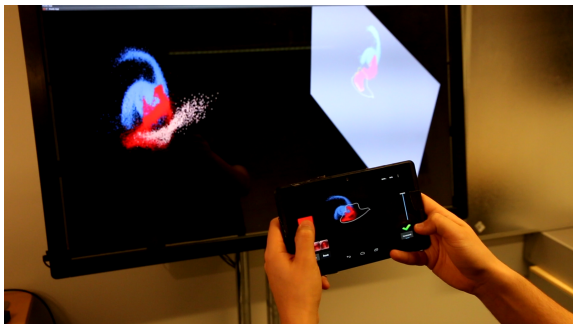


Figure 7. Illustration of Tangible Brush application which combines a spatial-aware multi-touch tablet and a remote large screen which shows different perspectives of the view shared with the tablet.

We discussed spatial selection techniques for three-dimensional datasets. Such 3D spatial selection is fundamental to exploratory data analysis. While 2D selection is efficient for datasets with explicit shapes and structures, it is less efficient for data without such properties.

We first proposed a new taxonomy of 3D selection techniques [12], focusing on the amount of control the user has to define the selection volume. We then described the 3D spatial selection technique Tangible Brush (Figure 7), which gives manual control over the final selection volume. It combines 2D touch with 6-DOF 3D tangible input to allow users to perform 3D selections in volumetric data. We use touch input to draw a 2D lasso, extruding it to a 3D selection volume based on the motion of a tangible, spatially-aware tablet. We described our approach and presented its quantitative and qualitative comparison to state-of-the-art structure-dependent selection. Our results show that, in addition to being dataset-independent, Tangible Brush is more accurate than existing dataset-dependent techniques, thus providing a trade-off between precision and effort.

7.5. Is there a reproducibility crisis around here? Maybe not, but we still need to change

Participants: Alex Holcombe [The University of Sydney], Charles Ludowici [The University of Sydney], Steve Haroz.

Those of us who study large effects may believe ourselves to be unaffected by the reproducibility problems that plague other areas [39]. However, we will argue that initiatives to address the reproducibility crisis, such as preregistration and data sharing, are worth adopting even under optimistic scenarios of high rates of replication success. We searched the text of articles published in the Journal of Vision from January through October of 2018 for URLs (our code is here: https://osf.io/cv6ed/) and examined them for raw data, experiment code, analysis code, and preregistrations. We also reviewed the articles' supplemental material. Of the 165 articles, approximately 12% provide raw data, 4% provide experiment code, and 5% provide analysis code. Only one article contained a preregistration. When feasible, preregistration is important because p-values are not interpretable unless the number of comparisons performed is known, and selective reporting appears to be common across fields. In the absence of preregistration, then, and in the context of the low rates of successful replication found across multiple fields, many claims in vision science are shrouded by uncertain credence. Sharing de-identified data, experiment code, and data analysis code not only increases credibility and ameliorates the negative impact of errors, it also accelerates science. Open practices allow researchers to build on others' work more quickly and with more confidence. Given our results and the broader context of concern by funders, evident in the recent NSF statement that "transparency is a necessary condition when designing scientifically valid research" and "pre-registration..." can help ensure the integrity and transparency of the proposed research", there is much to discuss.

EX-SITU Project-Team

7. New Results

7.1. Fundamentals of Interaction

Participants: Michel Beaudouin-Lafon [correspondant], Wendy Mackay, Cédric Fleury, Theophanis Tsandilas, Benjamin Bressolette, Julien Gori, Han Han, Yiran Zhang, Miguel Renom, Philip Tchernavskij, Martin Tricaud.

In order to better understand fundamental aspects of interaction, ExSitu conducts in-depth observational studies and controlled experiments which contribute to theories and frameworks that unify our findings and help us generate new, advanced interaction techniques. Our theoretical work also leads us to deepen or re-analyze existing theories and methodologies in order to gain new insights.

At the methodological level and in collaboration with University of Zurich (Switzerland), we have developed *Touchstone2* [19] (Best Paper award), a direct-manipulation interface for generating and examining trade-offs in experiment designs (Fig. 2). Based on interviews with experienced researchers, we developed an interactive environment for manipulating experiment design parameters, revealing patterns in trial tables, and estimating and comparing statistical power. We also developed TSL, a declarative language that precisely represents experiment designs. In two studies, experienced HCI researchers successfully used *Touchstone2* to evaluate design trade-offs and calculate how many participants are required for particular effect sizes. *Touchstone2* is freely available at https://touchstone2.org and we encourage the community to use it to improve the accountability and reproducibility of research by sharing TSL descriptions of their experimental designs.

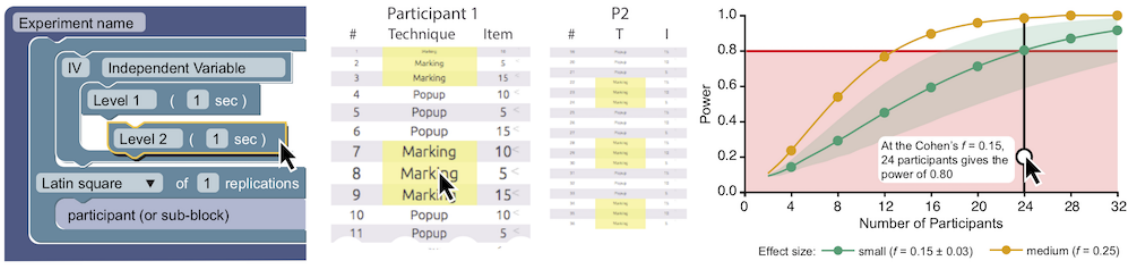


Figure 2. Touchstone2: visual language to specify an experimental design, trial table with fish-eye view, power plot.

The book "Sticky Creativity: Post-It Note Cognition, Interaction and Digitalization" [32], Academic Press, explores how the Post-It note has "become the most commonly used design material in creative design activities", with research and use cases to illustrate its role creative activities. Wendy Mackay converted her one-day Master Class on participatory design methods into a book chapter, shifting the designer's focus from static wireframes to prototyping how users will interact with a proposed new technology. The course takes the reader through a full interaction design cycle, with nine illustrated participatory design methods. It begins with a design brief: create an augmented sticky note inspired by observations of how people actually use paper sticky notes. Story-based interviews reveal both breakdowns and creative new uses of sticky notes. Brainstorming and video brainstorming, informed by the users' stories, generate new ideas. Paper prototyping a design concept related to augmented sticky notes lets designers explore ideas for a future system to address an untapped need or desire. Shooting a video prototype, guided by titlecards and a storyboard, illustrates how future users will interact with the proposed system. Finally, a design walkthrough identifies key problems and suggests ideas for improvement.

At the theoretical level, we have continued our exploration of Information Theory as a design tool for HCI by analyzing past and current applications of Shannon's theory to HCI research to identify areas where information-theoretic concepts can be used to understand, design and optimize human-computer communication [30]. We have also continued our long-standing strand of work on pointing by evaluating several models for assessing pointing performance by participants with motor impairments [27]. Namely, we studied the strengths of weaknesses of various models, from traditional Fitts' Law to the WHo model, the EMG regression and the method of Positional Variance Profiles (PVPs), on datasets from abled participants vs. participants with dyspraxia.

In the context of the ERC ONE project on Unified Principles of Interaction, Philip Tchernavskij defended his Ph.D. thesis on malleable software [40]. The goal of malleable software is to make is as easy as possible for users themselves to change software, or to have it changed on their behalf in response to their developing needs. Current approaches do not address this issue adequately: software engineering promotes flexible code, but in practice this does not help end-users effect change in their software. Based on a study of a network of communities working with biodiversity data, we found that the mode of software production, i.e. the technologies and economic relations that produce software, is biased towards centralized, one-size-fits-all systems. Instead, we should seek to create infrastructures for plurality, i.e. tools that help multiple communities collaborate without forcing them to consolidate around identical interfaces or data representations. Malleable software seeks to maximize the kinds of modifications that can take place through regular interactions, e.g. direct manipulation of interface elements. By generalizing existing control structures for interaction under the concepts of co-occurrences and entanglements, we created an environment where interactions can be dynamically created and modified. The *Tangler* prototype illustrates the power of these concepts to create malleable software.

In collaboration with Aarhus University (Denmark), we created *Videostrates* [22] to explore the notion of an *interactive substrate* for video data. *Videostrates* is based on our joint previous work on *Webstrates* (https://webstrates.net) and supports both live and recorded video composition with a declarative HTML-based notation, combining both simple and sophisticated editing tools that can be used collaboratively. *Videostrates* is programmable and unleashes the power of the modern web platform for video manipulation. We demonstrated its potential through three use scenarios (Fig. 3): collaborative video editing with multiple tools and devices; orchestration of multiple live streams that are recorded and broadcast to a popular streaming platform; and programmatic creation of video using WebGL and shaders for blue screen effects. These scenarios demonstrate *Videostrates*' potential for novel collaborative video editors with fully programmable interfaces.

Figure 3. Videostrates examples: A) Two users collaboratively edit the same videostrate, one with a timeline-based editor and the other with a subtitle editor. The results appear in a live, interactive preview on a large screen. B) Videostrates aggregates, broadcasts and records multiple live streams, here from a statically mounted camera and a smartphone. C) A Videostrate-based computational notebook uses Codestrates to programmatically create a WebGL animation and synchronize its playback with recorded video composited with a green screen.

We conducted an in-depth observational study of landscape architecture students to reveal a new phenomenon in pen-and-touch surface interaction: *interstices* [24]. We observed that bimanual interactions with a pen

and touch surface involved various sustained hand gestures, interleaved between their regular commands. Positioning of the non-preferred hand indicates anticipated actions, including: sustained hovering near the surface; pulled back but still floating above the surface; resting in their laps; and stabilizing the preferred hand while handwriting. These intersticial actions reveal anticipated actions and therefore should be taken into account in the design of novel interfaces.

We also started a study of blind or visually impaired people to better understand how they use graphical user interfaces [28]. The goal is to design multimodal interfaces for sighted users that do not rely on the visual channel as much as current GUIs.

In collaboration with the University of Paris Descartes and the ILDA Inria team, we investigated how to help users to query massive data series collections within interaction times. We demonstrated the importance of providing progressive whole-matching similarity search results on large time series collections (100 GB). Our experiments showed that there is a significant gap between the time the 1st Nearest Neighbor (1-NN) is found and the time when the search algorithm terminates [29]. In other words, users often wait without any improvement in their answers. We further showed that high-quality approximate answers are found very early, e.g., in less than one second, so they can support highly interactive visual analysis tasks. We discussed how to estimate probabilistic distance bounds, and how to help analysts evaluate the quality of their progressive results. The results of this collaboration have lead to Gogolou's Ph.D. thesis (ILDA Inria team) [38].

In the context of virtual reality, we explored how to integrate the real world surrounding users in the virtual environment. In many virtual reality systems, user physical workspace is superposed with a particular area of the virtual environment. This spatial consistency allows users to physically walk in the virtual environment and interact with virtual content through tangible objects. However, as soon as users perform virtual navigation to travel on a large scale (i.e. move their physical workspace in the virtual environment), they break this spatial consistency. We introduce two switch techniques to help users to recover the spatial consistency in some predefined virtual areas when using a teleportation technique for the virtual navigation [26]. We conducted a controlled experiment on a box-opening task in a CAVE-like system to evaluate the performance and usability of these switch techniques. The results highlight that helping the user to recover a spatially consistency ensures the accessibility of the entire virtual interaction space of the task. Consequently, the switch techniques decrease time and cognitive effort required to complete the task.

7.2. Human-Computer Partnerships

Participants: Wendy Mackay [correspondant], Baptiste Caramiaux, Téo Sanchez, Carla Griggio, Shu Yuan Hsueh, Wanyu Liu, Joanna Mcgrenere, Midas Nouwens.

ExSitu is interested in designing effective human-computer partnerships, in which expert users control their interaction with technology. Rather than treating the human users as the 'input' to a computer algorithm, we explore human-centered machine learning, where the goal is to use machine learning and other techniques to increase human capabilities. Much of human-computer interaction research focuses on measuring and improving productivity: our specific goal is to create what we call 'co-adaptive systems' that are discoverable, appropriable and expressive for the user.

In creative practices, human-centred machine learning facilitates the workflow for creatives to explore new ideas and possibilities. We compiled recent research and development advances in human-centred machine learning and artificial intelligence (AI), within the field of creative industries, in a white paper commissioned by the NEM (New European Media) initiative [35]. We explored the use of Deep Reinforcement Learning in the context of sound design with sound design experts [37]. We first conducted controlled studies where we compared manual exploration versus exploration by reinforcement. This helped us design a fully working system that we assessed in workshops with expert designers. We showed that an algorithmic sound explorer learning from human preferences enhances the creative process by allowing holistic and embodied exploration as opposed to analytic exploration afforded by standard interfaces.

We also explored how users create their own ecosystems of communication apps as a way to support rich, personalized forms of expression [12]. We wanted to gather data about how people customize apps to enable more personal forms of expression, and how such customizations shape their everyday communication. Given the increasing use of multiple apps with overlapping communication features, we were also interested in how customizing one app influences communication via other apps. We created a taxonomy of customization options based on interviews with 15 "extreme users" of communication apps. We found that participants tailored their apps to express their identities, organizational culture, and intimate bonds with others. They also experienced expression breakdowns: frustrations around barriers to transferring personal forms of expression across apps, which inspired inventive workarounds to maintain cross-app habits of expression, such as briefly switching apps to generate and export content for a particular conversation. We conclude with implications for personalized expression in ecosystems of communication apps.

We investigated the special communication practices between couples [20]. Research shows that sharing streams of contextual information, e.g. location and motion, helps couples coordinate and feel more connected. We studied how couples' communication changes when sharing multiple, persistent information streams. We designed *Lifelines*, a mobile-app technology probe that visualizes up to six streams on a shared timeline: closeness to home, battery level, steps, media playing, texts and calls. A month-long study with nine couples showed that partners interpreted information mostly from individual streams, but also combined them for more nuanced interpretations. Persistent streams allowed missing data to become meaningful and provided new ways of understanding each other. Unexpected patterns from any stream can trigger calls and texts, whereas seeing expected data can replace direct communication, which may improve or disrupt established communication practices.

Finally, we extended our earlier work on the *Expressive Keyboard* by adding animated emojis as a form of expressive output for messaging apps. An initial user study identified both the cumbersome nature of inserting emojis and the creative ways that users construct emoji sequences to convey rich, nuanced non-verbal expressions, including emphasis, change of expressions, and micro stories. We then developed *MojiBoard* [17], an emoji entry technique that lets users generate dynamic parametric emojis from a gesture keyboard. Here, the form of the user's gesture is transformed into an animation, allowing users to "draw" dynamic expressions through their own movements. *MojiBoard* lets users switch seamlessly between typing and parameterizing emojis. *MojiBoard* provides an example of how we can transform a user's gesture into an expressive output, which is reified into an emoji than can be interacted with again.

Wendy Mackay describes how the theoretical foundation of the CREATIV ERC Advance Grant, based on the principle of co-adaptation, influenced her research with musicians, choreographers, graphic designers and other creative professionals. The interview is published in the book "New Directions in Music and Human-Computer Interaction", Springer Nature, as a chapter entitled "HCI, Music and Art: An Interview with Wendy Mackay" [34]. Along the same lines, she contributed to a chapter "A Design Workbench for Interactive Music Systems" [33] that discusses possible links between the fields of computer music and human-computer interaction (HCI), particularly in the context of the MIDWAY project between Inria, France and McGill University, Canada. The goal of MIDWAY was to construct a "musical interaction design workbench" to facilitate the exploration and development of new interactive technologies for musical creation and performance by bringing together useful models, tools, and recent developments from computer music and HCI. These models and tools have helped expand the possibilities for enhancing musical expression, and provide HCI researchers with a better foundation for the design of tools for "extreme" users.

7.3. Creativity

Participants: Sarah Fdili Alaoui [correspondant], Carla Griggio, Shu Yuan Hsueh, Wendy Mackay, Baptiste Caramiaux, Joanna Mcgrenere, Midas Nouwens, Jean-Philippe Riviere, Nicolas Taffin, Philip Tchernavskij, Theophanis Tsandilas.

ExSitu is interested in understanding the work practices of creative professionals, particularly artists, designers, and scientists, who push the limits of interactive technology. We follow a multi-disciplinary participatory design approach, working with both expert and non-expert users in diverse creative contexts. We also create

situations that cause users to reflect deeply on their activities in situ and collaborate to articulate new design problems.

We conducted an interview study of 23 contemporary music composers and choreographers where we focused on the role that physical artifacts play in shaping creative collaborations with performers [13]. We found that creators and performers form relationships where the creator acts as a author, a curator, a planner, or a researcher and the performer acts as an interpreter, a creator, an improvisor, or an informant. Furthermore, we found that creators sculpt, layer and remix artifacts, moving fluidly across these different forms of interaction throughout the creative process.

We studied Kinaesthetic creativity which refers to the body's ability to generate alternate futures [21]. We probe such creative process by studying how dancers interact with technology to generate ideas. We developed a series of parameterized interactive visuals and asked dance practitioners to use them in generating movement materials. From our study, we define a taxonomy that comprises different relationships and movement responses dancers form with the visuals. We describe resulting types of interaction patterns and demonstrate how dance creativity is driven by the ability to shift between these patterns.

We used technology probes to understand how dancers learned dance fragments from videos [15]. We introduced *MoveOn*, which lets dancers decompose video into short, repeatable clips to support their learning. This served as an effective analysis tool for identifying the changes in focus and understanding dancers decomposition and recomposition processes. Additionally we compared the teacher's and dancers' decomposition strategies, and how dancers learn on their own compared to teacher-created decompositions. We found that they all ungroup and regroup dance fragments, but with different foci of attention, which suggests that teacher-imposed decomposition is more effective for introductory dance students, whereas personal decomposition is more suitable for expert dancers.

We ran a workshop [25] at ACM *Creativity and Cognition* that explored how distributed forms of creativity arising in play can help guide and foster supportive research, game design, and technology. We brought together researchers, game designers, and others to examine theories of creativity and play, game design practices, and methods for studying creativity.

We developed a taxonomy [18] on technologies using Defamiliarization to to support Co-Creation in choreographic practices. Regarding intersection of choreographic practice and HCI, Sarah Fdili Alaoui [16] describe her research and creation journey of an interactive choreographic dance piece called SKIN. This generated a set of research questions that she addresses through experience explicitation interviews of both audience and creative team members on the lived experience of making and attending the performance and the emergent relationships between dance, media and interaction as well as the tensions and negotiations that emerged from integrating technology in art. She discusses her approach as anti-solutionist and argue for more openness in HCI to allow artists to contribute.

Finally, we assessed the inter-rater reliability of the Laban Movement Analysis system used in choreography and dance notation [11].

7.4. Collaboration

Participants: Cédric Fleury [correspondant], Michel Beaudouin-Lafon, Wendy Mackay, Carla Griggio, Yujiro Okuya, Arthur Fages.

ExSitu explores new ways of supporting collaborative interaction and remote communication. In particular, we studied co-located collaboration on large wall-sized display, video-conferencing systems for remote collaboration, and collaboration between professional designers and developers during the design of interactive systems.

Multi-touch wall-sized displays, as those of the Digiscope network (http://digiscope.fr/, afford collaborative exploration of large datasets and re-organization of digital content. In the context of industrial design, computer-aided design (CAD) is now an essential part of the design process allowing experts to evaluate and adjust product design using digital mock-ups. We investigated how a wall-sized display could be used

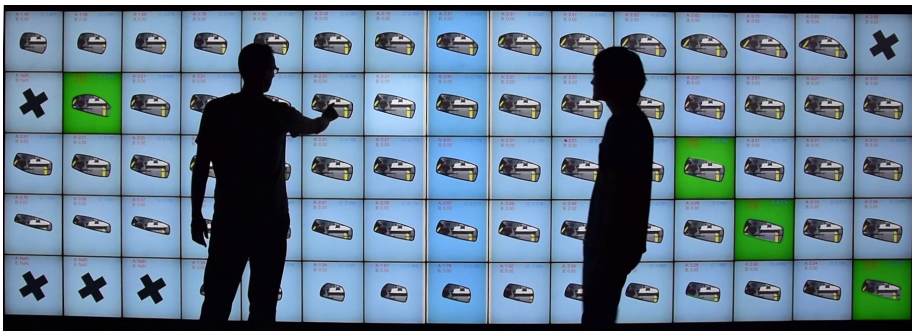


Figure 4. Collaborative exploration of multiple design alternatives of a car rear-view mirror on a wall-sized display.

to allow multidisciplinary collaborators (e.g. designers, engineers, ergonomists) to explore large number of design alternatives. In particular, we design a system which allows non-CAD expert to generate and distribute on a wall-sized display multiple various of a CAD model (Figure 4). We ran a usability study and a controlled experiment to assess the benefit of wall-sized displays in such context. Yujiro Okuya, under the supervision of Patrick Bourdot (LIMSI-CNRS) and Cédric Fleury, successfully defended his thesis *CAD Modification Techniques for Design Reviews on Heterogeneous Interactive Systems* [39] on this topic.

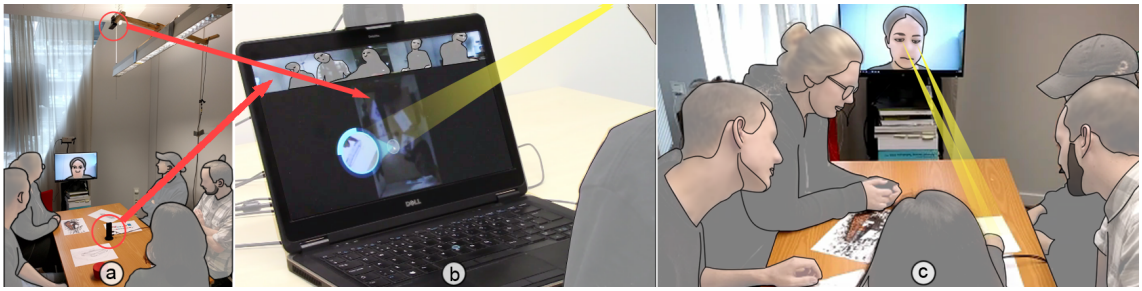


Figure 5. GazeLens system. (left) On the coworkers' side, a 360 camera on the table captures coworkers and a webcam mounted on the ceiling captures artifacts on the table. (middle) Video feeds from the two cameras are displayed on the screen of the remote satellite worker; a virtual lens strategically guides her/his attention towards a specific screen area according to the observed artifact. (right) The satellite's gaze, guided by the virtual lens, is aligned towards the observed artifact on the coworkers' space.

For remote collaboration using video, interpreting gaze direction is critical for communication between coworkers sitting around a table and a remote satellite colleague. However, 2D video distorts images and makes this interpretation inaccurate. We proposed GazeLens [23], a video conferencing system that improves coworkers' ability to interpret the satellite worker's gaze (Figure 5). A 360 camera captures the coworkers and a ceiling camera captures artifacts on the table. The system combines these two video feeds in an interface. Lens widgets strategically guide the satellite worker's attention toward specific areas of her/his screen allowing coworkers to clearly interpret her/his gaze direction. Controlled experiments showed that GazeLens increases coworkers' overall gaze interpretation accuracy in comparison to a conventional video conferencing system.

Finally, we also conducted an in-depth study of the collaboration patterns between designers and developers of interactive systems, and created a tool, *Enact*, to facilitate their work [14]. Professional designers and developers often struggle when transitioning between the design and implementation of an interactive system. We found that current practices induce unnecessary rework and cause discrepancies between design and implementation. We identified three recurring types of breakdowns: omitting critical details, ignoring edge cases, and disregarding technical limitations. We introduced four design principles to create tools that mitigate these problems: Provide multiple viewpoints, maintain a single source of truth, reveal the invisible and support design by enaction. We applied these principles to create *Enact*, a live environment for prototyping touch-based interactions (Fig. 6). We conducted two studies to assess *Enact* and compare it with current tools. Results suggest that *Enact* helps participants detect more edge cases, increases designers' participation and provides new opportunities for co-creation.

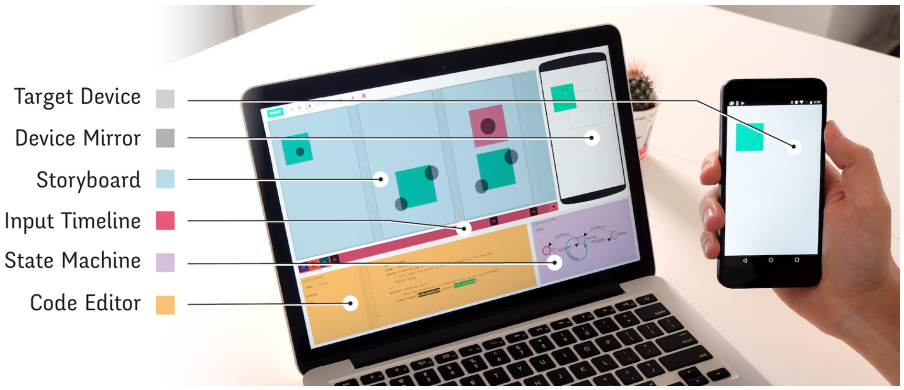
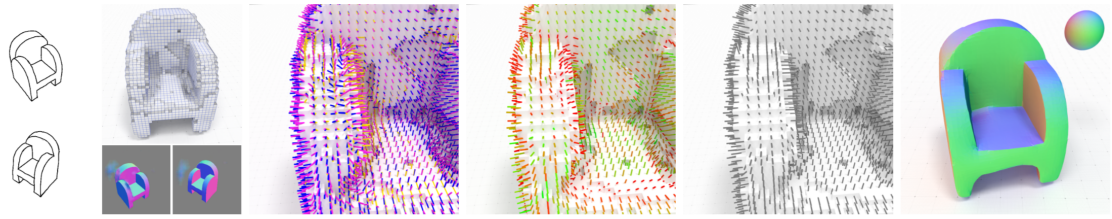


Figure 6. Enact uses a target mobile device and a desktop interface with five areas: a storyboard with consecutive screens, an event timeline with a handle for each screen, a state machine, a code editor and a device mirror.

GRAPHDECO Project-Team

6. New Results

6.1. Computer-Assisted Design with Heterogeneous Representations



(a) Input (b) CNNs predictions (c) Candidate normals (d) Aggregated normals (e) Piecewise-smooth normals (f) Final surface

Figure 4. Our method takes as input multiple sketches of an object (a). We first apply existing deep neural networks to predict a volumetric reconstruction of the shape as well as one normal map per sketch (b). We re-project the normal maps on the voxel grid in complement to the surface normal computed from the volumetric prediction (c). We aggregate these different normals into a distribution represented by a mean vector and a standard deviation (d). We optimize this normal field to make it piecewise smooth (e) and use it to regularize the surface (f). The final surface preserves the overall shape of the predicted voxel grid as well as the sharp features of the predicted normal maps.

6.1.1. Combining Voxel and Normal Predictions for Multi-View 3D Sketching

Participants: Johanna Delanoy, Adrien Bousseau.

Recent works on data-driven sketch-based modeling use either voxel grids or normal/depth maps as geometric representations compatible with convolutional neural networks. While voxel grids can represent complete objects – including parts not visible in the sketches – their memory consumption restricts them to low-resolution predictions. In contrast, a single normal or depth map can capture fine details, but multiple maps from different viewpoints need to be predicted and fused to produce a closed surface. We propose to combine these two representations to address their respective shortcomings in the context of a multi-view sketch-based modeling system. Our method predicts a voxel grid common to all the input sketches, along with one normal map per sketch. We then use the voxel grid as a support for normal map fusion by optimizing its extracted surface such that it is consistent with the re-projected normals, while being as piecewise-smooth as possible overall (Fig. 4). We compare our method with a recent voxel prediction system, demonstrating improved recovery of sharp features over a variety of man-made objects.

This work is a collaboration with David Coeurjolly from Université de Lyon and Jacques-Olivier Lachaud from Université Savoie Mont Blanc. The work was published in the journal Computer & Graphics and presented at the SMI conference [14].

6.1.2. Video Motion Stylization by 2D Rigidification

Participants: Johanna Delanoy, Adrien Bousseau.

Figure 5. Our method takes as input a video and its optical flow (a). We segment the video and optimize its pixel trajectories to produce a new video that exhibits piecewise-rigid motion (b). The resulting rigidified video can be stylized with existing algorithms (c) to produce animations where the style elements (brush strokes, paper texture) produce a strong sense of 2D motion.

We introduce a video stylization method that increases the apparent rigidity of motion. Existing stylization methods often retain the 3D motion of the original video, making the result look like a 3D scene covered in paint rather than a 2D painting of a scene. In contrast, traditional hand-drawn animations often exhibit simplified in-plane motion, such as in the case of cut-out animations where the animator moves pieces of paper from frame to frame. Inspired by this technique, we propose to modify a video such that its content undergoes 2D rigid transforms (Fig. 5). To achieve this goal, our approach applies motion segmentation and optimization to best approximate the input optical flow with piecewise-rigid transforms, and re-renders the video such that its content follows the simplified motion. The output of our method is a new video and its optical flow, which can be fed to any existing video stylization algorithm.

This work is a collaboration with Aaron Hertzmann from Adobe Research. It was presented at the ACM/EG Expressive Symposium [21].

6.1.3. Multi-Pose Interactive Linkage Design

Participant: Adrien Bousseau.

We introduce an interactive tool for novice users to design mechanical objects made of 2.5D linkages. Users simply draw the shape of the object and a few key poses of its multiple moving parts. Our approach automatically generates a one-degree-of-freedom linkage that connects the fixed and moving parts, such that the moving parts traverse all input poses in order without any collision with the fixed and other moving parts. In addition, our approach avoids common linkage defects and favors compact linkages and smooth motion trajectories. Finally, our system automatically generates the 3D geometry of the object and its links, allowing the rapid creation of a physical mockup of the designed object (Fig. 6).

This work was conducted in collaboration with Gen Nishida and Daniel G. Aliaga from Purdue University, was published in Computer Graphics Forum and presented at the Eurographics conference [18].

6.1.4. Extracting Geometric Structures in Images with Delaunay Point Processes

Participant: Adrien Bousseau.

We introduce Delaunay Point Processes, a framework for the extraction of geometric structures from images. Our approach simultaneously locates and groups geometric primitives (line segments, triangles) to form extended structures (line networks, polygons) for a variety of image analysis tasks. Similarly to traditional point processes, our approach uses Markov Chain Monte Carlo to minimize an energy that balances fidelity to the input image data with geometric priors on the output structures. However, while existing point processes struggle to model structures composed of inter-connected components, we propose to embed the point process into a Delaunay triangulation, which provides high-quality connectivity by construction. We further leverage key properties of the Delaunay triangulation to devise a fast Markov Chain Monte Carlo sampler. We

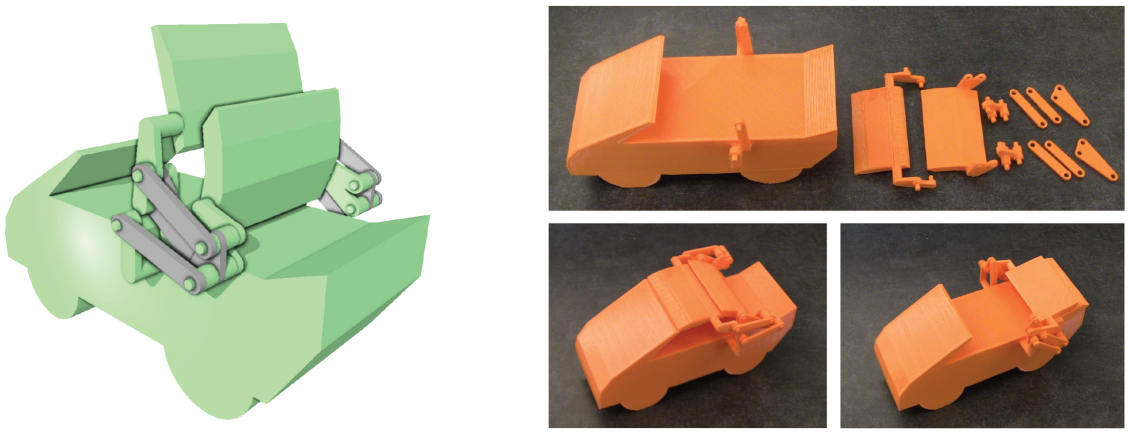


Figure 6. Our interactive system facilitates the creation (left) and fabrication (right) of mechanical objects.

demonstrate the flexibility of our approach on a variety of applications, including line network extraction, object contouring, and mesh-based image compression (see Fig. 7).

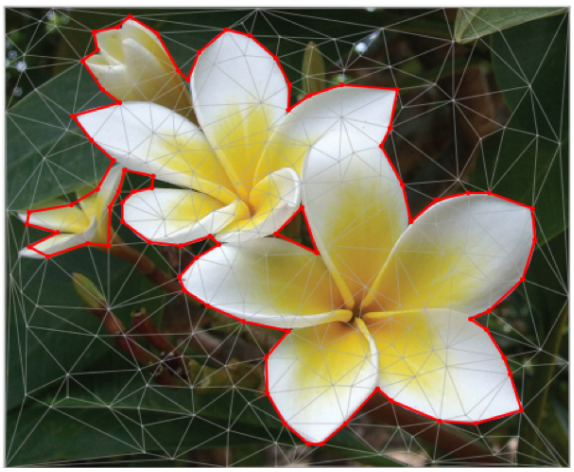


Figure 7. Our method extracts geometric structures like the countour of these flowers by optimizing a dynamic Delaunay triangulation.

This work was conducted in collaboration with Jean-Dominique Favreau and Florent Lafarge (TITANE group), and published in IEEE PAMI [16].

6.1.5. Integer-Grid Sketch Vectorization

Participants: Tibor Stanko, Adrien Bousseau.

A major challenge in line drawing vectorization is segmenting the input bitmap into separate curves. This segmentation is especially problematic for rough sketches, where curves are depicted using multiple overdrawn

strokes. Inspired by feature-aligned mesh quadrangulation methods in geometry processing, we propose to extract vector curve networks by parametrizing the image with local drawing-aligned integer grids. The regular structure of the grid facilitates the extraction of clean line junctions; due to the grid's discrete nature, nearby strokes are implicitly grouped together. Our method successfully vectorizes both clean and rough line drawings, whereas previous methods focused on only one of those drawing types.

This work is an ongoing collaboration with David Bommes from University of Bern and Mikhail Bessmeltsev from University of Montreal. It is currently under review.

6.1.6. Surfacing Sparse Unorganized 3D Curves using Global Parametrization

Participants: Tibor Stanko, Adrien Bousseau.

Designers use sketching to quickly externalize ideas, often using a handful of curves to express complex shapes. Recent years have brought a plethora of new tools for creating designs directly in 3D. The output of these tools is often a set of sparse, unorganized curves. We propose a novel method for automatic conversion of such unorganized curves into clean curve networks ready for surfacing. The core of our method is a global curve-aligned parametrization, which allows us to automatically aggregate information from neighboring curves and produce an output with valid topology.

This work is an ongoing collaboration with David Bommes from University of Bern, Mikhail Bessmeltsev from University of Montreal, and Justin Solomon from MIT.

6.1.7. OpenSketch: A Richly-Annotated Dataset of Product Design Sketches

Participants: Yulia Gryaditskaya, Adrien Bousseau, Fredo Durand.

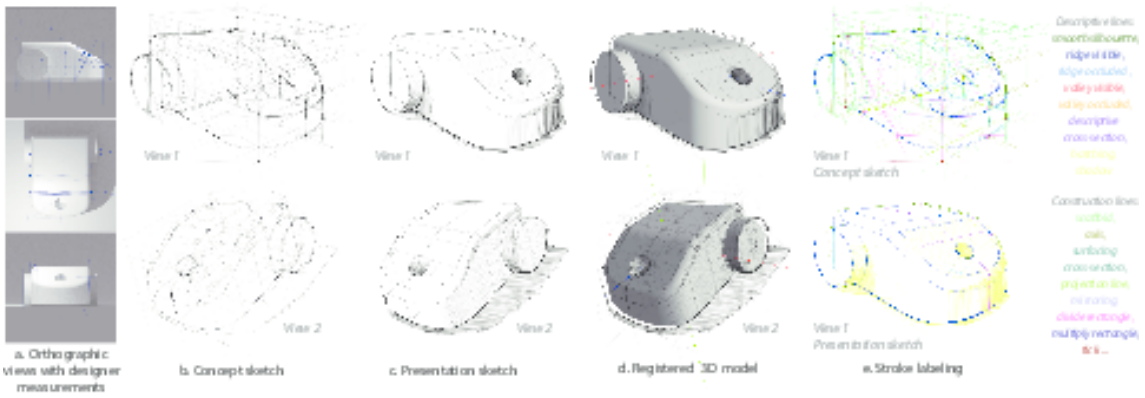


Figure 8. We showed designers three orthographic views (a) of the object and asked them to draw it from two different perspective views (b). We also asked to replicate each of their sketches as a clean presentation drawing (c). We semi-automatically registered 3D models to each sketch (d), and we manually labeled different types of lines in all concept sketches and presentation drawings from the first viewpoint (e).

Product designers extensively use sketches to create and communicate 3D shapes and thus form an ideal audience for sketch-based modeling, non-photorealistic rendering and sketch filtering. However, sketching requires significant expertise and time, making design sketches a scarce resource for the research community. We introduce *OpenSketch*, a dataset of product design sketches aimed at offering a rich source of information for a variety of computer-aided design tasks. *OpenSketch* contains more than 400 sketches representing 12 man-made objects drawn by 7 to 15 product designers of varying expertise. We provided participants with front, side and top views of these objects (Fig. 8 a), and instructed them to draw from two *novel* perspective

viewpoints(Fig. 8 b). This drawing task forces designers to *construct the shape* from their mental vision rather than directly copy what they see. They achieve this task by employing a variety of sketching techniques and methods not observed in prior datasets. Together with industrial design teachers, we distilled a taxonomy of line types and used it to label each stroke of the 214 sketches drawn from one of the two viewpoints (Fig. 8 e). While some of these lines have long been known in computer graphics, others remain to be reproduced algorithmically or exploited for shape inference. In addition, we also asked participants to produce clean presentation drawings from each of their sketches, resulting in aligned pairs of drawings of different styles (Fig. 8 c). Finally, we registered each sketch to its reference 3D model by annotating sparse correspondences (Fig. 8 d). We provide an analysis of our annotated sketches, which reveals systematic drawing strategies over time and shapes, as well as a positive correlation between presence of construction lines and accuracy. Our sketches, in combination with provided annotations, form challenging benchmarks for existing algorithms as well as a great source of inspiration for future developments. We illustrate the versatility of our data by using it to test a 3D reconstruction deep network trained on synthetic drawings, as well as to train a filtering network to convert concept sketches into presentation drawings. We distribute our dataset under the Creative Commons CC0 license: https://ns.inria.fr/d3/OpenSketch.

This work is a collaboration with Mark Sypesteyn, Jan Willem Hoftijzer and Sylvia Pont from TU Delft, Netherlands. This work was published at ACM Transactions on Graphics, and presented at SIGGRAPH Asia 2019 [17].

6.1.8. Intersection vs. Occlusion: a Discrete Formulation of Line Drawing 3D Reconstruction

Participants: Yulia Gryaditskaya, Adrien Bousseau, Felix Hähnlein.

The popularity of sketches in design stems from their ability to communicate complex 3D shapes with a handful of lines. Yet, this economy of means also makes sketch interpretation a challenging task, as global 3D understanding needs to emerge from scattered pen strokes. To tackle this challenge, many prior methods cast 3D reconstruction of line drawings as a global optimization that seeks to satisfy a number of geometric criteria, including orthogonality, planarity, symmetry. However, all of these methods require users to distinguish line intersections that exist in 3D from the ones that are only due to occlusions. These user annotations are critical to the success of existing algorithms, since mistakenly treating an occlusion as a true intersection would connect distant parts of the shape, with dramatic consequences on the overall optimization procedure. We propose a line drawing 3D reconstruction method that automatically discriminates 3D intersections from occlusions. This automation not only reduces user burden, it also allows our method to scale to real-world sketches composed of hundreds of pen strokes, for which the number of intersections is too high to make existing user-assisted methods practical. Our key idea is to associate each 2D intersection with a binary variable that indicates if the intersection should be preserved in 3D. Our algorithm then searches for the assignment of binary values that yields the best 3D shape, as measured with similar criteria as the ones used by prior work for 3D reconstruction. However, the combinatorial nature of this binary assignment problem prevents trying all possible configurations. Our main technical contribution is an efficient search algorithm that leverages principles of how product designers draw to reconstruct complex 3D drawings within minutes.

This work is a collaboration with Alla Sheffer (Professor at University of British Columbia) and Chenxi Liu (PhD student at University of British Columbia).

6.1.9. Data-driven sketch segmentation

Participants: Yulia Gryaditskaya, Felix Hähnlein, Adrien Bousseau.

Deep learning achieves impressive performance on image segmentation, which has motivated the recent development of deep neural networks for the related task of sketch segmentation, where the goal is to assign labels to the different strokes that compose a line drawing. However, while natural images are well represented as bitmaps, line drawings can also be represented as vector graphics, such as point sequences and point clouds. In addition to offering different trade-offs on resolution and storage, vector representations often come with additional information, such as stroke ordering and speed.

In this project, we evaluate three crucial design choices for sketch segmentation using deep-learning: which sketch representation to use, which information to encode in this representation, and which loss function to optimize. Our findings suggest that point clouds represent a competitive alternative to bitmaps for sketch segmentation, and that providing extra-geometric information improves performance.

6.1.10. Stroke-based concept sketch generation

Participants: Felix Hähnlein, Yulia Gryaditskaya, Adrien Bousseau.

State-of-the-art non-photorealisting rendering algorithms can generate lines representing salient visual features on objects. However, very few methods exist for generating lines outside of an object, as is the case for most construction lines, used in technical drawings and design sketches. Furthermore, most methods do not generate human-like strokes and do not consider the drawing order of a sketch.

In this project, we address these issues by proposing a reinforcement learning framework, where a virtual agent tries to generate a construction sketch of a given 3D model. One key element of our approach is the study and the mathematical formalization of drawing strategies used by industrial designers.

6.1.11. Designing Programmable, Self-Actuated Structures

Participants: David Jourdan, Adrien Bousseau.

Self-actuated structures are material assemblies that can deform from an initially simpler state to a more complex, curved one, by automatically deforming to shape. Most relevant to applications in manufacturing are self-actuated shapes that are fabricated flat, considerably reducing the cost and complexity of manufacturing curved 3D surfaces. While there are many ways to design self-actuated materials (e.g. using heat or water as actuation mechanisms), we use 3D printing to embed rigid patterns into prestressed fabric, which is then released and assumes a shape matching a given target when reaching static equilibrium.

While using a 3D printer to embed plastic curves into prestressed fabric is a technique that has been experimented on before, it has been mostly restricted to piecewise minimal surfaces, making it impossible to reproduce most shapes. By using a dense packing of 3-pointed stars, we are able to create convex shapes and positive gaussian curvature, moreover we found a direct link between the stars dimensions and the induced curvature, allowing us to build an inverse design tool that can faithfully reproduce some target shapes.

This is a collaboration with Mélina Skouras of Inria Rhône Alpes and Etienne Vouga of the University of Texas at Austin.

6.2. Graphics with Uncertainty and Heterogeneous Content

6.2.1. Multi-view relighting using a geometry-aware network

Participants: Julien Philip, George Drettakis.

We propose the first learning-based algorithm that can relight images in a plausible and controllable manner given multiple views of an outdoor scene. In particular, we introduce a geometry-aware neural network that utilizes multiple geometry cues (normal maps, specular direction, etc.) and source and target shadow masks computed from a noisy proxy geometry obtained by multi-view stereo. Our model is a three-stage pipeline: two subnetworks refine the source and target shadow masks, and a third performs the final relighting. Furthermore, we introduce a novel representation for the shadow masks, which we call RGB shadow images. They reproject the colors from all views into the shadowed pixels and enable our network to cope with inacuraccies in the proxy and the non-locality of the shadow casting interactions. Acquiring large-scale multi-view relighting datasets for real scenes is challenging, so we train our network on photorealistic synthetic data. At train time, we also compute a noisy stereo-based geometric proxy, this time from the synthetic renderings. This allows us to bridge the gap between the real and synthetic domains. Our model generalizes well to real scenes. It can alter the illumination of drone footage, image-based renderings, textured mesh reconstructions, and even internet photo collections (see Fig. 9).

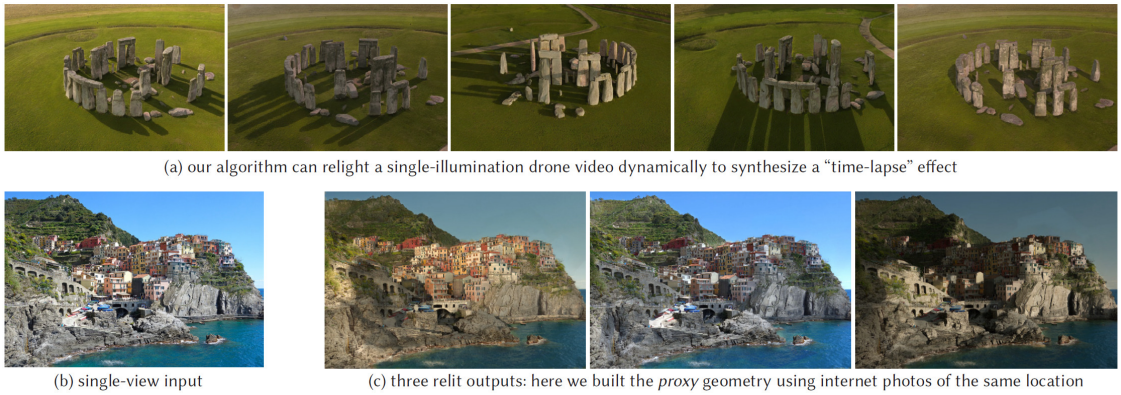


Figure 9. Results of our method: multi-view relighting using a geometry-aware network.

This work was in collaboration with M. Gharbi of Adobe Research and A. Efros and T. Zhang of UC Berkeley, and was published in ACM Transactions on Graphics and presented at SIGGRAPH 2019 [19].

6.2.2. Flexible SVBRDF Capture with a Multi-Image Deep Network

Participants: Valentin Deschaintre, Frédo Durand, George Drettakis, Adrien Bousseau.

Empowered by deep learning, recent methods for material capture can estimate a spatially-varying reflectance from a single photograph. Such lightweight capture is in stark contrast with the tens or hundreds of pictures required by traditional optimization-based approaches. However, a single image is often simply not enough to observe the rich appearance of real-world materials. We present a deep-learning method capable of estimating material appearance from a variable number of uncalibrated and unordered pictures captured with a handheld camera and flash. Thanks to an order-independent fusing layer, this architecture extracts the most useful information from each picture, while benefiting from strong priors learned from data. The method can handle both view and light direction variation without calibration. We show how our method improves its prediction with the number of input pictures, and reaches high quality reconstructions with as little as 1 to 10 images – a sweet spot between existing single-image and complex multi-image approaches – see Fig. 10 .

This work is a collaboration with Miika Aittala from MIT CSAIL. This work was published in Computer Graphics Forum, and presented at EGSR 2019 [15].

A short paper and poster summarizing this work together with our 2018 "Single-Image SVBRDF Capture with a Rendering-Aware Deep Network" was published in the Siggraph Asia doctoral consortium 2019 [22].

6.2.3. Guided Acquisition of SVBRDFs

Participants: Valentin Deschaintre, George Drettakis, Adrien Bousseau.

Another project is under development to capture a large-scale SVBRDF from a few pictures of a planar surface. Many existing lightweight methods for SVBRDF capture take as input flash pictures, which need to be acquired close to the surface of interest restricting the scale of capture. We complement such small-scale inputs with a picture of the entire surface, taken under ambient lighting. Our method then fuses these two sources of information to propagate the SVBRDFs estimated from each close-up flash picture to all pixels of the large image. Thanks to our two-scale approach, we can capture surfaces several meters wide, such as walls, doors and furniture. In addition, our method can also be used to create large SVBRDFs from internet pictures, where we use artist-designed SVBRDFs as exemplars of the small-scale behavior of the surface.

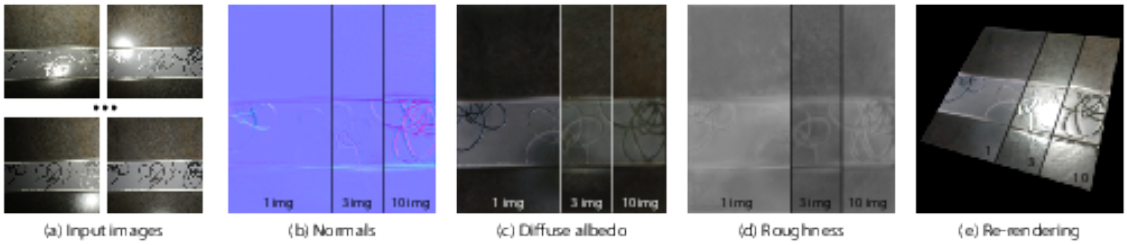


Figure 10. Our deep learning method for SVBRDF capture supports a variable number of input photographs taken with uncalibrated light-view directions (a, rectified). While a single image is enough to obtain a first plausible estimate of the SVBRDF maps, more images provide new cues to our method, improving its prediction. In this example, adding images reveals fine normal variations (b), removes highlight residuals in the diffuse albedo (c), and reveals the difference of roughness between the stone, the stripe, and the thin pattern (d).

6.2.4. Mixed rendering and relighting for indoor scenes

Participants: Julien Philip, Michaël Gharbi, George Drettakis.

We are investigating a mixed image rendering and relighting method that allows a user to move freely in a multi-view interior scene while altering its lighting. Our method uses a deep convolutional network trained on synthetic photo-realistic images. We adapt classical path tracing techniques to approximate complex lighting effects such as color bleeding and reflections.

6.2.5. DiCE: Dichoptic Contrast Enhancement for VR and Stereo Displays

Participant: George Drettakis.

In stereoscopic displays, such as those used in VR/AR headsets, our eyes are presented with two different views. The disparity between the views is typically used to convey depth cues, but it could be also used to enhance image appearance. We devise a novel technique that takes advantage of binocular fusion to boost perceived local contrast and visual quality of images. Since the technique is based on fixed tone curves, it has negligible computational cost and it is well suited for real-time applications, such as VR rendering. To control the trade-of between contrast gain and binocular rivalry, we conducted a series of experiments to explain the factors that dominate rivalry perception in a dichoptic presentation where two images of different contrasts are displayed (see Fig. 11). With this new finding, we can effectively enhance contrast and control rivalry in mono- and stereoscopic images, and in VR rendering, as conirmed in validation experiments.

This work was in collaboration with Durham University (G. Koulieris, past postdoc of the group), Cambridge (F. Zhong, R. Mantiuk), UC Berkeley (M. Banks) and ENS Renne (M. Chambe), and was published in ACM Transactions on Graphics and presented at SIGGRAPH Asia 2019 [20].

6.2.6. Compositing Real Scenes using a relighting Network

Participants: Baptiste Nicolet, Julien Philip, George Drettakis.

Image-Based Rendering (IBR) allows for fast rendering of photorealistic novel viewpoints of real-world scenes captured by photographs. While it facilitates the very tedious traditional content creation process, it lacks user control over the appearance of the scene. We propose a novel approach to create novel scenes from a composition of multiple IBR scenes. This method relies on the use of a relighting network, which we first use to match the lighting conditions of each scene, and then to synthesize shadows between scenes in the final composition. This work has been submitted for publication.

6.2.7. Image-based Rendering of Urban Scenes based on Semantic Information

Participants: Simon Rodriguez, Siddhant Prakash, George Drettakis.

Figure 11. Comparison of standard stereo images and the images with enhanced perceived contrast using our DiCE method. They can be cross-fused with the assistance of the dots above the images. Notice the enhanced contrast in the shadows and highlights of the scene. The stereo images are from Big Buck Bunny by Blender Foundation.

Cityscapes exhibit many hard cases for image-based rendering techniques, such as reflective and transparent surfaces. Pre-existing information about the scene can be leveraged to tackle these difficult cases. By relying on semantic information, it is possible to address those regions with tailored algorithms to improve reconstruction and rendering. This project is a collaboration with Peter Hedman from University College of London. This work has been submitted for publication.

6.2.8. Synthetic Data for Image-based Rendering

Participants: Simon Rodriguez, Thomas Leimkühler, George Drettakis.

This project explores the potential of Image-based rendering techniques in the context of real-time rendering for synthetic scenes. Accurate information can be precomputed from the input synthetic scene and used at run-time to improve the quality of approximate global illumination effects while preserving performance. This project is a collaboration with Chris Wyman and Peter Shirley from NVIDIA Research.

6.2.9. Densified Surface Light Fields for Human Capture Video

Participants: Rada Deeb, George Drettakis.

In this project, we focus on video-based rendering for mid-scale platforms. Having a mid-scale platform introduces one important problem for image-based rendering techniques due to low angular resolution. This leads to unrealistic view-dependent effects. We propose to use the temporal domain in a multidimensional surface light field approach in order to enhance the angular resolution. In addition, our approach provides a compact representation essential to dealing with the large amount of data introduced by videos compared to image-based techniques. In addition, we evaluate the use of deep encoder-decoder networks to learn a more compact representation of our multidimensional surface light field. This work is in collaboration with Edmond Boyer, MORPHEO team, Inria Grenoble.

6.2.10. Deep Bayesian Image-based Rendering

Participants: Thomas Leimkühler, George Drettakis.

Deep learning has permeated the field of computer graphics and continues to be instrumental in producing state-of-the-art research results. In the context of image-based rendering, deep architectures are now routinely used for tasks such as blending weight prediction, view extrapolation, or re-lighting. Current algorithms, however, do not take into account the different sources of uncertainty arising from the several stages of the image-based rendering pipeline. In this project, we investigate the use of Bayesian deep learning models to estimate and exploit these uncertainties. We are interested in devising principled methods which combine the expressive power of modern deep learning with the well-groundedness of classical Bayesian models.

6.2.11. Path Guiding for Metropolis Light Transport

Participants: Stavros Diolatzis, George Drettakis.

Path guiding has been proven to be an effective way to achieve faster convergence in Monte Carlo renderings by learning the incident radiance field. However, current path guiding techniques could be beaten by unguided path tracing due to their overhead or inability to incorporate the BSDF distribution factor. In our work, we improve path guiding and Metropolis light transport algorithms with low overhead product sampling between the incoming radiance and BSDF values. We demonstrate that our method has better convergence compared to the previous state-of-the-art techniques. Moreover, combining path guiding with MLT solves the global exploration issues ensuring convergence to the stationary distribution.

This work is an ongoing collaboration with Wenzel Jakob from Ecole Polytechnique Fédérale de Lausanne and Adrien Gruson from McGill University.

6.2.12. Improved Image-Based Rendering with Uncontrolled Capture

Participants: Siddhant Prakash, George Drettakis.

Current state-of-the-art Image Based Rendering (IBR) algorithms, such as Deep Blending, use per-view geometry to render candidate views and machine learning to improve rendering of novel views. The casual capture process employed introduces in visible color artifacts during rendering due to automated camera settings, and incur significant computational overhead when using per-view meshes. We aim to find a global solution to harmonize color inconsistency across the entire set of images in a given dataset, and also improve the performance of IBR algorithms by limiting the use of more advanced techniques only to regions where they are required.

6.2.13. Practical video-based rendering of dynamic stationary environments

Participants: Théo Thonat, George Drettakis.

The goal of this work is to extend traditional Image Based Rendering to capture subtle motions in real scenes. We want to allow free-viewpoint navigation with casual capture, such as a user taking photos and videos with a single smartphone and a tripod. We focus on stochastic time-dependent textures such as waves, flames or waterfalls. We have developed a video representation able to tackle the challenge of blending unsynchronized videos.

This work is a collaboration with Sylvain Paris from Adobe Research, Miika Aittala from MIT CSAIL, and Yagiz Aksoy from ETH Zurich, and has been submitted for publication.

HYBRID Project-Team

7. New Results

7.1. Virtual Reality Tools and Usages

7.1.1. Studying the Mental Effort in Virtual Versus Real Environments

Participants: Tiffany Luong, Ferran Argelaguet, Anatole Lécuyer [contact].

Is there an effect of Virtual Reality (VR) Head-Mounted Display (HMD) on the user's mental effort? In this work, we compare the mental effort in VR versus in real environments [26]. An experiment (N=27) was conducted to assess the effect of being immersed in a virtual environment (VE) using a HMD on the user's mental effort while performing a standardized cognitive task (the wellknown N-back task, with three levels of difficulty (1,2,3)). In addition to test the effect of the environment (i.e., virtual versus real), we also explored the impact of performing a dual task (i.e., sitting versus walking) in both environments on mental effort. The mental effort was assessed through self-reports, task performance, behavioural and physiological measures. In a nutshell, the analysis of all measurements revealed no significant effect of being immersed in the VE on the users' mental effort. In contrast, natural walking significantly increased the users' mental effort. Taken together, our results support the fact that there is no specific additional mental effort related to the immersion in a VE using a VR HMD.

7.1.2. Influence of Personality Traits and Body Awareness on the Sense of Embodiment in VR

Participants: Diane Dewez, Rebecca Fribourg, Ferran Argelaguet, Anatole Lécuyer [contact].

With the increasing use of avatars in virtual reality, it is important to identify the factors eliciting the sense of embodiment. This work reports an exploratory study aiming at identifying internal factors (personality traits and body awareness) that might cause either a resistance or a predisposition to feel a sense of embodiment towards a virtual avatar. To this purpose, we conducted an experiment (n=123) in which participants were immersed in a virtual environment and embodied in a gender-matched generic virtual avatar through a head-mounted display [16]. After an exposure phase in which they had to perform a number of visuomotor tasks, a virtual character entered the virtual scene and stabbed the participants' virtual hand with a knife (see Figure 4). The participants' sense of embodiment was measured, as well as several personality traits (Big Five traits and locus of control) and body awareness, to evaluate the influence of participants' personality on the acceptance of the virtual body. The major finding is that the locus of control is linked to several components of embodiment: the sense of agency is positively correlated with an internal locus of control and the sense of body ownership is positively correlated with an external locus of control. Taken together, our results suggest that the locus of control could be a good predictor of the sense of embodiment. Yet, further studies are required to confirm these results.
This work was done in collaboration with the MimeTIC team.

7.1.3. Consumer perceptions and purchase behavior of imperfect fruits and vegetables in VR

Participants: Jean-Marie Normand, Guillaume Moreau [contact].

This study investigates the effects of fruits and vegetables (FaVs) abnormality on consumer perceptions and purchasing behavior [9]. For the purposes of this study, a virtual grocery store was created with a fresh FaVs section, where 142 participants became immersed using an Oculus Rift DK2 Head-Mounted Display (HMD) software. Participants were presented either normal, slightly misshapen, moderately misshapen or severely misshapen FaVs. The study findings indicate that shoppers tend to purchase a similar number of FaVs whatever their level of deformity. However, perceptions of the appearance and quality of the FaVs depend on the degree of abnormality. Moderately misshapen FaVs are perceived as significantly better than those that are heavily misshapen but also "slightly" misshapen (except for the appearance of fruits).

Figure 4. From left to right: an example of a trajectory to draw during the experimental task; A view of the scene from behind; Another virtual character stabbing the participants' virtual hand at the end of the experiment to measure their response to the threat on their virtual body.

This work was done in collaboration with Audecia Recherce, the University of Reading and the University of Tokyo.

7.1.4. Am I better in VR with a real audience?

Participants: Romain Terrier, Valérie Gouranton [contact], Bruno Arnaldi.

We designed an experimental study to investigate the effects of a real audience on social inhibition [33]. The study is a virtual reality (VR) and multiuser application (see Figure 5). The experience is locally or remotely shared. The application engages one user and a real audience (i.e., local or remote conditions). A control condition is designed where the user is alone (i.e., alone condition). The objective performance (i.e., type and answering time) of users, when performing a categorization of numbers task in VR, is used to explore differences between conditions. In addition to this, the perceptions of others, the stress, the cognitive workload, and the presence of each user have been compared in relation to the location of the real audience. The results showed that in the presence of a real audience (in the local and remote conditions), user performance is affected by social inhibitions. Furthermore, users are even more influenced when the audience does not share the same room, despite others are less perceived.
This work was done in collaboration with IRT B COM.

Figure 5. Experimental setup for the social inhibition experiment in Virtual Reality.

7.1.5. Create by Doing – Action sequencing in VR

Participants: Flavien Lécuyer, Valérie Gouranton [contact], Adrien Reuzeau, Ronan Gaugne, Bruno Arnaldi.

In every virtual reality application, there are actions to perform, often in a logical order. This logical ordering can be a predefined sequence of actions, enriched with the representation of different possibilities, which we refer to as a scenario. Authoring such a scenario for virtual reality is still a difficult task, as it needs both the expertise from the domain expert and the developer. We propose [28] to let the domain expert create in virtual reality the scenario by herself without coding, through the paradigm of creating by doing (see Figure 6). The domain expert can run an application, record the sequence of actions as a scenario, and then reuse this scenario for other purposes, such as an automatic replay of the scenario by a virtual actor to check the obtained scenario, the injection of this scenario as a constraint or a guide for a trainee, or the monitoring of the scenario unfolding during a procedure.

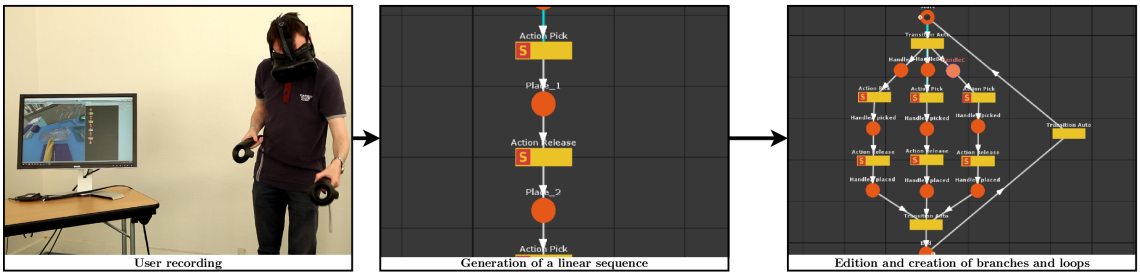


Figure 6. The proposed workflow for the creation of scenarios

7.1.6. Help! I Need a Remote Guide in my Mixed Reality Collaborative Environment

Participants: Valérie Gouranton [contact], Bruno Arnaldi.

The help of a remote expert in performing a maintenance task can be useful in many situations, and can save time as well as money. In this context, augmented reality (AR) technologies can improve remote guidance thanks to the direct overlay of 3D information onto the real world. Furthermore, virtual reality (VR) enables a remote expert to virtually share the place in which the physical maintenance is being carried out. In a traditional local collaboration, collaborators are face-to-face and are observing the same artifact, while being able to communicate verbally and use body language such as gaze direction or facial expression. These interpersonal communication cues are usually limited in remote collaborative maintenance scenarios, in which the agent uses an AR setup while the remote expert uses VR. Providing users with adapted interaction and awareness features to compensate for the lack of essential communication signals is therefore a real challenge for remote MR collaboration. However, this context offers new opportunities for augmenting collaborative abilities, such as sharing an identical point of view, which is not possible in real life. Based on the current task of the maintenance procedure, such as navigation to the correct location or physical manipulation, the remote expert may choose to freely control his/her own viewpoint of the distant workspace, or instead may need to share the viewpoint of the agent in order to better understand the current situation. In this work, we first focus on the navigation task, which is essential to complete the diagnostic phase and to begin the maintenance task in the correct location [8]. We then present a novel interaction paradigm, implemented in an early prototype, in which the guide can show the operator the manipulation gestures required to achieve a physical task that is necessary to perform the maintenance procedure. These concepts are evaluated, allowing us to provide guidelines for future systems targeting efficient remote collaboration in MR environments.
This work was done in collaboration with IRT B COM and UMR Lab-STICC, France.

7.1.7. Learning procedural skills with a VR simulator: An acceptability study

Participants: Valérie Gouranton [contact], Bruno Arnaldi.

Virtual Reality (VR) simulation has recently been developed and has improved surgical training. Most VR simulators focus on learning technical skills and few on procedural skills. Studies that evaluated VR simulators focused on feasibility, reliability or easiness of use, but few of them used a specific acceptability measurement tool. The aim of the study was to assess acceptability and usability of a new VR simulator for procedural skill training among scrub nurses, based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model. The simulator training system was tested with a convenience sample of 16 non-expert users and 13 expert scrub nurses from the neurosurgery department of a French University Hospital. The scenario was designed to train scrub nurses in the preparation of the instrumentation table for a craniotomy in the operating room (OR). Acceptability of the VR simulator was demonstrated with no significant difference between expert scrub nurses and non-experts. There was no effect of age, gender or expertise. Workload, immersion and simulator sickness were also rated equally by all participants. Most participants stressed its pedagogical interest, fun and realism, but some of them also regretted its lack of visual comfort. This VR simulator designed to teach surgical procedures can be widely used as a tool in initial or vocational training [2], [43].
This work was achieved in collaboration with Univ. Rennes 2-LP3C, LTSI and the Hycomes team.

7.1.8. The Anisotropy of Distance Perception in VR

Participants: Etienne Peillard, Anatole Lécuyer, Ferran Argelaguet, Jean-Marie Normand, Guillaume Moreau [contact].

The topic of distance perception has been widely investigated in Virtual Reality (VR). However, the vast majority of previous work mainly focused on distance perception of objects placed in front of the observer. Then, what happens when the observer looks on the side? In this work, we study differences in distance estimation when comparing objects placed in front of the observer with objects placed on his side [31]. Through a series of four experiments (n=85), we assessed participants' distance estimation and ruled out potential biases. In particular, we considered the placement of visual stimuli in the field of view, users' exploration behavior as well as the presence of depth cues. For all experiments a two-alternative forced choice (2AFC) standardized psychophysical protocol was employed, in which the main task was to determine the stimuli that seemed to be the farthest one. In summary, our results showed that the orientation of virtual stimuli with respect to the user introduces a distance perception bias: objects placed on the sides are systematically perceived farther away than objects in front. In addition, we could observe that this bias increases along with the angle, and appears to be independent of both the position of the object in the field of view as well as the quality of the virtual scene. This work sheds a new light on one of the specificities of VR environments regarding the wider subject of visual space theory. Our study paves the way for future experiments evaluating the anisotropy of distance perception in real and virtual environments.

7.1.9. Study of Gaze and Body Segments Temporal Reorientation Behaviour in VR

Participants: Hugo Brument, Ferran Argelaguet [contact].

This work investigates whether the body anticipation synergies in real environments (REs) are preserved during navigation in virtual environments (VEs). Experimental studies related to the control of human locomotion in REs during curved trajectories report a top-down body segments reorientation strategy, with the reorientation of the gaze anticipating the reorientation of head, the shoulders and finally the global body motion [12]. This anticipation behavior provides a stable reference frame to the walker to control and reorient his/her body segments according to the future walking direction. To assess body anticipation during navigation in VEs, we conducted an experiment where participants, wearing a head-mounted display, performed a lemniscate trajectory in a virtual environment (VE) using five different navigation techniques, including walking, virtual steering (head, hand or torso steering) and passive navigation. For the purpose of this experiment, we designed a new control law based on the power-law relation between speed and curvature during human walking. Taken together, our results showed a similar ordered top-down sequence of reorientation of the gaze, head and shoulders during curved trajectories for all the evaluated techniques. However, the anticipation mechanism was significantly higher for the walking condition compared to the

others. Finally, the results work pave the way to the better understanding of the underlying mechanisms of human navigation in VEs and to the design of navigation techniques more adapted to humans.

This work was done in collaboration with the MimeTIC team and the Interactive Media Systems Group (TU Wien, Vienna, Austria).

7.1.10. User-centered design of a multisensory power wheelchair simulator

Participants: Guillaume Vailland, Valérie Gouranton [contact].

Autonomy and social inclusion can reveal themselves everyday challenges for people experiencing mobility impairments. These people can benefit from technical aids such as power wheelchairs to access mobility and overcome social exclusion. However, power wheelchair driving is a challenging task which requires good visual, cognitive and visuo-spatial abilities. Besides, a power wheelchair can cause material damage or represent a danger of injury for others or oneself if not operated safely. Therefore, training and repeated practice are mandatory to acquire safe driving skills to obtain power wheelchair prescription from therapists. However, conventional training programs may reveal themselves insufficient for some people with severe impairments. In this context, Virtual Reality offers the opportunity to design innovative learning and training programs while providing realistic wheelchair driving experience within a virtual environment. In line with this, we propose a user-centered design of a multisensory power wheelchair simulator [34]. This simulator addresses classical virtual experience drawbacks such as cybersickness and sense of presence by combining 3D visual rendering, haptic feedback and motion cues. The simulator was showcased in the SOFMER conference [37].
This work has been done in collaboration with Rainbow team.

Figure 7. Wheelchair simulator.

7.1.11. Machine Learning Based Interaction Technique Selection For 3D User Interfaces

Participant: Bruno Arnaldi [contact].

A 3D user interface can be adapted in multiple ways according to each user's needs, skills and preferences. Such adaptation can consist in changing the user interface layout or its interaction techniques. Personalization systems which are based on user models can automatically determine the configuration of a 3D user interface in order to fit a particular user. In this work, we proposed to explore the use of machine learning in order to propose a 3D selection interaction technique adapted to a target user [23]. To do so, we built a dataset with 51 users on a simple selection application in which we recorded each user profile, his/her results to a

2D Fitts Law based pre-test and his/her preferences and performances on this application for three different interaction techniques. Our machine learning algorithm based on Support Vector Machines (SVMs) trained on this dataset proposes the most adapted interaction technique according to the user profile or his/her result to the 2D selection pre-test. Our results suggest the interest of our approach for personalizing a 3D user interface according to the target user but it would require a larger dataset in order to increase the confidence about the proposed adaptations.

7.1.12. The 3DUI Contest 2019

Participants: Hugo Brument, Rebecca Fribourg, Gerard Gallagher, Thomas Howard, Flavien Lécuyer, Tiffany Luong, Victor Mercado, Etienne Peillard, Xavier de Tinguy, Maud Marchal [contact].

Pyramid Escape: Design of Novel Passive Haptics Interactions for an Immersive and Modular Scenario
In this work, we present the design of ten different 3D user interactions using passive haptics and embedded in an escape game scenario in which users have to escape from a pyramid in a limited time [11]. Our solution is innovative by its modularity, allowing interactions with virtual objects using tangible props manipulated either directly using the hands and feet or indirectly through a single prop held in the hand, in order to perform several interactions with the virtual environment (VE). We also propose a navigation technique based on the "impossible spaces" design, allowing users to naturally walk through several overlapping rooms of the VE. All together, our different interaction techniques allow the users to solve several enigmas built into a challenging scenario inside a pyramid.

7.2. Augmented Reality Tools and Usages

7.2.1. Authoring AR by AR, abstraction and libraries

Participants: Flavien Lécuyer, Valérie Gouranton [contact], Adrien Reuzeau, Ronan Gaugne, Bruno Arnaldi.

The demand for augmented reality applications is rapidly growing. In many domains, we observe a new interest for this technology, stressing the need for more efficient ways of producing augmented content. Similarly to virtual reality, interactive objects in augmented reality are a powerful means to improve the experience. While it is now well democratized for virtual reality, interactivity is still finding its way into augmented reality. To open the way to this interactive augmented reality, we designed a new methodology for the management of the interactions in augmented reality, supported by an authoring tool for the use by designers and domain experts [27]. This tool makes the production of interactive augmented content faster, while being scalable to the needs of each application. Usually in the creation of applications, a large amount of time is spent through discussions between the designer (or the domain expert), carrying the needs of the application, and the developer, holding the knowledge to create it (see Figure 8). Thanks to our tool, we reduce this time by allowing the designer to create an interactive application, without having to write a single line of code.

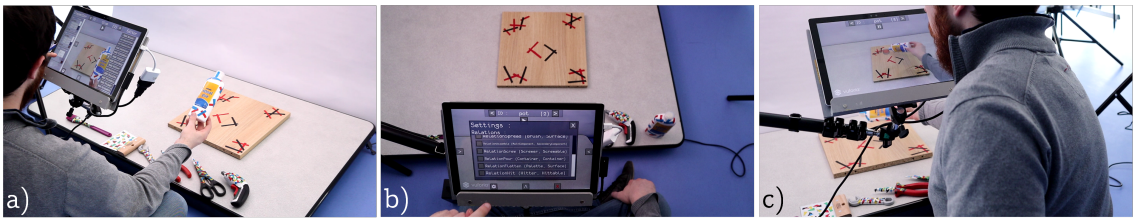


Figure 8. From left to right, the user (a) adds an interactive behaviour on a bottle of glue, (b) imports the interactions in the environment, and (c) uses the interaction to pour the virtual glue from the real bottle into a virtual pot

7.2.2. Studying Exocentric Distance Perception in Optical See-Through AR

Participants: Etienne Peillard, Ferran Argelaguet, Jean-Marie Normand, Anatole Lécuyer, Guillaume Moreau [contact].

While perceptual biases have been widely investigated in Virtual Reality (VR), very few studies have considered the challenging environment of Optical See-through Augmented Reality (OST-AR). Moreover, regarding distance perception, existing works mainly focus on the assessment of egocentric distance perception, i.e. distance between the observer and a real or a virtual object. In this work, we studied exocentric distance perception in AR, hereby considered as the distance between two objects, none of them being directly linked to the user. We report a user study (n=29) aiming at estimating distances between two objects lying in a frontoparallel plane at 2.1m from the observer (i.e. in the medium-field perceptual space). Four conditions were tested in our study: real objects on the left and on the right of the participant (called real-real), virtual objects on both sides (virtual-virtual), a real object on the left and a virtual one on the right (real-virtual) and finally a virtual object on the left and a real object on the right (virtual-real). Participants had to reproduce the distance between the objects by spreading two real identical objects presented in front of them (see Figure 9). The main findings of this study are the overestimation (20%) of exocentric distances for all tested conditions. Surprisingly, the real-real condition was significantly more overestimated (by about 4%,p=.0166) compared to the virtual-virtual condition, i.e. participants obtained better estimates of the exocentric distance for the virtual-virtual condition. Finally, for the virtual-real/real-virtual conditions, the analysis showed a non-symmetrical behavior, which suggests that the relationship between real and virtual objects with respect to the user might be affected by other external factors. Considered together, these unexpected results illustrate the need for additional experiments to better understand the perceptual phenomena involved in exocentric distance perception with real and virtual objects [30].

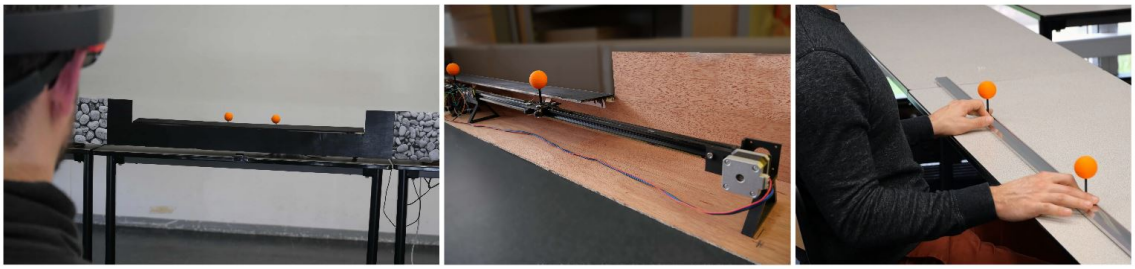


Figure 9. Left, bench displaying two real spheres. The hinge-actuated moving panel, opened here, could be automatically opened/closed to reveal/hide the visual stimuli. Center, one of the two rails of the bench, seen from behind. An orange sphere is attached on top of a trolley that can slide on the rail. The trolley is moved by a stepper motor through a belt. The other half of the bench is symmetrical. Right, participants could provide the perceived exocentric distance by placing two sliding spheres. After the participants placed the spheres the system automatically took a picture of both spheres which was used to measure the distance between both spheres.

7.2.3. Influence of virtual objects' shadows and lighting coherence in AR

Participants: Etienne Peillard, Jean-Marie Normand, Guillaume Moreau [contact].

This work focuses on how virtual objects' shadows as well as differences in alignment between virtual and real lighting influence distance perception in optical see-through (OST) augmented reality (AR) [5]. Four hypotheses are pro-posed: (H1) Participants underestimate distances in OST AR; (H2) Virtual objects' shadows improve distance judgment accuracy in OST AR; (H3) Shadows with different realism levels have different influence on distance perception in OST AR; (H4) Different levels of lighting misalignment between

real and virtual lights have different influence on distance perception in OST AR scenes. Two experiments were designed with an OST head mounted display(HMD), the Microsoft HoloLens. Participants had to match the position of a virtual object displayed in the OST-HMD with a real target. Distance judgment accuracy was recorded under the different shadows and lighting conditions.The results validate hypotheses H2 and H4 but surprisingly showed no impact of the shape of virtual shadows on distance judgment accuracy thus rejecting hypothesis H3. Regarding hypothesis H1, we detected a trend toward underestimation; given the high variance of the data, more experiments are needed to confirm this result. Moreover, the study also reveals that perceived distance errors and completion time of trials increase along with targets' distance.

7.2.4. A study on differences in human perception in AR

Participants: Jean-Marie Normand, Guillaume Moreau [contact].

With the recent growth in the development of augmented reality (AR) technologies, it is becoming important to study human perception of AR scenes. In order to detect whether users will suffer more from visual and operator fatigue when watching virtual objects through optical see-through head-mounted displays (OST-HMDs), compared with watching real objects in the real world, we propose a comparative experiment including a virtual magic cube task and a real magic cube task [4]. The scores of the subjective questionnaires (SQ) and the values of the critical flicker frequency (CFF) were obtained from 18 participants. In our study, we use several electrooculogram (EOG) and heart rate variability (HRV) measures as objective indicators of visual and operator fatigue. Statistical analyses were performed to deal with the subjective and objective indicators in the two tasks. Our results suggest that participants were very likely to suffer more from visual and operator fatigue when watching virtual objects presented by the OST-HMD. In addition, the present study provides hints that HRV and EOG measures could be used to explore how visual and operator fatigue are induced by AR content. Finally, three novel HRV measures are proposed to be used as potential indicators of operator fatigue.
This work was done in collaboration with the Beijing Engineering Research Center of Mixed Reality and Advanced Display (School of Optics and Photonics, Beijing Institute of Technology, Beijing, China) and AICFVE (Beijing Film Academy, Beijing, China).

7.3. Physically-Based Simulation and Haptic Feedback

7.3.1. Design of haptic guides for pre-positioning assistance of a comanipulated needle

Participant: Maud Marchal [contact].

In minimally-invasive procedures like biopsy, the physician has to insert a needle into the tissues of a patient to reach a target. Currently, this task is mostly performed manually and under visual guidance. However, manual needle insertion can result in a large final positioning error of the tip that might lead to misdiagnosis and inadequate treatment. A way to solve this limitation is to use shared control; a gesture assistance paradigm that combines the cognitive skills of the operator with the precision, stamina and repeatability of a robotic or haptic device. In this paper, we propose to assist the physician with a haptic device that holds the needle and generates mechanical guides during the phase of manual needle pre-positioning. In the latter, the physician has to place the tip of the needle on a planned entry point, with a pre-defined angle of incidence. From this pre-operative information and also from intra-operative measurements, we propose to generate haptic cues, known as virtual fixtures, to guide the physician towards the desired position and orientation of the needle. It takes the form of five haptic guides, each one implementing virtual fixtures. We conducted a user study where those guides were compared to the unassisted reference gesture. The most constraining guide, in terms of assisted degrees of freedom, was highlighted as the one that provides the best results in terms of performance and user experience [20], [21].
This work was done in collaboration with the Inria Rainbow team.

7.3.2. An Interactive Physically-based Model for Active Suction Phenomenon Simulation

Participants: Antonin Bernardin, Maud Marchal [contact].

While suction cups are widely used in Robotics, the literature is underdeveloped when it comes to the modelling and simulation of the suction phenomenon (see Figure 10). In this work, we present a novel physically-based approach to simulate the behavior of active suction cups. Our model relies on a novel formulation which assumes the pressure exerted on a suction cup during active control is based on constraint resolution. Our algorithmic implementation uses a classification process to handle the contacts during the suction phenomenon of the suction cup on a surface. Then, we formulate a convenient way for coupling the pressure constraint with the multiple contact constraints. We propose an evaluation of our approach through a comparison with real data, showing the ability of our model to reproduce the behavior of suction cups. Our approach paves the way for improving the design as well as the control of robotic actuators based on suction cups such as vacuum grippers.

This work was done in collaboration with the Inria Defrost team.

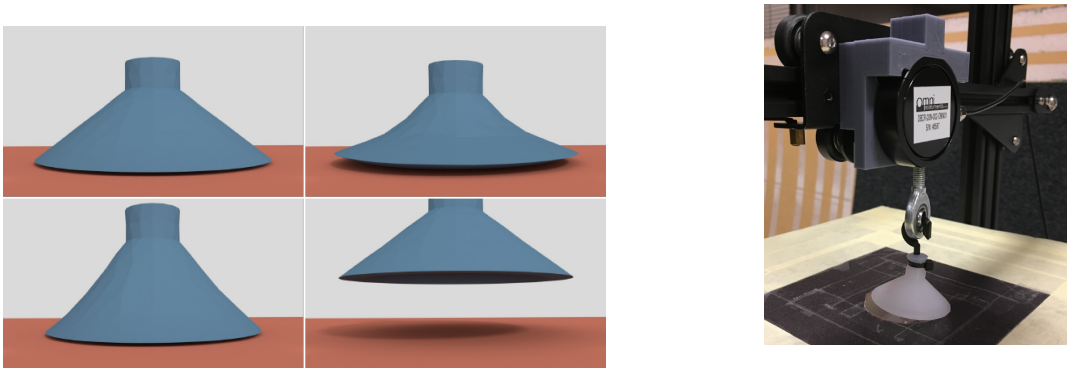


Figure 10. Left, Illustration of our constraint-based physically-based approach for simulating active suction cup phenomenon: (Top) the suction cup is actively stuck to the surface, (Bottom) is is then release until being completely in the air. Right, experimental setup for the force measurements. The suction cup is attached to a force sensor. When it is positioned on a flat surface, its cavity is linked to a vacuum pump with a regulator inbetween.

7.3.3. How different tangible and virtual objects can be while still feeling the same?

Participants: Xavier de Tinguy, Anatole Lécuyer, Maud Marchal [contact].

Tangible objects are used in Virtual Reality to provide human users with distributed haptic sensations when grasping virtual objects. To achieve a compelling illusion, there should be a good correspondence between the haptic features of the tangible object and those of the corresponding virtual one, i.e., what users see in the virtual environment should match as much as possible what they touch in the real world. This work [14] aims at quantifying how similar tangible and virtual objects need to be, in terms of haptic perception, to still feel the same. As it is often not possible to create tangible replicas of all the virtual objects in the scene, it is important to understand how different tangible and virtual objects can be without the user noticing (see Figure 11). This paper reports on the just-noticeable difference (JND) when grasping, with a thumb-index pinch, a tangible object which differ from a seen virtual one on three important haptic features: width, local orientation, and curvature. Results show JND values of 5.75%, 43.8%, and 66.66% of the reference shape for the width, local orientation, and local curvature features, respectively. These results will enable researchers in the field of Virtual Reality to use a reduced number of tangible objects to render multiple virtual ones.
This work was done in collaboration with the Inria Rainbow team.

7.3.4. Toward Universal Tangible Objects

Participants: Xavier de Tinguy, Maud Marchal, Anatole Lécuyer [contact].

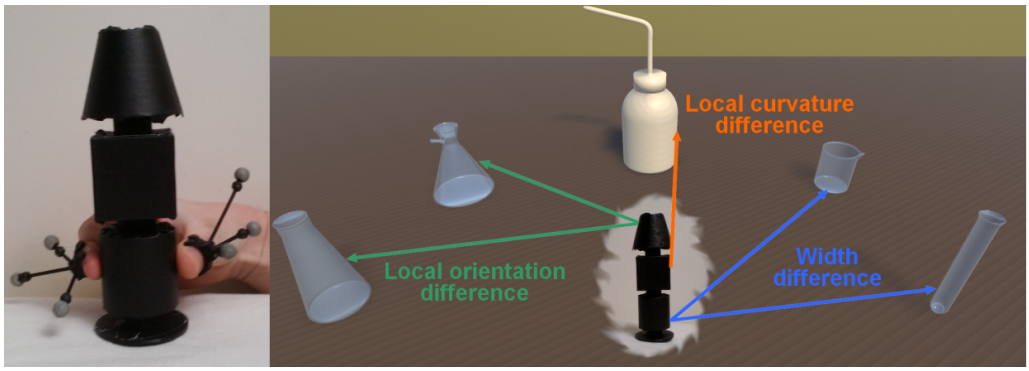


Figure 11. Understanding how different a tangible object (left) can be from virtual objects (right) without the user noticing the mismatch. We focused our study on three specific criteria: width, local orientation, and curvature.

Tangible objects are a simple yet effective way for providing haptic sensations in Virtual Reality. For achieving a compelling illusion, there should be a good correspondence between what users see in the virtual environment and what they touch in the real world. The haptic features of the tangible object should indeed match those of the corresponding virtual one in terms of, e.g., size, local shape, mass, texture. A straightforward solution is to create perfect tangible replicas of all the virtual objects in the scene. However, this is often neither feasible nor desirable. This work [15] presents an innovative approach enabling the use of few tangible objects to render many virtual ones (see Figure 12). The proposed algorithm analyzes the available tangible and virtual objects to find the best grasps in terms of matching haptic sensations. It starts by identifying several suitable pinching poses on the considered tangible and virtual objects. Then, for each pose, it evaluates a series of haptically-salient characteristics. Next, it identifies the two most similar pinching poses according to these metrics, one on the tangible and one on the virtual object. Finally, it highlights the chosen pinching pose, which provides the best matching sensation between what users see and touch. The effectiveness of our approach is evaluated through a user study. Results show that the algorithm is able to well combine several haptically-salient object features to find convincing pinches between the given tangible and virtual objects.
This work was done in collaboration with the Inria Rainbow team.

7.3.5. Investigating the recognition of local shapes using mid-air ultrasound haptics

Participants: Thomas Howard, Gerard Gallagher, Anatole Lécuyer, Maud Marchal [contact].

Mid-air haptics technologies are able to convey haptic sensations without any direct contact between the user and the haptic interface. One representative example of this technology is ultrasound haptics, which uses ultrasonic phased arrays to deliver haptic sensations. Research on ultrasound haptics is only in its beginnings, and the literature still lacks principled perception studies in this domain. This work [22] presents a series of human subject experiments investigating important perceptual aspects related to the rendering of 2D shapes by an ultrasound haptic interface (the Ultrahaptics STRATOS platform, see Figure 13). We carried out four user studies aiming at evaluating (i) the absolute detection threshold for a static focal point rendered via amplitude modulation, (ii) the absolute detection and identification thresholds for line patterns rendered via spatiotemporal modulation, (iii) the ability to discriminate different line orientations, and (iv) the ability to perceive virtual bumps and holes. These results shed light on the rendering capabilities and limitations of this novel technology for 2D shapes.
This work was done in collaboration with the Inria Rainbow team.

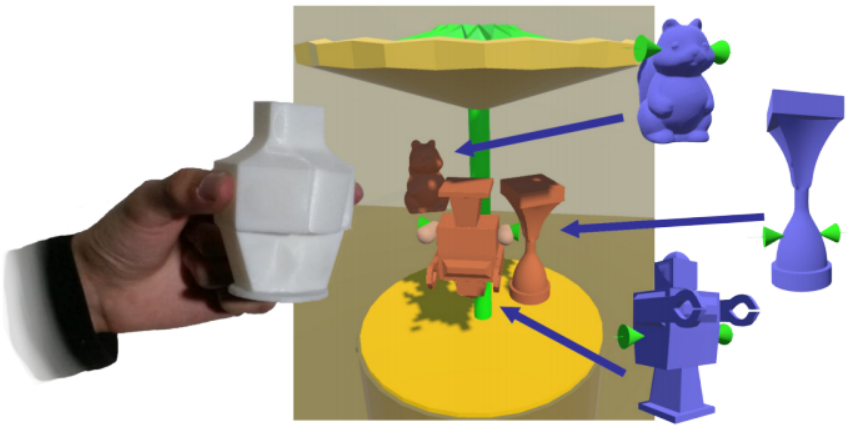


Figure 12. Illustration of our approach through a carousel of virtual objects that can be grasped using a single "universal" tangible object. The user is able to turn the virtual carousel and manipulate the three virtual objects using the suggested pinch poses (in green). These poses are proposed by our algorithm to best match the corresponding haptic pinching sensations on the tangible object.

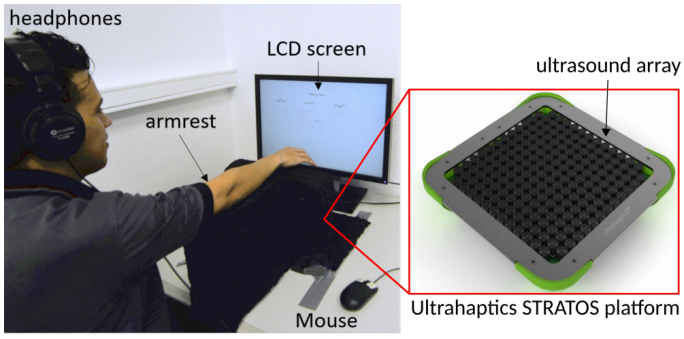


Figure 13. Experimental setup to investigate the recognition of local shapes using mid-air ultrasound haptics.

7.3.6. Touchy: Tactile Sensations on Touchscreens Using a Cursor and Visual Effects

Participants: Antoine Costes, Ferran Argelaguet, Anatole Lécuyer [contact].

Haptic enhancement of touchscreens usually involves vibrating motors that produce limited sensations or custom mechanical actuators that are difficult to widespread. In this work, we propose an alternative approach called "Touchy" to induce haptic sensations in touchscreens through purely visual effects [3]. Touchy introduces a symbolic cursor under the user's finger which shape and motion are altered in order to evoke haptic properties. This novel metaphor enables to address four different perceptual dimensions, namely: hardness, friction, fine roughness and macro roughness. Our metaphor comes with a set of seven visual effects that we compared with real texture samples within a user study conducted with 14 participants. Taken together our results show that Touchy is able to elicit clear and distinct haptic properties: stiffness, roughness, reliefs, stickiness and slipperiness.
This work was achieved in collaboration with InterDigital.

7.3.7. Investigating Tendon Vibration Illusions

Participants: Salomé Lefranc [contact], Mélanie Cogné, Mathis Fleury, Anatole Lécuyer.

Illusion of movement induced by tendon vibration can be useful in applications such as rehabilitation of neurological impairments. In [40], we investigated whether a haptic proprioceptive illusion induced by a tendon vibration of the wrist congruent to the visual feedback of a moving hand could increase the overall illusion of movement. Tendon vibration was applied on the non-dominant wrist during 3 visual conditions: a moving virtual hand corresponding to the movement that the subjects could feel during the tendon vibration (Moving condition), a static virtual hand (Static condition), or no virtual hand at all (Hidden condition). There was a significant difference between the 3 visual feedback conditions, and the Moving condition was found to induce a higher intensity of illusion of movement and higher sensation of wrist's extension. Therefore, our study demonstrated the potentiation of illusion by visual cues congruent to the illusion of movement. Further steps will be to test the same hypothesis with stroke patients and use our results to develop EEG-based Neurofeedback including vibratory feedback to improve upper limb motor function after a stroke.
This work was achieved in collaboration with CHU Rennes and Inria EMPENN team.

7.4. Brain-Computer Interfaces

7.4.1. Defining Brain-Computer Interfaces: A Human-Computer Interaction Perspective

Participants: Hakim Si Mohammed, Ferran Argelaguet, Anatole Lécuyer [contact].

Regardless of the term used to designate them, Brain-Computer Interfaces (BCIs) are "Interfaces" between a user and a computer in the broad sense of the term. This paper aims to discuss how BCIs have been defined in the literature from the day the term was introduced by Jacques Vidal. In [32], from a Human-Computer Interaction perspective, we propose a new definition of Brain-Computer Interfaces as : "any artificial systems that directly converts brain activity into input of a computer process". As they are interfaces, such definition should not include the finality and objective of the system they are used to interact with. To illustrate this, we compared BCIs with other widely used Human-Computer Interfaces, and drew analogies in their conception and purpose.

This work was done in collaboration with the Inria LOKI team.

7.4.2. A conceptual space for EEG-based brain-computer interfaces

Participant: Anatole Lécuyer [contact].

Brain-Computer Interfaces have become more and more popular these last years. Researchers use this technology for several types of applications, including attention and workload measures but also for the direct control of objects by the means of BCIs. In [7] we present a first, multidimensional feature space for EEG-based BCI applications to help practitioners to characterize, compare and design systems, which use EEG-based BCIs. Our feature space contains 4 axes and 9 sub-axes and consists of 41 options in total as well as their different combinations. In addition we present the axes of our feature space and we position our feature space regarding the existing BCI and HCI taxonomies. We also showed how our work integrates the past works, and/or complements them.

7.4.3. The use of haptic feedback in Brain-Computer Interfaces and Neurofeedback

Participants: Mathis Fleury, Anatole Lécuyer [contact].

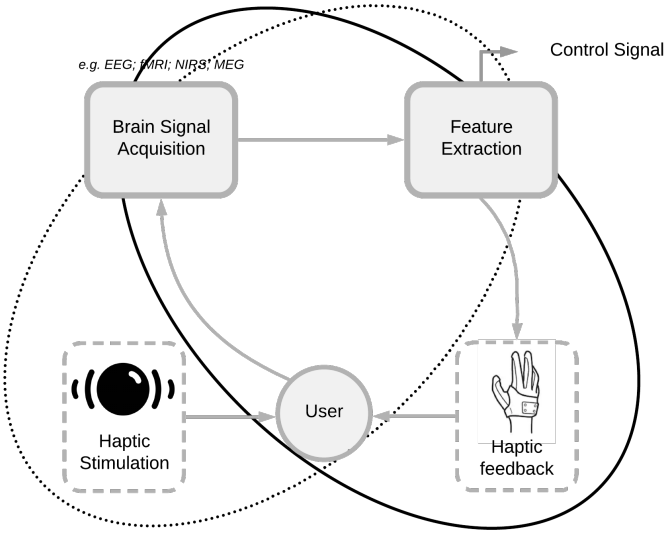


Figure 14. Using haptic feedback in active and reactive Brain-Computer Interfaces (BCI). In active BCI, haptics provide feedback from user's neural activity (black ellipse). In reactive BCI, haptics provide a stimulation to elicit a specific brain activity (black doted ellipse).

Neurofeedback (NF) and brain-computer interfaces are based on the recording of the cerebral activity associated with the requested task and the presentation of a feedback. The subject relies on the given feedback (visual, auditory or haptic) to learn and improve his mental strategy. It is therefore of crucial importance that it must be transmitted optimally. Historically, vision is the most used sensory modality in BCI/NF applications, but its use is raising potential issues. The more and more frequent use of haptic as a feedback modality reveals the limits of visual feedback; indeed, a visual feedback is not suitable in some cases, for individuals with an impaired visual system or during a mental motor imagery task (e.g. requiring a great abstraction). In such case, a haptic feedback would seem more appropriate. Haptic feedback has also been reported to be more engaging than visual feedback. This feedback could also contribute to close the sensory-motor loop. Haptic-based BCI/NF is a promising alternative for the design of the feedback and potentially improve the clinical efficacy of NF. In [38], [39] we have therefore surveyed the recent studies exploiting haptic feedback in BCI and NF.

This work was achieved in collaboration with the Inria EMPENN team.

7.4.4. Efficacy of EEG-fMRI Neurofeedback for stroke rehabilitation: a pilot study

Participants: Giulia Lioi, Mathis Fleury, Anatole Lécuyer [contact].

Recent studies have shown the potential of neurofeedback for motor rehabilitation after stroke. The majority of these NF approaches have relied solely on one imaging technique: mostly on EEG recordings. Recent study have gone further, revealing the potential of integrating complementary techniques such as EEG and fMRI to achieve a more specific regulation. In this exploratory work, multi-session bimodal EEG-fMRI NF for upper limb motor recovery was tested in four stroke patients.The feasibility of the NF training was investigated [41] with respect to the integrity of the cortico-spinal tract (CST), a well-established predictor of the potential for

clinical improvement. Results indicated that patients exhibiting a high degree of integrity of the ipsilesional CST showed significant increased activation of the ipsilesional M1 at the end of the training. These preliminary findings confirm the critical role of the CST integrity for stroke motor recovery and indicate that this is importantly related also to functional brain regulation of the ipsilesional motor cortex.
This work was achieved in collaboration with Inria EMPENN team.

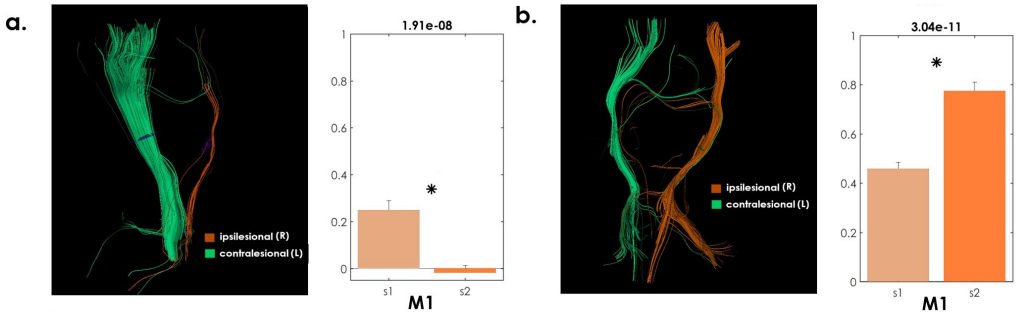


Figure 15. Example of CST reconstruction and primary motor cortex (M1) activation in two patients (a. and b.). Ipsilesional CST is plotted in orange and controlesional CST in green. The bar plot on the right hand side of the figure show the average (and standard error across NF training blocks) of BOLD contrast activation in the primary motor cortex in the first (s1) and second (s2) training session, with relative statistics.

7.4.5. A multi-target motor imagery training using EEG-fMRI Neurofeedback

Participants: Giulia Lioi, Mathis Fleury, Anatole Lécuyer [contact].

Upper limb recovery after stroke is a complex process. Recent studies have revealed the potential of neurofeedback training as an alternative or an aid to traditional therapies. Studies on cerebral plasticity and recovery after stroke indicate that premotor areas should be a preferred target for NF in the most severe patients while M1 stimulation may be more ef for patients with better recovery potential. Moreover, fMRI-NF studies (also on stroke patients) have shown that SMA is a robust correlate of motor imagery, while the activation of M1 is more dif to achieve, especially for short training sessions. Based on these results, in an exploratory work [13], we tested a dynamic NF training more strongly rewarding SMA activation in the NF training session and then increasing the M1 activation contribution in the NF session. We tested this novel approach on four stroke patients in a multisession bimodal EEG-fMRI NF training. To this end, we used an adaptive cortical region of interest (ROI) equal to a weighted combination of ipsilesional SMA and M1 activities and then varied the weights in order guide the patient training towards an improved activation of M1. Four chronic stroke patients with left hemiparesis participated to the study. The experimental protocol included an alternation of bimodal EEG-fMRI NF and unimodal EEG-only NF sessions. Preliminary results, on a short training duration, reveal the potential of a dynamic, multi-target/multimodal NF training approach.
This work was achieved in collaboration with Inria EMPENN team.

7.4.6. Bimodal EEG-fMRI Neurofeedback for upper motor limb rehabilitation

Participants: Giulia Lioi, Mathis Fleury, Anatole Lécuyer [contact].

There is a growing interest in Neurofeedback or Brain computer interfaces for stroke rehabilitation. Integrating EEG and fMRI, two highly complementary imaging modalities, has potential to provide a more specific and efficient stimulation of motor areas. In this exploratory work [25], we tested the feasibility of a multi-session EEG-fMRI NF protocol on four chronic stroke patients, and its potential for upper-limb recovery. All the patients were able to upregulate their activity during NF training with respect to rest in the ipsilesional SMA and M1. Three over four patients showed a significant increase in ipsilesional M1 activation at the end of the protocol. Of these three individuals, two exhibited an increase in FMA-UE score. Preliminary results from this pilot study showed feasibility of bimodal EEG-fMRI in chronic stroke patients and indicated the potential of this training protocol for upper-limb recovery.
This work was achieved in collaboration with Inria EMPENN team.

7.5. Cultural Heritage

7.5.1. Expressive potentials of motion capture in the Vis Insita musical performance

Participants: Ronan Gaugne [contact], Florian Nouviale, Valérie Gouranton.

The electronic music performance project Vis Insita [10] implements the design of experimental instrumental interfaces based on optical motion capture technology with passive infrared markers (MoCap), and the analysis of their use in a real scenic presentation context (Figure 16). Because of MoCap's predisposition to capture the movements of the body, a lot of research and musical applications in the performing arts concern dance or the sonification of gesture. For our research, we wanted to move away from the capture of the human body to analyse the possibilities of a kinetic object handled by a performer, both in terms of musical expression, but also in the broader context of a multimodal scenic interpretation.
This work was done in collaboration with Univ. Rennes 2, France.

Figure 16. The Vis Insita performance.

7.5.2. Interactive and Immersive Tools for Point Clouds in Archaeology

Participants: Ronan Gaugne [contact], Quentin Petit, Valérie Gouranton.

A framework is presented for an immersive and interactive 3D manipulation of large point clouds, in the context of an archaeological study [19]. The framework was designed in an interdisciplinary collaboration with archaeologists. We first applied this framework for the study of an 17th-century building of a Real Tennis court (Figure 17). We propose a display infrastructure associated with a set of tools that allows archaeologists to interact directly with the point cloud within their study process. The resulting framework allows an immersive navigation at scale 1:1 in a dense point cloud, the manipulation and production of cut plans and cross sections, and the positioning and visualisation of photographic views. We also apply the same framework to three other archaeological contexts with different purposes, a 13th century ruined chapel, a 19th-century wreck and a cremation urn from the Iron Age.
This work was done in collaboration with UMR CREAAH, France.

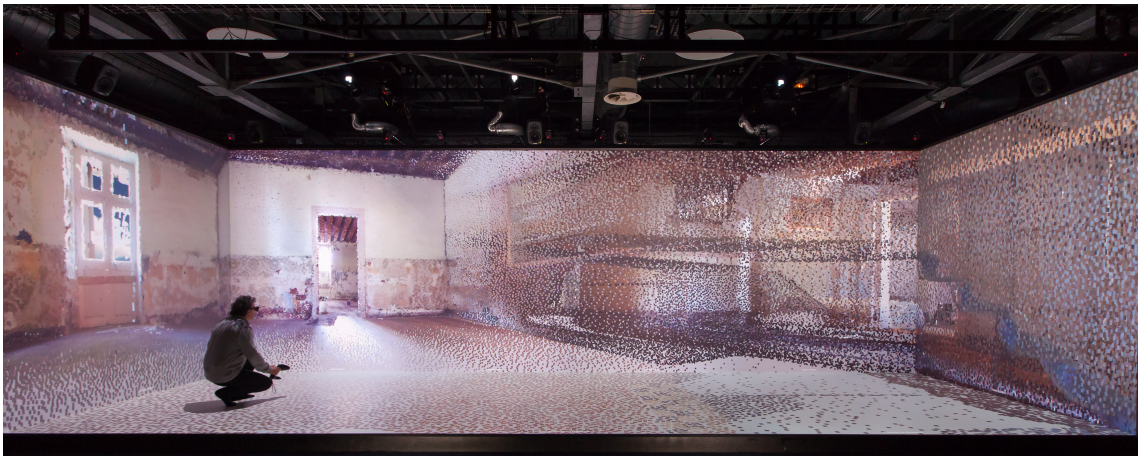


Figure 17. Immersive view of the point cloud of the Real Tennis building.

7.5.3. Making virtual archeology great again (without scientific compromise)

Participants: Ronan Gaugne, Valérie Gouranton [contact].

In the past two decades or so, digital tools have been slowly integrated as part of the archaeological process of information acquisition, analysis, and dissemination. We are now entering a new era, adding the missing piece to the puzzle in order to complete this digital revolution and take archaeology one step further into virtual reality (VR). The main focus of this work is the methodology of digital archaeology that fully integrates virtual reality, from beta testing to interdisciplinary teamwork. After data acquisition and processing necessary to construct the 3D model, we explore the analysis that can be conducted during and after the making or creation of the 3D environment and the dissemination of knowledge. We explain the relevance of this methodology through the case study on the intendant's palace, an 18th century archaeological site in Quebec City, Canada (Figure 18 left). With this experience, we believe that VR can prompt new questions that would never have occurred otherwise and can provide technical advantages in terms of gathering data in the same virtual space (Figure 18 right). We conclude that multidisciplinary input in archaeological research is once again proven essential in this new, inclusive and vast digital structure of possibilities [29].
This work was done in collaboration with UMR CREAAH, Inrap, France and Univ. Laval, Canada.

7.5.4. Evaluation of a Mixed Reality based Method for Archaeological Excavation Support

Participants: Ronan Gaugne [contact], Quentin Petit, Valérie Gouranton.

Figure 18. Left, model of the XVIIth century Palais de l'Intendant. Right, study of the reconstitution of the Palais de L'intendant and its neighborhood inside Immersia

In the context of archaeology, most of the time, micro-excavation for the study of furniture (metal, ceramics...) or archaeological context (incineration, bulk sampling) is performed without complete knowledge of the internal content, with the risk of damaging nested artifacts during the process. The use of medical imaging coupled with digital 3D technologies, has led to significant breakthroughs by allowing to refine the reading of complex artifacts. However, archaeologists may have difficulties in constructing a mental image in 3 dimensions from the axial and longitudinal sections obtained during medical imaging, and in the same way to visualize and manipulate a complex 3D object on screen, and an inability to simultaneously manipulate and analyze a 3D image, and a real object. Thereby, if digital technologies allow a 3D visualization (stereoscopic screen, VR headset ...), they are not without limiting the natural, intuitive and direct 3D perception of the archaeologist on the material or context being studied. We therefore propose a visualization system based on optical see-through augmented reality that associates real visualization of archaeological material with data from medical imaging [18] (see Figure 19). This represents a relevant approach for composite or corroded objects or contexts associating several objects such as cremations. The results presented in the paper identify adequate visualization modalities to allow archaeologist to estimate, with an acceptable error, the position of an internal element in a particular archaeological material, an Iron-Age cremation block inside a urn.
This work was done in collaboration with Inrap, France and AIST (National Institute of Advanced Industrial Science and Technology), Japan.

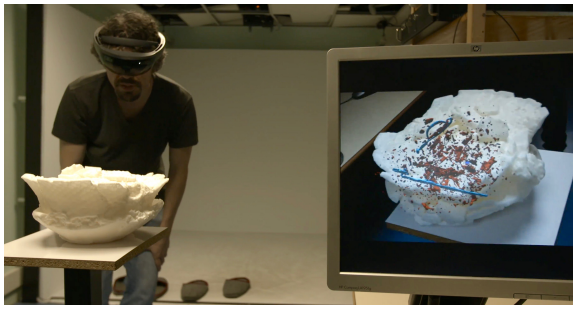


Figure 19. Evaluation of the mixed reality system

ILDA Project-Team

7. New Results

7.1. Digital Ink and Data Manipulation

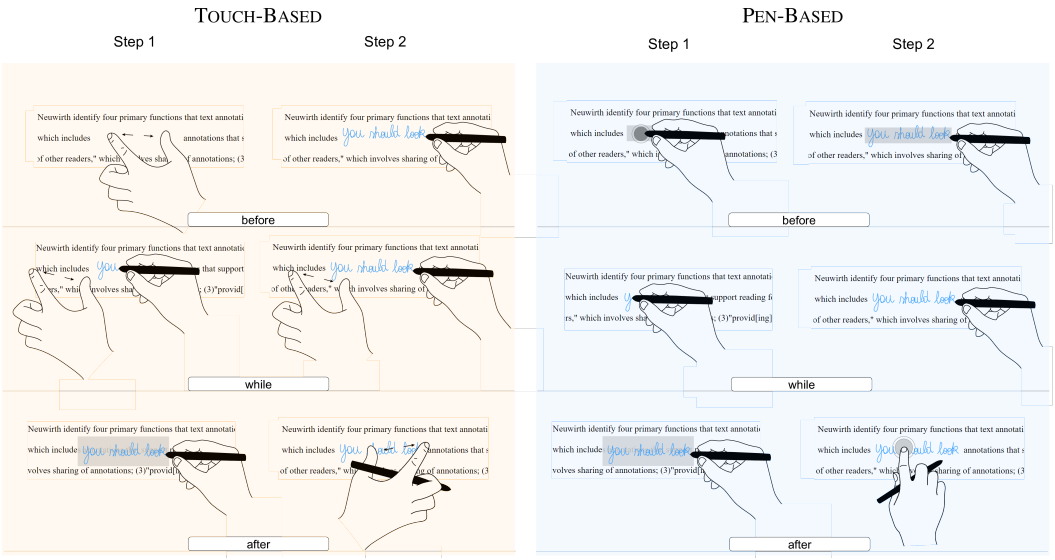


Figure 3. In the SpaceInk design space, both touch-based and pen-based techniques let users specify the strategy for creating white space at different moments: either before, while, or after annotating. With touch-based techniques, users are free to both gesture and write with a single hand (as illustrated in the after condition) or with two hands (as illustrated in the before condition).

We investigated how pen and touch could be best combined to facilitate the digital annotation of documents. When editing or reviewing a document, people directly overlay ink marks on content. For instance, they underline words, or circle elements in a figure. These overlay marks often accompany in-context annotations in the form of handwritten footnotes and marginalia. People tend to put annotations close to the content that elicited them, but have to compose with the often-limited whitespace. Based on these observations, we explored a design space – which we call SpaceInk (UIST 2019 [9]) – of pen+touch techniques that make room for in-context annotations by dynamically reflowing documents. We identified representative techniques in this design space, spanning both new ones and existing ones, as illustrated in Figure 3 . We evaluated them in a user study. The results of this study then informed the design of a prototype system which lets users concentrate on capturing fleeting thoughts, streamlining the overall annotation process by enabling the fluid inverleaving of space-making gestures with freeform ink.

Together with colleagues from the EPIC team at Microsoft Research (see Section 9.3.2.1), we also investigated the potential of digital inking for exploring heterogeneous datasets and trying to make sense of them. During sensemaking, people annotate insights: underlining sentences in a document or circling regions on a map. They jot down their hypotheses: drawing correlation lines on scatterplots or creating personal legends to track patterns. Based on these observations, we designed ActiveInk (CHI 2019 [22]), a system enabling people to seamlessly transition between exploring data and externalizing their thoughts using pen and touch as input

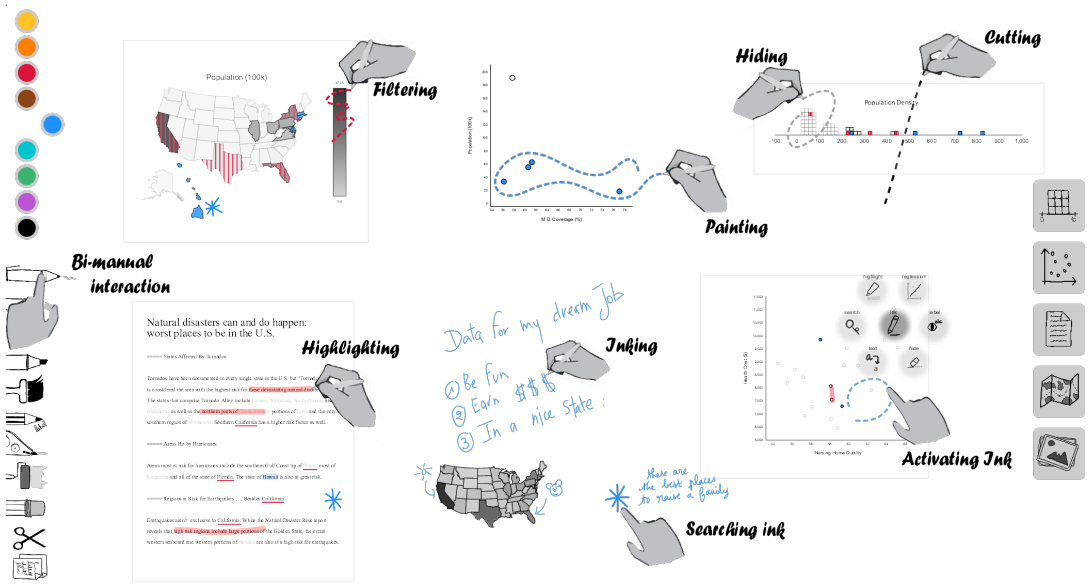


Figure 4. ActiveInk, a collaboration with Microsoft Research, affords smooth transition between using a digital pen for high-precision selections of heterogeneous data coming from multiple sources, and for externalizing thinking via notes and annotations. Ink strokes are leveraged to perform operations on underlying data.

channels. ActiveInk, illustrated in Figure 4 , enables the natural use of pen for active reading behaviors, while supporting analytic actions by activating any of these ink strokes. Through a qualitative study with eight participants, we observed active reading behaviors during data exploration and design principles to support sensemaking.

7.2. Novel Forms of Input in Immersive Environments

ILDA researchers have started to investigate input techniques for the specific context of immersive environments, based on, e.g., virtual or augmented reality. These hardware devices enable displaying large amounts of data in space to better support data analysis, and there is a growing group of research focusing on Immersive Analytics in the HCI community. The question of input for data manipulation in these environments is crucial, but challenging because users must be able to activate various commands or adjust various values while remaining free to move. In this context, using the whole body as an input device offers several advantages: 1) the body provides physical support as an interactive surface, which improves accuracy and makes it less tiring to interact; 2) using the body does not impair mobility and avoids handling devices; 3) proprioception makes it possible to interact eyes-free, including when choosing values in a range; 4) by leveraging spatial memory, the body helps memorizing commands, thus interacting in expert mode (i.e., perform quick actions without visual feedback). ILDA team members participated to a position paper on this topic, which analyzes various ways of interacting with the body, discussing their pros and cons as well as associated challenges for immersive analytics [23].

7.3. Multivariate Network Visualization

Edges in networks often represent transfer relationships between vertices. When visualizing such networks as node-link diagrams, animated particles flowing along the links can effectively convey this notion of transfer. Variables that govern the motion of particles, their speed in particular, may be used to visually represent

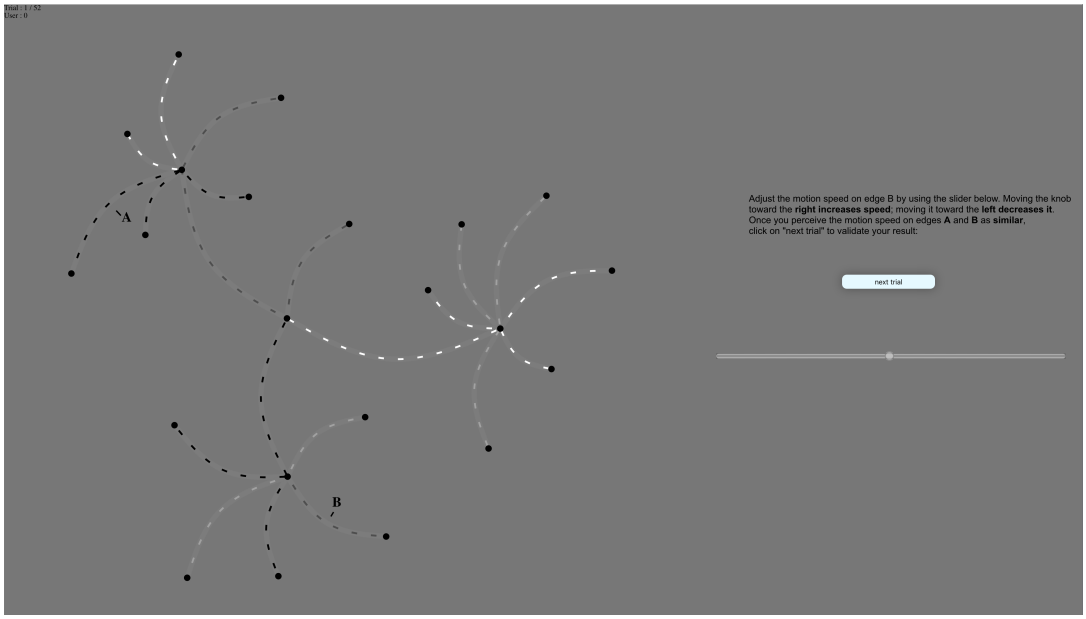


Figure 5. Experimental task used in the study about the influence of color and size of moving particles on their perceived speed in node-link diagrams. Participants had to adjust the speed of particles on a specific edge so that it would match that of particles on another edge, using a slider.

edge data attributes. Few guidelines exist to inform the design of these particle-based network visualizations, however. Following up on our initial investigation of motion as an encoding channel for edge attributes in multivariate network visualization [75], we investigated the influence of color and size of moving particles on their perceived speed in node-link diagrams (INTERACT 2019 [20]). Empirical studies so far had only looked at the different motion variables in isolation, independently from other visual variables controlling the appearance of particles, such as their color or size. We ran a study of the influence of several visual variables on users' perception of the speed of particles. Part of the experimental setup is illustrated in Figure 5 . Our results show that particles' luminance, chromaticity and width do not interfere with their perceived speed. But variations in their length make it more difficult for users to compare the relative speed of particles across edges.

Figure 6. Graphies running on a tablet with support for pen+touch. Graphies is developed entirely using Web technologies.

Beyond questions of perception of information in multivariate network visualizations, we also investigated the problem of creating environments for the design of multivariate network visualizations, with a focus on expressive design. Expressive design environments enable visualization designers not only to specify chart types and visual mappings, but also to customize individual graphical marks, as they would in a vector graphics drawing tool. Prior work had mainly investigated how to support the expressive design of a wide range of charts generated from tabular data: bar charts, scatterplots, maps, *etc*. But multivariate network data structures raise specific challenges and opportunities in terms of visual design and interactive authoring. Together with company TKM (see Section 8.1), we developed an expressive design environment for node-link diagrams generated from multivariate networks called Graphies (TVCG 2019, [15]), illustrated in Figure 6 . We followed a user-centered design approach, involving expert analysts from TKM, and validated the approach through a study in which participants successfully reproduced several expressive designs, and created their own designs as well.

7.4. Visualization in Specific Application Areas

Finally, we worked in collaboration with other researchers on projects aimed at investigating how visualization can support experts in different application areas.

In the area of geovisualization, we performed a comparison of visualization techniques to help analysts identify correlation between variables over space and time (TVCG/InfoVis 2019, [6]). Observing the relationship between two or more variables over space and time is essential in many application domains. For instance, looking, for different countries, at the evolution of both the life expectancy at birth and the fertility rate will give an overview of their demographics. The choice of visual representation for such multivariate data is key to enabling analysts to extract patterns and trends. We conducted a study comparing three techniques that are representative of different strategies to visualize geo-temporal multivariate data. Participants performed a series of tasks that required them to identify if two variables were correlated over time and if there was a pattern in their evolution. Our results showed that a visualization's effectiveness depends strongly on the task to be carried out. Based on this study's findings, we derived a set of design guidelines about geo-temporal visualization techniques for communicating correlation.

Together with researchers from INRA, we performed an exploratory study about the visual exploration of model simulations for a range of experts (CHI 2019, [16]). Experts in different domains rely increasingly on simulation models of complex processes to reach insights, make decisions, and plan future projects. These models are often used to study possible trade-offs, as experts try to optimize multiple conflicting objectives in a single investigation. Understanding all the model intricacies, however, is challenging for a single domain expert. This project introduced a simple approach to support multiple experts when exploring complex model results, working concurrently on a shared visualization surface. The results of an observational study focusing on the link between expertise and insight generation during the analysis process, revealed the different exploration strategies and multi-storyline approaches that domain experts adopt during trade-off analysis. This eventually led to recommendations for collaborative model exploration systems.

We collaborated with researchers in databases from Université Paris Descartes on progressive similarity search on time-series data (BigVis 2019 workshop, [24]). Time-series data are increasing at a dramatic rate, yet their analysis remains highly relevant in a wide range of human activities. Due to their volume, existing systems dealing with time-series data cannot guarantee interactive response times, even for fundamental tasks such as similarity search. This paper presented our vision to develop analytic approaches that support exploration and decision making by providing progressive results, before the final and exact ones have been computed. Findings from our experiment indicated that there is a gap between the time the most similar answer is found and the time when the search algorithm terminates, resulting in inflated waiting times without any improvement. These findings led to preliminary ideas about computing probabilistic estimates of the final results that could help users decide when to stop the search process.

In the field of Education, we contributed to EduClust, an online visualization application for teaching clustering algorithms (EuroGraphics 2019, [18]). EduClust combines visualizations, interactions, and animations to facilitate the understanding and teaching of clustering steps, parameters, and procedures. Traditional classroom settings aim for cognitive processes like remembering and understanding. We designed EduClust for expanded educational objectives like applying and evaluating. The application can be used by both educators to prepare teaching material and examples, and by students to explore clustering differences and discover algorithmic subtleties.

IMAGINE Project-Team

7. New Results

7.1. Star-Shaped Metrics for Mechanical Metamaterial Design

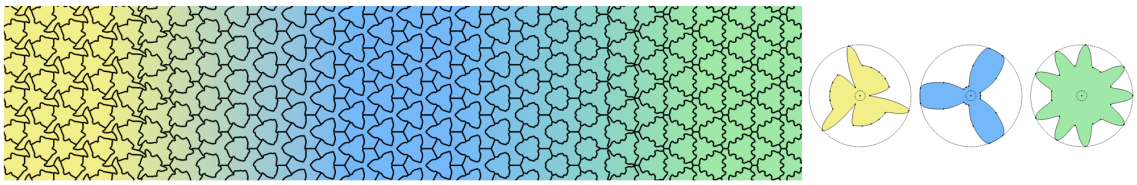


Figure 1. Our method generates a smoothly-graded pattern (left) when interpolating between three star-shaped distance functions (right) on a regular honeycomb lattice. Each distance function is compactly parameterized with polar coordinates, allowing for simple interpolation in metric space as indicated by color-coding.

We present a method for designing mechanical metamaterials based on the novel concept of Voronoi diagrams induced by star-shaped metrics. As one of its central advantages, our approach supports interpolation between arbitrary metrics, as depicted in Figure 1 . This capability opens up a rich space of structures with interesting aesthetics and a wide range of mechanical properties, including isotropic, tetragonal, orthotropic, as well as smoothly graded materials. We evaluate our method by creating large sets of example structures, provided as accompanying material. We validate the mechanical properties predicted by simulation through tensile tests on a set of physical prototypes.

7.2. Computational Design of Fabric Formwork

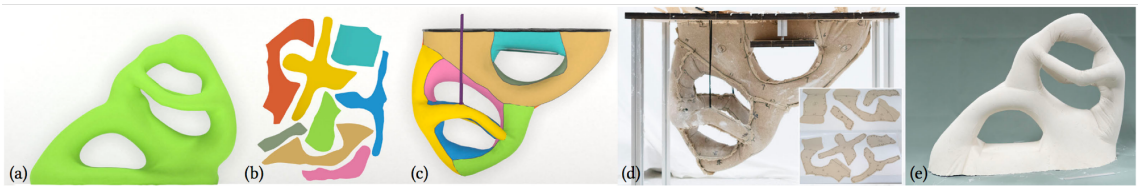


Figure 2. A fertility model designed and fabricated using our computational approach. For a target 3D model (a), our system can automatically compute a set of flat panels (b) that can be sewn together to serve as fabric containers to form a target shape by pressure of liquid plaster poured in – see (c) for the simulation under force equilibrium of membrane tension, liquid pressure and external supports. The generated flat panels are used to conduct the physical fabrication of fabric formwork (d). After drying and unwrapping the fabric container, a sculpture with the designed target shape has been fabricated (e).

This work (illustrated in Figure 2) presents an inverse design tool for fabric formwork - a process where flat panels are sewn together to form a fabric container for casting a plaster sculpture. Compared to 3D printing techniques, the benefit of fabric formwork is its properties of low-cost and easy transport. The process of fabric formwork is akin to molding and casting but having a soft boundary. Deformation of the fabric container is governed by force equilibrium between the pressure forces from liquid fill and tension in the stretched fabric. The final result of fabrication depends on the shapes of the flat panels, the fabrication orientation and the placement of external supports. Our computational framework generates optimized flat panels and fabrication orientation with reference to a target shape, and determines effective locations for external supports. We demonstrate the function of this design tool on a variety of models with different shapes and topology. Physical fabrication is also demonstrated to validate our approach.

7.3. Spatial Motion Doodles

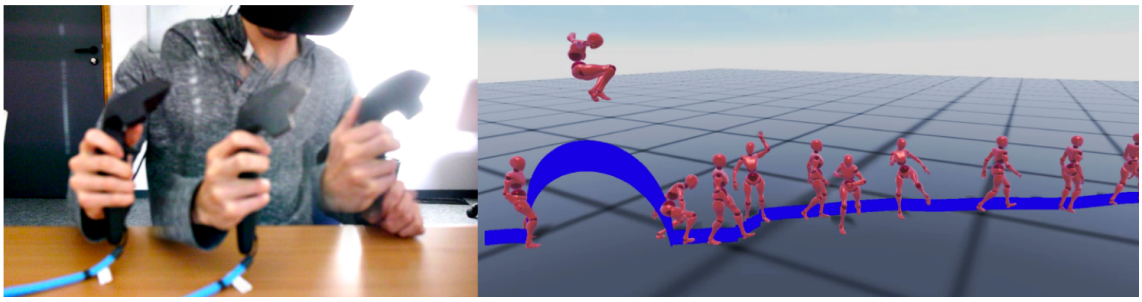


Figure 3. Left: A user drawing a spatial motion doodle (SMD) which is the six-dimensional trajectory of a moving frame (position and orientation), here attached to the HTC Vive controller. Right: The SMD is parsed into a string of motion tokens,allowing to recognize actions and extract the associated motion qualities. This information is transferred to an articulated character to generate an expressive 3D animation sequence.

We present a method for easily drafting expressive character animation by playing with instrumented rigid objects (see Figure 3). We parse the input 6D trajectories (position and orientation over time) – called spatial motion doodles – into sequences of actions and convert them into detailed character animations using a dataset of parameterized motion clips which are automatically fitted to the doodles in terms of global trajectory and timing. Moreover, we capture the expressiveness of user-manipulation by analyzing Laban effort qualities in the input spatial motion doodles and transferring them to the synthetic motions we generate. We validate the ease of use of our system and the expressiveness of the resulting animations through a series of user studies, showing the interest of our approach for interactive digital storytelling applications dedicated to children and non-expert users, as well as for providing fast drafting tools for animators.

7.4. Text-to-Movie Authoring of Anatomy Lessons

With popular use of multimedia and 3D content in anatomy teaching there is a need for a simple yet comprehensive tool to create and edit pedagogical anatomy video lessons. This work introduces an automated video authoring tool (shown in Figure 4) created for teachers. It takes text written in a novel domain specific language (DSL) called the Anatomy Storyboard Language (ASL) as input and translates it to real time 3D animation. Preliminary results demonstrates the ease of use and effectiveness of the tool for quickly drafting video lessons in realistic medical anatomy teaching scenarios.

7.5. Approximate Reconstruction of 3D Scenes From Bas-Reliefs

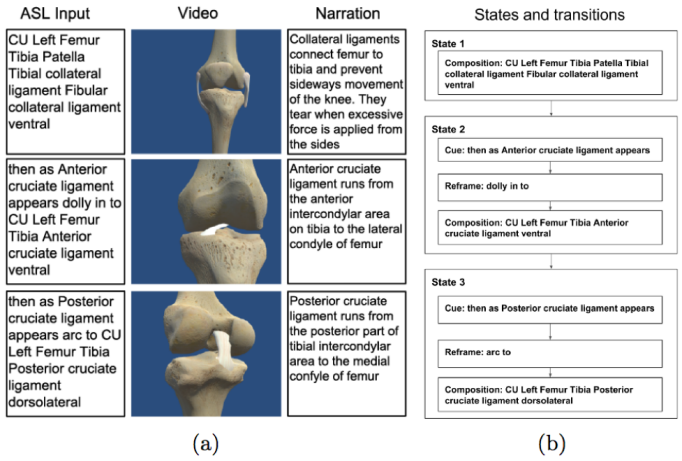


Figure 4. Text-to-movie generation example with hierarchical finite state machines representation.

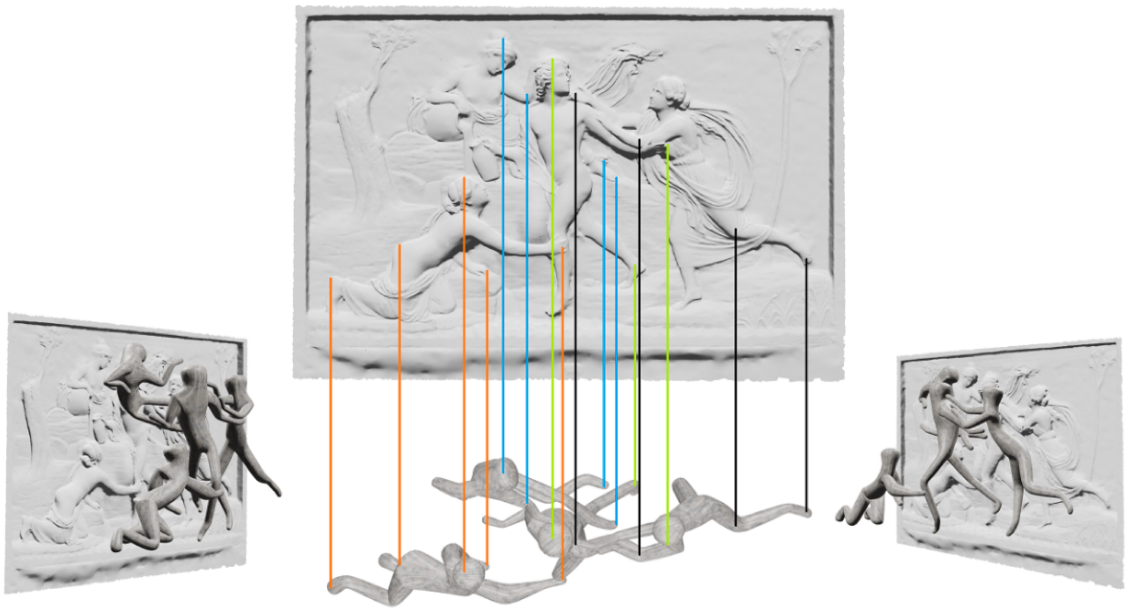


Figure 5. 3D interpretation of the mythological story of Hylas and the Water Nymphs, after a bas-relief marble by Bertel Thorvaldsen (1833). Hylas was sent to fetch water for the camp. Finding a pool in a clearing, he was encircled by water nymphs reaching up to kiss him and there disappeared with them forever. Using hand-drawn silhouette shapes and 2D skeletons of the four characters, we compute a plausible 3D reconstruction of the scene with rigged and skinned models suitable for 3D animation.

For thousands of years, bas-reliefs such as the one depicted in Figure 5 have been used to depict scenes of everyday life, mythology and historic events. Yet, the precise geometry of those scenes remains difficult to interpret and reconstruct. Over the past decade, methods have been developed for generating bas-reliefs from 3D scenes. With this work, we investigate the inverse problem of interpreting and reconstructing 3D scenes from their bas-relief depictions. Even approximate reconstructions can be useful for art historians and museum exhibit designers, as a first entry to the complete interpretation of the narratives told in stone or marble. To create such approximate reconstructions, we present methods for extracting 3D base mesh models of all characters depicted in a bas-relief. We take advantages of the bas-relief geometry and high-level knowledge of human body proportions to recover body parts and their three-dimensional structure, even in severe cases of contact and occlusion. We present experimental results for 6 bas-relief depictions of Greek mythological and historical scenes involving 18 characters and draw conclusions for future work.

LOKI Project-Team

7. New Results

7.1. Introduction

According to our research program, in the next two to five years, we will study dynamics of interaction along three levels depending on interaction time scale and related user's perception and behavior: *Micro-dynamics*, *Meso-dynamics*, and *Macro-dynamics*. Considering phenomena that occur at each of these levels as well as their relationships will help us to acquire the necessary knowledge (Empowering Tools) and technological bricks (Interaction Machine) to reconcile the way interactive systems are designed and engineered with human abilities. Our strategy is to investigate issues and address challenges for all of the three levels of dynamics. Last year we focused on micro-dynamics since it concerns very fundamental knowledge about interaction and relates to very low-level parts of interactive systems. In 2019 we were able to build upon those results (micro), but also to enlarge the scope of our studies within larger interaction time scales, especially at the meso-dynamic level. Some of these results have also contributed to our objective of defining the basic principles of an Interaction Machine.

7.2. Micro-dynamics

Participants: Géry Casiez [contact person], Sylvain Malacria, Mathieu Nancel, Thomas Pietrzak.

7.2.1. Latency & Transfer functions

End-to-end latency in interactive systems is detrimental to performance and usability, and comes from a combination of hardware and software delays. While these delays are steadily addressed by hardware and software improvements, it is at a decelerating pace. In parallel, short-term input prediction has recently shown promising results to compensate for latency, in both research and industry.

In the context of the collaborative TurboTouch project, we proposed a method based on a frequency-domain approximation of a non-causal ideal predictor with a finite impulse response filter. Given a sufficiently rich dataset, the parameters of the filter can be either optimized off-line or tuned on-line with the proposed adaptive algorithm. The performance of the proposed solution is evaluated in an experimental study consisting of drawings on a touchscreen [13].

On the related topic of transfer functions, we proposed a switched dynamic model to model indirect pointing tasks with a computer mouse. The model contains a ballistic movement phase governed by a nonlinear model in Lurie form and a corrective movement phase described by a linear visual-feedback system. The stability of the model was evaluated and the derived model was then validated with experimental data acquired in a pointing task with a mouse. Numerical comparison to pointing models available in the literature is also provided [12].

7.2.2. 3D interaction

Raycasting is the most common target pointing technique in virtual reality environments. However, performance on small and distant targets is impacted by the accuracy of the pointing device and the user's motor skills. Current pointing facilitation techniques are currently only applied in the context of the virtual hand, i.e. for targets within reach. We proposed enhancements to Raycasting: filtering the ray, and adding a controllable cursor on the ray to select the nearest target (Figure 2). We ran a series of studies for the design of the visual feedforward, filtering technique, as well as a comparative study between different 3D pointing techniques. Our results show that highlighting the nearest target is one of the most efficient visual feedforward technique. We also show that filtering the ray reduces error rate in a drastic way. Finally we show the benefits of RayCursor compared to Raycasting and another technique from the literature [19], [14].

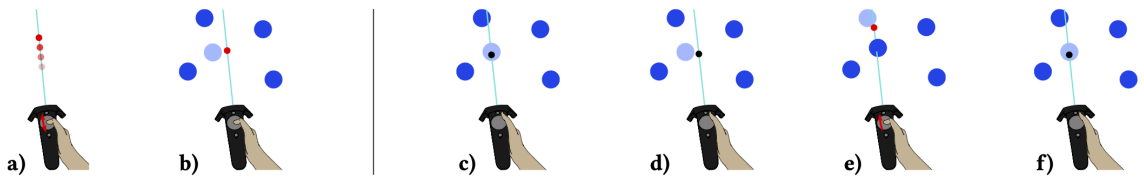


Figure 2. Illustration of manual RayCursor: a) the user controls a cursor along the ray using relative displacements of their thumb on the controller's touchpad; b) the target closest to the cursor is highlighted. Illustration of semi-auto RayCursor: c) by default, it works like Raycasting. The cursor (in black) is positioned at the intersection with a target; d) the target remains selected if the cursor moves out of the target, until it is closer to another target; e) the user can manually move the cursor using the controller's touchpad, to select another target (the cursor turns red to indicate manual mode); f) if the user does not touch the touchpad for 1s, the cursor returns to its behaviour described in c).

7.3. Meso-dynamics

Participants: Axel Antoine, Marc Baloup, Géry Casiez, Stéphane Huot, Edward Lank, Sylvain Malacria, Mathieu Nancel, Thomas Pietrzak [contact person], Thibault Raffaillac, Marcelo Wanderley.

7.3.1. Production of illustrative supports

Trace figures are contour drawings of people and objects that capture the essence of scenes without the visual noise of photos or other visual representations. Their focus and clarity make them ideal representations to illustrate designs or interaction techniques. In practice, creating those figures is a tedious task requiring advanced skills, even when creating the figures by tracing outlines based on photos. To mediate the process of creating trace figures, we introduce the open-source tool Esquisse (Figure 3). Informed by our taxonomy of 124 trace figures, Esquisse provides an innovative 3D model staging workflow, with specific interaction techniques that facilitate 3D staging through kinematic manipulation, anchor points and posture tracking. Our rendering algorithm (including stroboscopic rendering effects) creates vector-based trace figures of 3D scenes. We validated Esquisse with an experiment where participants created trace figures illustrating interaction techniques, and results show that participants quickly managed to use and appropriate the tool [18].

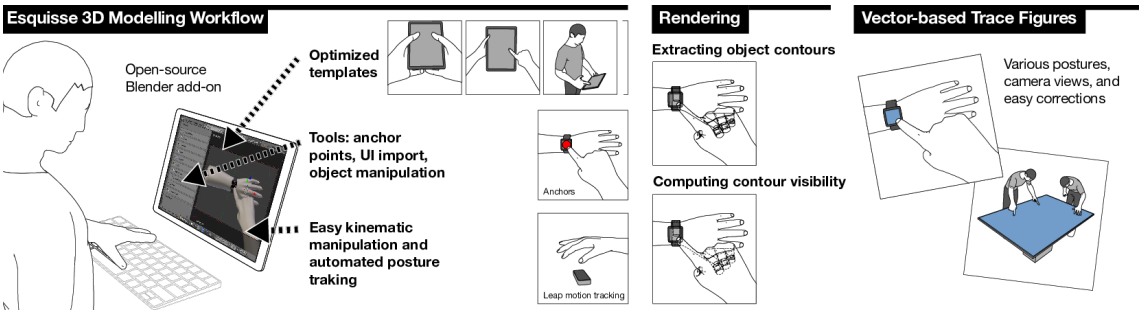


Figure 3. Workflow of Esquisse: (left) facilitating staging of 3D scenes with templates, anchor points, direct UI import, kinematic manipulation and automated posture tracking, (center) Esquisse's algorithm rendering the tracing images, and (right) final vector-based trace figures.

7.3.2. Impact of confirmation modes on expert interaction techniques adoption

Expert interaction techniques such as gestures or keyboard shortcuts are more efficient than traditional WIMP techniques because it is often faster to recall a command than to navigate to it. However, many users seem to be reluctant to switch to expert interaction. We hypothesized the cause might be the aversion to making errors. To test this, we designed two intermediate modes for the FastTap interaction technique, allowing quick confirmation of what the user has retrieved from memory, and quick adjustment if she made an error. We investigated the impact of these modes and of various error costs in a controlled study, and found that participants adopted the intermediate modes, that these modes reduced error rate when the cost of errors was high, and that they did not substantially change selection times. However, while it validates the design of our intermediate modes, we found no evidence of greater switch to memory-based interaction, suggesting that reducing error rate is not sufficient to motivate the adoption of expert use of techniques [25].

7.3.3. Effect of the context on mobile interaction

7.3.3.1. Pointing techniques for eyewear using a simulated pedestrian environment

Eyewear displays allow users to interact with virtual content displayed over real-world vision, in active situations like standing and walking. Pointing techniques for eyewear displays have been proposed, but their social acceptability, efficiency, and situation awareness remain to be assessed. Using a novel street-walking simulator, we conducted an empirical study of target acquisition while standing and walking under different levels of street crowdedness. Results showed that indirect touch was the most efficient and socially acceptable technique, and that in-air pointing was inefficient when walking. Interestingly, the eyewear displays did not improve situation awareness compared to the control condition [23].

7.3.3.2. Studying smartphone motion gestures in private or public contexts

We also investigated the effect of social exposure on smartphone motion gestures. We conducted a study where participants performed sets of motion gestures on a smartphone in both private and public locations. Using data from the smartphone's accelerometer, we found that the location had a significant effect on both the duration and intensity of the participants' gestures. We concluded that it may not be sufficient for gesture input systems to be designed and calibrated purely in private lab settings. Instead, motion gesture input systems for smartphones may need to be aware of the changing context of the device and to account for this in algorithms that interpret gestural input [26].

7.4. Macro-dynamics

Participants: Stéphane Huot, Sylvain Malacria [contact person], Nicole Pong.

7.4.1. Awareness, usage and discovery of hidden controls

Revealing a hidden widget with a dedicated sliding gesture is a common interaction design in today's handheld devices. Such "Swhidgets" (for swipe-revealed hidden widgets) provide a fast (and sometime unique) access to some commands. Interestingly, swhidgets do not follow conventional design guidelines in that they have no explicit signifiers, and users have to discover their existence before being able to use them. We conducted the first two studies specifically targeted to this type of interface design, investigating iOS users' experience with swhidgets. The first study conducted in a laboratory setting investigated which Swhidgets are spontaneously used by participants when prompted to perform certain operations on an iOS device. The second study conducted via an online survey platform, investigated which Swhidgets users reported to know and use. Combined, our studies provide the following main insights on awareness, usage and discovery of Swhidgets by middle-aged and technology-friendly users. Our results suggest that Swhidgets are moderately but unevenly known by participants, yet the awareness and the discovery issues of this design is worthy of further discussion [21].

7.5. Interaction Machine

Two of our contributions this year relate specifically to our Interaction Machine project.

7.5.1. Definition of Brain-Computer Interfaces

Regardless of the term used to designate them, Brain-Computer Interfaces are “Interfaces” between a user and a computer in the broad sense of the term. We provided a perspective to discuss how BCIs have been defined in the literature from the day the term was introduced by Jacques Vidal. From a Human-Computer Interaction perspective, we propose a new definition of Brain-Computer Interfaces as “any artificial system that transforms brain activity into input of a computer process” [24]. As they are interfaces, their definition should not include the finality and objective of the system they are used to interact with. To illustrate this, we compared BCIs with other widely used Human-Computer Interfaces, and draw analogies in their conception and purpose. This definition would help better encompassing for such interfaces in systems design, and more generally inform on how to better manage diverse forms of input in an Interaction Machine.

7.5.2. Software architecture for interactive systems

On the software engineering side, we have proposed a new Graphical User Interface (GUI) and Interaction framework based on the Entity-Component-System model (ECS) [22]. In this model, interactive elements (Entities) are characterized only by their data (Components). Behaviors are managed by continuously running processes (Systems) which select entities by the Components they possess. This model facilitates the handling of behaviors and promotes their reuse. It provides developers with a simple yet powerful composition pattern to build new interactive elements with Components. It materializes interaction devices as Entities and interaction techniques as a sequence of Systems operating on them. We have implemented these principles in the Polyphony toolkit in order to experiment the ECS model in the context of GUIs programming. It has proven to be useful and efficient for modeling standard interaction techniques, and we are now exploring its benefits for prototyping and implementing more advanced methods in a modular way. It also raises some interesting challenges about performance and scalability that we will explore further.

7.5.3. From the dynamics of interaction to an Interaction Machine

Several of our new results this year also informed our global objective of building an Interaction Machine. At the micro-dynamics level, as last year, our work on prediction algorithms and transfer functions highlighted the need for accessing low-level input data and to have flexible input management to be able to reliably predict current finger position and compensate for latency. Our work on new selection methods in 3D also highlighted the importance of easing the combination of input events from multiple sources and of data filtering to achieve better interaction. As it also leverages the real time aspect of the perception-action coupling for efficient interaction, it also confirms the need for efficient and low-latency input management stacks. These results give us the first leads to redefine input management and input events propagation in order to better account for human factors in interactive systems, and to extend the possibilities for designing more efficient and expressive interaction methods.

At the meso-dynamics level, our studies on the adoption of expert interaction techniques and of the impact of the context in performing interaction gestures highlighted the need for both adaptable and adaptive systems (e.g. context-based calibration of gesture recognition algorithms), which require more modular and flexible system architectures in order to enable real-time parametrization or even switching interaction techniques. These results also resonate with those at the micro-dynamics level, since they suggest strong links between users’ behaviors and strategies (meso) and their low-level perception mechanisms (micro) that should be better taken into account in the design of interactive systems.

These conclusions and observations will be the basis for our investigations on the topic next year. We will in particular focus on the redefinition of the input stack and on applying the ECS model to the whole architecture of an interactive system.

MANAO Project-Team

7. New Results

7.1. Analysis and Simulation

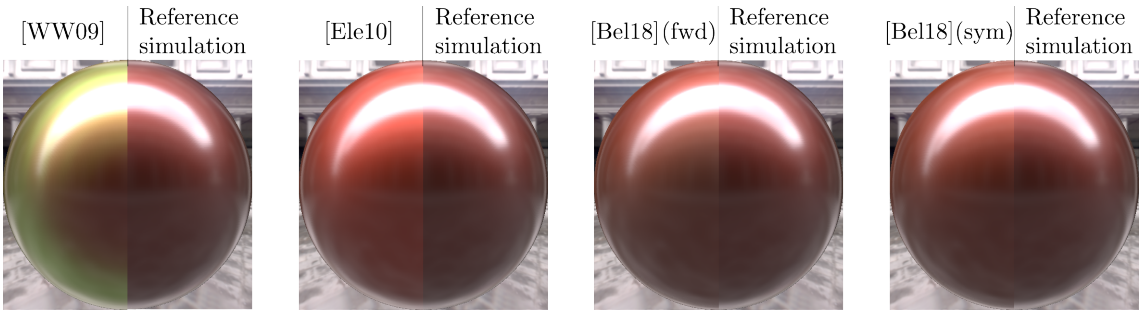


Figure 8. We study how the approximations made by layered material models impact their accuracy, and ultimately material appearance. Here we compare four models side by side with our reference simulation on a frosted metal – one of the 60 material configurations we have considered in our study. This specific choice is particularly problematic for the model of Weidlich and Wilkie [WW09], which creates oddly-colored reflections away from normal incidence. The variant of Elek [Ele10] is devoid of these artefacts, but clearly overestimates the intensity of the metallic base. Belcour's models [Bel18] (forward and symmetric) produce more accurate results, even though the intensity of the metallic base remains slightly higher. They still deviate from the reference simulation, especially at grazing angles as seen for instance at the bottom of the spheres. Our analysis in BRDF (and BTDF) space provides explanations for such departures from the reference.

7.1.1. Numerical Analysis of Layered Materials Models ,

Publications: [12], [14]

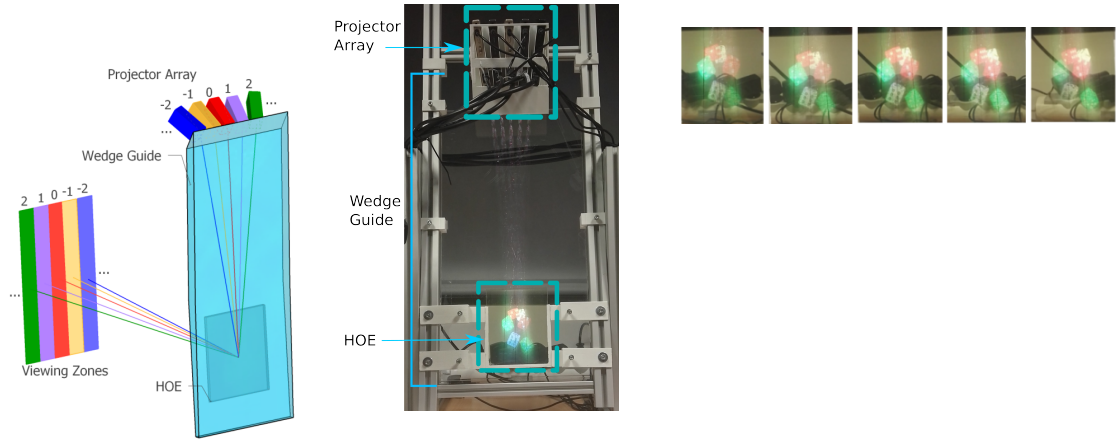
Most real-world materials are composed of multiple layers, whose physical properties impact the appearance of objects. The accurate reproduction of layered material properties is thus an important part of physically-based rendering applications. Since no exact analytical model exists for arbitrary configurations of layer stacks, available models make a number of approximations. In this technical report, we propose to evaluate these approximations with a numerical approach: we simulate BRDFs and BTDFs for layered materials in order to compare existing models against a common reference. More specifically, we consider 60 layered material configurations organized in three categories: plastics, metals and transparent slabs. Our results (see Figure 8) show that: (1) no single model systematically outperforms the others on all categories; and (2) significant discrepancies remain between simulated and modeled materials. We analyse the reasons for these discrepancies and introduce immediate corrections that improve models accuracy with little effort. Finally, we provide a few challenging cases for future layered material models.

7.1.2. A systematic approach to testing and predicting light-material interactions

Publication: [11]

Photographers and lighting designers set up lighting environments that best depict objects and human figures to convey key aspects of the visual appearance of various materials, following rules drawn from experience. Understanding which lighting environment is best adapted to convey which key aspects of materials is an important question in the field of human vision. The endless range of natural materials and lighting environments poses a major problem in this respect. Here we present a systematic approach to make this problem tractable for lighting–material interactions, using optics-based models composed of canonical lighting and material modes. In two psychophysical experiments, different groups of inexperienced observers judged the material qualities of the objects depicted in the stimulus images. In the first experiment, we took photographs of real objects as stimuli under canonical lightings. In a second experiment, we selected three generic natural lighting environments on the basis of their predicted lighting effects and made computer renderings of the objects. The selected natural lighting environments have characteristics similar to the canonical lightings, as computed using a spherical harmonic analysis. Results from the two experiments correlate strongly, showing (a) how canonical material and lighting modes associate with perceived material qualities; and (b) which lighting is best adapted to evoke perceived material qualities, such as softness, smoothness, and glossiness. Our results demonstrate that a system of canonical modes spanning the natural range of lighting and materials provides a good basis to study lighting–material interactions in their full natural ecology.

7.2. From Acquisition to Display



(a) Illustration (b) Photograph (c) 5 views

Figure 9. The autostereoscopic transparent display: light beams from multiple laser beam steering picoprojectors are coupled into a transparent wedge guide, and then the light from each projector is redirected to separate viewing zones using a transparent HOE.

7.2.1. Autostereoscopic transparent display using a wedge light guide and a holographic optical element

Publications: [8], [13]

We designed and developed a novel transparent autostereoscopic display consisting of laser picoprojectors, a wedge light guide, and a custom holographic optical element (HOE). Such a display can superimpose 3D data on the real world without any wearable.

The principle of our display, as depicted in Figure 9 , is to couple beams from multiple laser beam steering picoprojectors into a transparent wedge guide and then to redirect each beam to separate viewing zones using a transparent HOE. The HOE is wavelength-multiplexed for full-color efficiency, but only one angular grating is recorded and multiple viewing zones are reconstructed with several projector positions due to the high angular bandwidth. Our current prototype has 5 views but is theoretically able to generate 9 views. The views are located 50cm in front of the display, they are 3cm wide and 10cm high. These values are fixed once the HOE is recorded; they result from our choices and can be changed in the recording step.

This display has great potential for augmented reality applications such as augmented exhibitions in museums or shops, head-up displays for vehicles or aeronautics, and industrial maintenance, among others.

7.2.2. Wedge cameras for minimally invasive archaeology

Publication: [9]

Acquiring images of archaeological artifacts is an essential step for the study and preservation of cultural heritage. In constrained environments, traditional acquisition techniques may fail or be too invasive. We present an optical device including a camera and a wedge waveguide that is optimized for imaging within confined spaces in archeology. The major idea is to redirect light by total internal reflection to circumvent the lack of room, and to compute the final image from the raw data. We tested various applications onsite during an archaeological mission in Medamoud (Egypt). Our device was able to successfully record images of the underground from slim trenches, including underwater trenches, and between rocks composing a wall temple. Experts agreed that the acquired images were good enough to get useful information that cannot be obtained as easily with traditional techniques.

7.2.3. Study of contrast variations with depth in focused plenoptic cameras

Publication: [10]

A focused plenoptic camera has the ability to record and separate spatial and directional information of the incoming light. Combined with the appropriate algorithm, a 3D scene could be reconstructed from a single acquisition, over a depth range called plenoptic depth-of-field. We have studied the contrast variations with depth as a way to assess plenoptic depth-of-field. We take into account the impact of diffraction, defocus, and magnification on the resulting contrast. We measure the contrast directly on both simulated and acquired images. We demonstrate the importance of diffraction and magnification in the final contrast. Contrary to classical optics, the maximum of contrast is not centered around the main object plane, but around a shifted position, with a fast and nonsymmetric decrease of contrast.

7.2.4. Unifying the refocusing algorithms and parameterizations for traditional and focused plenoptic cameras

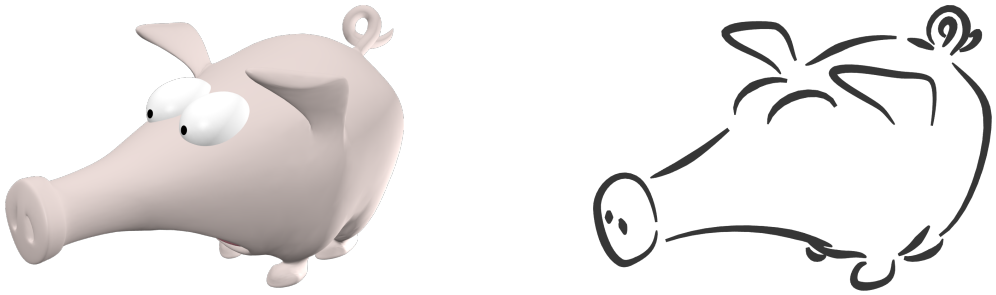
Publication: [16]

We propose a unique parameterization of the light rays in a plenoptic setup, allowing the development of a unique refocusing algorithm valid for any plenoptic configurations, based on this parameterization. With this method we aim at refocusing images at any distances from the camera, without previous discontinuity due to change of optical configuration. We aim to obtain reconstructed images visually similar to the results of the other algorithms, but quantitatively more accurate.

7.3. Rendering, Visualization and Illustration

7.3.1. Line drawings from 3D models: a tutorial

Publication: [7]



(a) 3D object with diffuse shading (b) Stylized curves

Figure 10. The occluding contours of the 3D model "Origins of the Pig" by Keenan Crane, shown in (a) with diffuse shading, are depicted in (b) with calligraphic brush strokes.

This tutorial describes the geometry and algorithms for generating line drawings from 3D models, focusing on occluding contours. The geometry of occluding contours on meshes and on smooth surfaces is described in detail, together with algorithms for extracting contours, computing their visibility, and creating stylized renderings and animations. Exact methods and hardware-accelerated fast methods are both described, and the trade-offs between different methods are discussed. The tutorial brings together and organizes material that, at present, is scattered throughout the literature. It also includes some novel explanations, and implementation tips. A thorough survey of the field of non-photorealistic 3D rendering is also included, covering other kinds of line drawings and artistic shading (Figure 10). In addition, we provide an interactive viewer at https://benardp.github.io/contours_viewer/.

7.4. Editing and Modeling

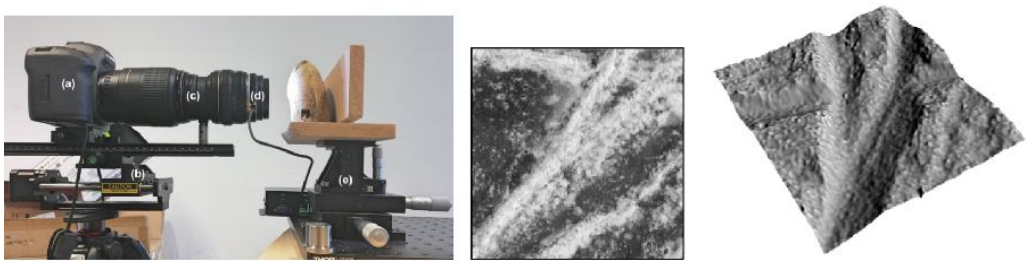


Figure 11. Left: our low-cost depth-from-focus acquisition setup. Right: result of our reconstruction algorithm for the sample shown in the middle image. The width of grooves are about 500 micrometers.

7.4.1. Depth from focus stacks at micrometer scale

In this work we designed a low-cost acquisition setup and a new algorithm for the digitalization of micro reliefs. The setup is based on a common digital camera equipped with a special assembly of different lenses designed to enable a $\times 2$ magnification factor with a very shallow depth of field (fig. 11). A micro-metric motor rail allows us to acquire dense focus stacks with depth information that can be reconstructed through image processing and analysis techniques. To enhance the accuracy of this reconstruction step, we designed

novel focus estimators as well as novel focus-point analysis algorithms exploiting novel 3D invariants. Our initial results show that we are able to reconstruct depth maps with sub-step length accuracy.

MAVERICK Project-Team

6. New Results

6.1. Texture synthesis

6.1.1. Procedural Phasor Noise

Participants: Thibault Tricard, Semyon Efremov, Cédric Zanni, Fabrice Neyret, Jonàs Martínez, Sylvain Lefebvre.

Procedural pattern synthesis is a fundamental tool of Computer Graphics, ubiquitous in games and special effects. By calling a single procedure in every pixel – or voxel – large quantities of details are generated at low cost, enhancing textures, producing complex structures within and along surfaces. Such procedures are typically implemented as pixel shaders. We propose a novel procedural pattern synthesis technique that exhibits desirable properties for modeling highly contrasted patterns, that are especially well suited to produce surface and microstructure details. In particular, our synthesizer affords for a precise control over the profile, orientation and distribution of the produced stochastic patterns, while allowing to grade all these parameters spatially. Our technique defines a stochastic smooth phase field – a phasor noise – that is then fed into a periodic function (e.g. a sine wave), producing an oscillating field with prescribed main frequencies and preserved contrast oscillations. In addition, the profile of each oscillation is directly controllable as shown Figure 2 . Our technique builds upon a reformulation of Gabor noise in terms of a phasor field that affords for a clear separation between local intensity and phase. Applications range from texturing to modeling surface displacements, as well as multi-material microstructures in the context of additive manufacturing.

This paper was published in ACM TOG [6] and presented at Siggraph 2019.

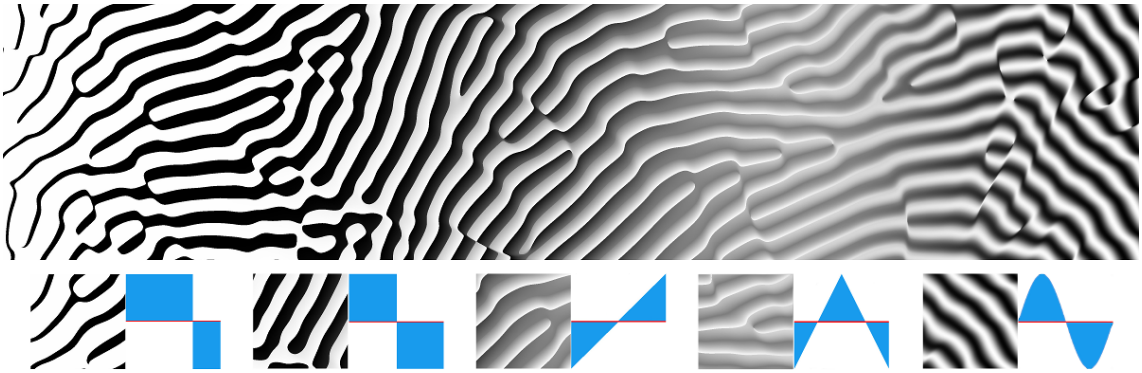


Figure 2. High-contrast patterns produced by our approach. Note how the profile of the oscillations smoothly transition from a rectangular wave (20% black), to a square wave, to a triangular profile and finally a sine wave. At the same time, the orientation of the waves changes from left to right. The field visualized here is purely procedural. It is obtained by feeding our phasor noise into periodic profile functions (shown in blue), that are interpolated from left to right.

6.1.2. Making Gabor Noise Fast and Normalized

Participants: Vincent Tavernier, Fabrice Neyret, Romain Vergne, Joëlle Thollot.

Gabor Noise is a powerful procedural texture synthesis technique, but it has two major drawbacks: It is costly due to the high required splat density and not always predictable because properties of instances can differ from those of the process. We bench performance and quality using alternatives for each Gabor Noise ingredient: point distribution, kernel weighting and kernel shape. For this, we introduce 3 objective criteria to measure process convergence, process stationarity, and instance stationarity. We show that minor implementation changes allow for 17-24× speed-up with same or better quality.

This paper has been presented at Eurographics-short 2019 [11].

6.2. Illumination simulation and materials

6.2.1. Harmonic Analysis of the Light Transport Operator

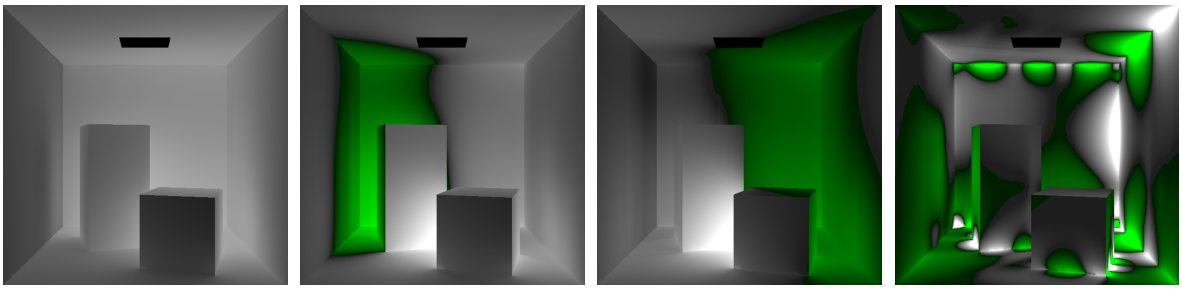
Participants: Ronak Molazem, Cyril Soler.

In this work we study the eigenvalues and eigenfunctions of the light transport operator. While computing the spectrum of the light transport operator is a simple task in Lambertian scenes by applying a traditional eigensolver to the linear system obtained from discretized geometry, it becomes a real challenge in general environments where discretizing the geometry is not possible anymore. “Diagonalizing” light transport however can be a very effective way to perform re-lighting and rapidely compute light transport solutions.

In this work we propose an analysis of the properties of the spectrum of the light transport operator, connecting the calculation of eigenvalues to resolvent theory. We show that the eigenfunctions are generally not orthogonal nor positive, but they can still be used to efficiently represent light distributions.

We analyse the performance of different methods to compute eigenvalues and images of their eigenfunctions using path tracing. We prove in particular that it is possible to compute the eigenfunctions of the light transport operator by integrating "circular" light paths of various lengths accross the scene.

This work is part of the PhD of Ronak Molazem and is funded by the ANR project "CaLiTrOp". At the time of writing this (Dec. 2019), we’re about to submit a paper to ACM Transactions on Graphics.



Λ_0 Λ_1 Λ_2 Λ_{144}

Figure 3. Path-traced images of four eigenfunctions of the light transport operator in the Cornell Box. A green scale is used to represent negative values.

6.2.2. Low Dimension Approximations of Light Transport

Participants: Ronak Molazem, Cyril Soler.

Light transport is known to be a low rank linear operator: the vector space formed by solutions of a light transport problem for different initial conditions is of low dimension. Approximating this space using appropriate bases is therefore of primordial help to efficiently compute solutions to light transport problems.

In this work, we're interested into generating such approximations using *ad-hoc* methods that rely on deep learning. The goal is to be able to efficiently generate a sensible basis for light transport solutions on which we can efficiently project a noisy image. Other applications of this work include relighting pictures, in which an approximate geometry is used to project the illumination in the image, that can further be manipulated while staying in the space of expected light transport solutions.

This work is an ongoing collaboration with Unity Research Grenoble, and part of the PhD of Ronak Molazem, currently in her second year of PhD, and is funded by the ANR project "CaLiTrOp".

6.2.3. Precomputed Multiple Scattering for Rapid Light Simulation in Participating Media

Participants: Nicolas Holzschuch, Liangsheng Ge, Beibei Wang.

Rendering translucent materials is costly: light transport algorithms need to simulate a large number of scattering events inside the material before reaching convergence. The cost is especially high for materials with a large albedo or a small mean-free-path, where higher-order scattering effects dominate. In [7], we present a new method for fast computation of global illumination with participating media. Our method uses precomputed multiple scattering effects, stored in two compact tables. These precomputed multiple scattering tables are easy to integrate with any illumination simulation algorithm. We give examples for virtual ray lights (VRL), photon mapping with beams and paths (UPBP), Metropolis Light Transport with Manifold Exploration (MEMLT). The original algorithms are in charge of low-order scattering, combined with multiple scattering computed using our table. Our results show significant improvements in convergence speed and memory costs, with negligible impact on accuracy.

6.2.4. Fast Computation of Single Scattering in Participating Media with Refractive Boundaries using Frequency Analysis

Participants: Nicolas Holzschuch, Yulin Liang, Lu Wang, Beibei Wang.

Many materials combine a refractive boundary and a participating media on the interior. If the material has a low opacity, single scattering effects dominate in its appearance. Refraction at the boundary concentrates the incoming light, resulting in an important phenomenon called volume caustics. This phenomenon is hard to simulate. Previous methods used point-based light transport, but attributed point samples inefficiently, resulting in long computation time. In [3], we use frequency analysis of light transport to allocate point samples efficiently. Our method works in two steps: in the first step, we compute volume samples along with their covariance matrices, encoding the illumination frequency content in a compact way. In the rendering step, we use the covariance matrices to compute the kernel size for each volume sample: small kernel for high-frequency single scattering, large kernel for lower frequencies. Our algorithm computes volume caustics with fewer volume samples, with no loss of quality. Our method is both faster and uses less memory than the original method. It is roughly twice as fast and uses one fifth of the memory. The extra cost of computing covariance matrices for frequency information is negligible.

6.2.5. Reparameterizing discontinuous integrands for differentiable rendering

Participants: Nicolas Holzschuch, Wenzel Jakob, Guillaume Loubet.

Differentiable rendering has recently opened the door to a number of challenging inverse problems involving photorealistic images, such as computational material design and scattering-aware reconstruction of geometry and materials from photographs. Differentiable rendering algorithms strive to estimate partial derivatives of pixels in a rendered image with respect to scene parameters, which is difficult because visibility changes are inherently non-differentiable.

We propose [5] a new technique for differentiating path-traced images with respect to scene parameters that affect visibility, including the position of cameras, light sources, and vertices in triangle meshes. Our algorithm computes the gradients of illumination integrals by applying changes of variables that remove or strongly reduce the dependence of the position of discontinuities on differentiable scene parameters. The underlying parameterization is created on the fly for each integral and enables accurate gradient estimates using standard Monte Carlo sampling in conjunction with automatic differentiation. Importantly, our approach does not rely

on sampling silhouette edges, which has been a bottleneck in previous work and tends to produce high-variance gradients when important edges are found with insufficient probability in scenes with complex visibility and high-resolution geometry. We show that our method only requires a few samples to produce gradients with low bias and variance for challenging cases such as glossy reflections and shadows. Finally, we use our differentiable path tracer to reconstruct the 3D geometry and materials of several real-world objects from a set of reference photographs.

6.3. Expressive rendering

6.3.1. Procedural Stylization

Participants: Maxime Isnel, Mohamed Amine Farhat, Romain Vergne, Joëlle Thollot.

Stylizing 3D scenes is a long term goal for the expressive rendering community. During the master internship of Maxime Isnel we have worked on a procedurel approach based on a procedural solid noise used in image space to generate brush strokes or 2.5D visual primitives, such as fur. The overview of the approach is shown Figure 4 . This project is still in progress and will continue with a post-doc in 2020.

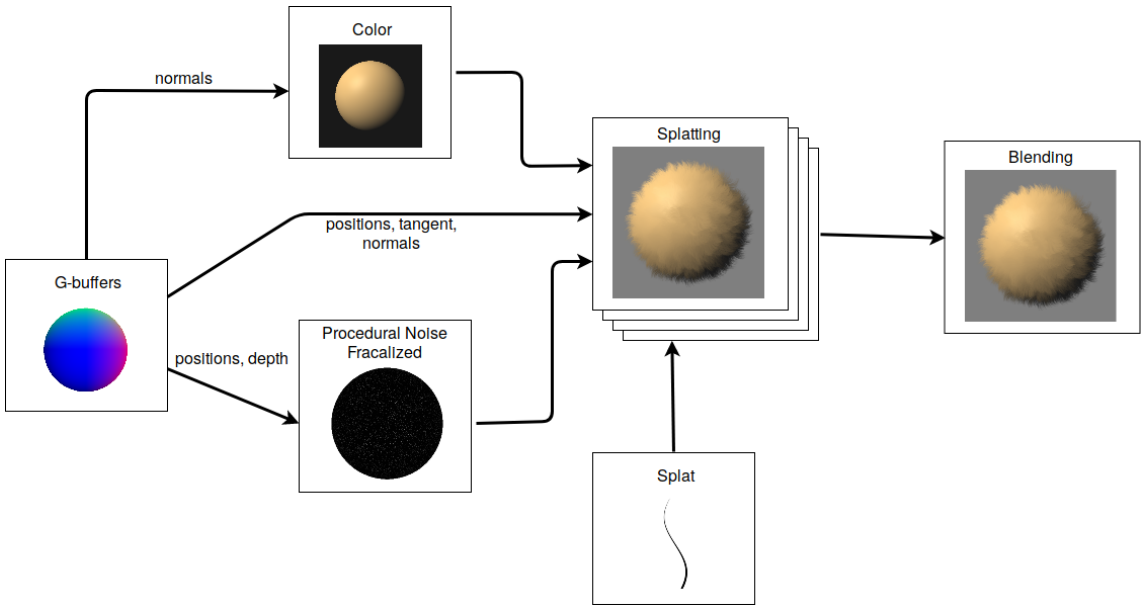


Figure 4. Based on a procedural solid noise and the use of geometry buffers, we propose an image-space approach to stylize a 3D object on the GPU.

MFX Project-Team

7. New Results

7.1. Star-shaped Metrics for Mechanical Metamaterial Design

Participants: Jonàs Martínez, Mélina Skouras, Christian Schumacher, Samuel Hornus, Sylvain Lefebvre, Bernhard Thomaszewski.

Digital manufacturing technologies such as 3D printing and laser cutting enable us to fabricate designs with great geometric detail. One particular way of exploiting this capability is to create patterned sheet materials whose geometric structures can be tailored to control their macro-mechanical behavior.

A typical approach to model and analyze structured sheet materials is centered around the concept of a representative element—a tile—which is repeated, transformed, and laid out so as to generate a regular spatial tiling. Changing the shape of the representative tile allows to control macro-mechanical properties such as isotropy or negative Poisson's ratios. Generalizing this material design principle from a single representative tile to *families* of tiles that can be combined in a spatially-varying manner opens the door to structures with progressively-graded material properties.

At SIGGRAPH 2019 we have presented a method for designing mechanical metamaterials [14]. It is based on the novel concept of Voronoi diagrams induced by star-shaped metrics. As one of its central advantages, our approach supports interpolation between arbitrary metrics (see Figure 1). This capability opens up a rich space of tile geometries with interesting aesthetics and a wide range of mechanical properties. They include isotropic, tetragonal, orthotropic, as well as smoothly graded materials. We have validated the mechanical properties predicted by simulation through tensile tests on a set of physical prototypes. An open source C++ implementation of the technique can be found at https://github.com/mfx-inria/starshaped2d

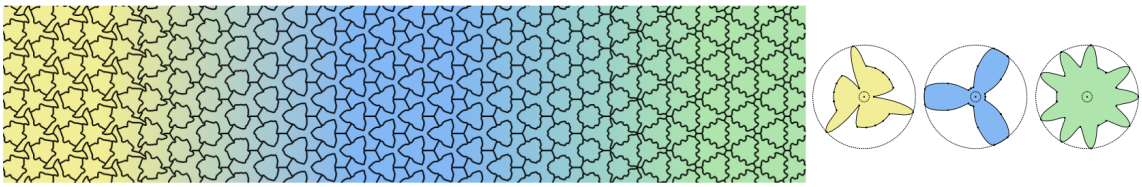


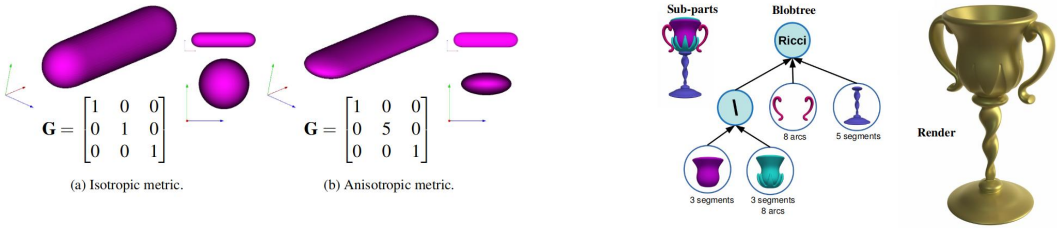
Figure 1. Our method generates a smoothly-graded pattern (left) when interpolating between three star-shaped distance functions (regular) on a regular honeycomb lattice. Each distance function is compactly parameterized with polar coordinates, allowing for simple interpolation in metric space as indicated by color-coding.

7.2. Anisotropic convolution surfaces

Participants: Alvaro Javier Fuentes Suárez, Evelyne Hubert, Cédric Zanni.

Skeletons, as a set of curves and/or surfaces centered inside a shape, provide a compact representation of the shape structure. Due to this property, skeletons have proved useful in many applications that range from shape analysis to 3D modeling and deformation. Convolution surfaces associate radii information to the skeleton and provide a simple way for users to rapidly define a shape. A convolution surface is an implicit surface defined as a level set of a scalar field, the convolution field, that is obtained by integrating a kernel function over the skeleton. This technique allows us to build a complex shape by modeling parts that combine into a smooth surface, independently of the smoothness of the skeleton itself. They also represent a volume with the convolution surface as its boundary and can therefore be combined with other composition operators from implicit modeling frameworks.

We have introduced anisotropic convolution surfaces [12], an extension that increases the modeling freedom by providing ellipse-like normal sections around 1D skeletons. It increases the diversity of shapes that can be generated from 1D skeletons, and reduces the need for 2D skeletons, while it still retains smoothness. We achieve anisotropy not just in the normal sections but also in the tangential direction. This allows sharper and steeper radius variation, and the control of thickness at skeleton endpoints (see Figure 2). The method is applied to skeletons represented by bi-arcs. It allows us to control precisely the orientation of anisotropy thanks to rotation minimizing frames. This work was presented at Shape Modeling International 2019.



(a) Isotropic metric. (b) Anisotropic metric.

Figure 2. Our method, based on an anisotropic metric, allows us to generate an ellipse-like cross section around 1D skeletons (segments, bi-arcs). The thickness around the skeleton can be controlled precisely both in the orthogonal cross-section and in the tangential direction giving finer control to the user. The density field generated can then be used in a classical implicit modeling framework.

7.3. Procedural Phasor Noise

Participants: Thibault Tricard, Semyon Efremov, Cédric Zanni, Fabrice Neyret, Jonàs Martínez, Sylvain Lefebvre.

Procedural pattern synthesis is a fundamental tool of Computer Graphics. In 2019 we introduced a new formulation that generates a wide range of patterns that could not be produced by other techniques. Our procedural *phasor noise* is based on a prior technique called Gabor noise, which creates oscillating patterns with accurate control over their frequency content (power spectrum). Gabor noise achieves this by summing a large number of Gabor kernels — Gaussian weighted sinewaves — distributed pseudo-randomly in space. Unfortunately Gabor noise suffers from local loss of contrast and lacks control over the shape of the oscillations (which always have a sinewave profile).

Our method solves these limitations by reformulating Gabor noise to expose its instantaneous phase. Once the phase obtained we can directly remap a periodic profile function onto it, to obtain an oscillating pattern of constant contrast and controlled profile geometry, while retaining all desirable properties of Gabor noise (see Figure 3). This unlocks two main applications. The first is in texture synthesis for computer graphics, to generate color, displacement and normal fields. The second is in additive manufacturing, where our method can be applied in 3D to generate a wide range of microstructures.

This work was done in collaboration with Fabrice Neyret (MAVERICK, Inria) and has been published in ACM Transactions on Graphics, in 2019 [17]. Thibault Tricard and Semyon Efremov did a joint presentation at ACM SIGGRAPH 2019.

7.4. Ribbed support vaults for 3D printing of hollowed objects

Participants: Thibault Tricard, Frédéric Claux, Sylvain Lefebvre.

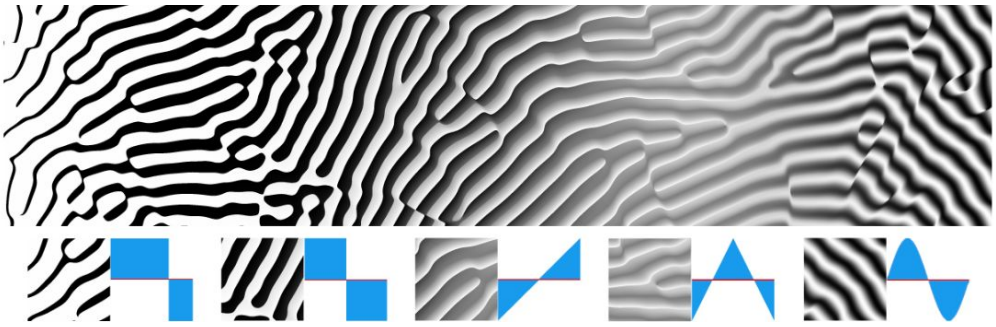


Figure 3. Phasor noise is a novel procedural function generating strongly oriented, coherent stripe patterns. The profiles of the oscillations are controlled (here: square, triangular, sine).

In additive manufacturing, and in particular with the popular filament-based fabrication, the printing time remains a major constraint. In a typical print, most of the time is spent filling the interior of the object. Based on this observation we explored how to print an object as empty as possible. The difficulty is that any deposited material has to be supported from below to prevent the object from collapsing.

We developed a simple, yet very efficient, algorithm that generates a lightweight ribbed support vault infill (see Figure 4). Our algorithm sweeps once through the slices from top to bottom, detects non-supported points, and connects them with a segment to the closest already supported points. The endpoints of open segments are eroded from one slice to the next. This process generates hierarchical ribbed support vaults that quickly retract and merge with the enclosing walls, while offering printability guarantees.

Our approach greatly reduces material usage (reaching parts as empty as 98%) and thus strongly reduces print time. Nevertheless it guarantees printability, and scales to very large inputs.

This work originated from the University of Limoges and was the master topic of Thibault Tricard, under the supervision of Frédéric Claux and in collaboration with Sylvain Lefebvre. The work was published in Computer Graphics Forum in June 2019 [16].

Figure 4. A 3D bunny model printed with our internal ribbed supports. It is mostly empty, with the ribbed vaults providing just enough support to prevent filament to fall during manufacturing.

7.5. CurviSlicer: Slightly curved slicing for 3-axis printers

Participants: Jimmy Étienne, Nicolas Ray, Daniele Panozzo, Samuel Hornus, Charlie C.l. Wang, Jonàs Martínez, Sara Mcmains, Marc Alexa, Brian Wyvill, Sylvain Lefebvre.

Most additive manufacturing processes fabricate objects by stacking planar layers of solidified material. As a result, produced parts exhibit a so-called staircase effect, which results from sampling slanted surfaces with parallel planes. Using thinner slices reduces this effect, but it always remains visible where layers almost align with the input surfaces. In this research we exploit the ability of some additive manufacturing processes to deposit material slightly out of plane to dramatically reduce these artifacts.

We focused in particular on the widespread Fused Filament Fabrication (FFF) technology, since most printers in this category can deposit along slightly curved paths, under deposition slope and thickness constraints. Our algorithm curves the layers, making them either follow the natural slope of the input surface or on the contrary, make them intersect the surfaces at a steeper angle thereby improving the sampling quality.

Rather than directly computing curved layers, our algorithm deforms the input model before slicing it with a standard planar approach. The deformation is optimized for reproduction accuracy. We demonstrate that this approach enables us to encode all fabrication constraints, including the guarantee of generating collision-free toolpaths, in a convex optimization that can be solved using a QP solver.

This work emerged from a problem solving session between its co-authors at the 17th international Bellairs Workshop on Computational Geometry (2018). It was then pursued during 2019 in the context of the PhD thesis of Jimmy Étienne and as a collaboration with Nicolas Ray (PIXEL, Inria). The work was published in ACM Transactions on Graphics in 2019 [11] and presented at ACM SIGGRAPH 2019 by Jimmy Étienne.

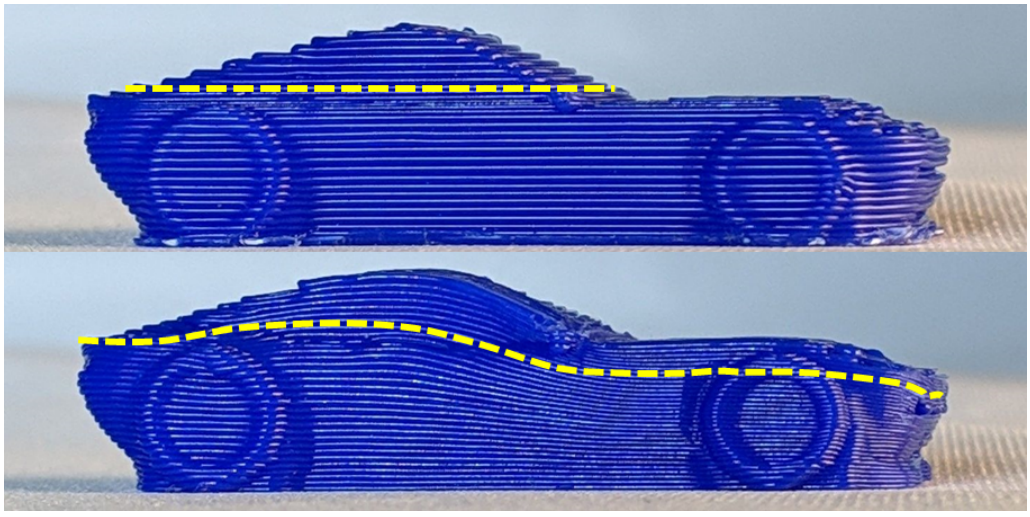


Figure 5. A 3D model printed with our technique. The algorithm automatically generates curved slices (right) to better reproduce the slanted surfaces, removing the staircase defect created by standard planar layers (top-left: standard, bottom-left: curved).

7.6. Extrusion-Based Ceramics Printing with Strictly-Continuous Deposition

Participants: Jean Hergel, Kevin Hinz, Bernhard Thomaszewski, Sylvain Lefebvre.

3D printing with extruded ceramic paste induces constraints that deviate significantly from standard thermoplastic materials. In particular existing path generation methods for thermoplastic materials rely on transfer moves to navigate between different print paths in a given layer. However, when printing with clay, these transfer moves can lead to severe artifacts and failure.

We explored how to eliminate transfer moves altogether by generating deposition paths that are continuous within and across layers. In each layer, we optimize a continuous support path with respect to length, smoothness, and distance to the model. Comparisons to existing path generation methods designed for thermoplastic materials show that our method substantially improves print quality and often makes the difference between success and failure.

This work was primarily done at the University of Montréal in collaboration with Sylvain Lefebvre. It was published in ACM Transactions on Graphics 2019 [13], and presented at SIGGRAPH Asia 2019 by Jean Hergel.

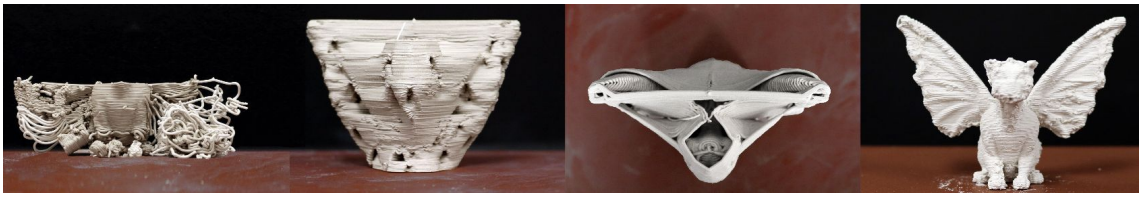


Figure 6. Our technique greatly improves the reliability of 3D printing with extruded clay.

MIMETIC Project-Team

7. New Results

7.1. Outline

In 2019, MimeTIC has maintained his activity in motion analysis, modelling and simulation, to support the idea that these approaches are strongly coupled in a motion analysis-synthesis loop. This idea has been applied to the main application domains of MimeTIC:

- Animation, autonomous characters and Digital Storytelling,
- Fidelity of Virtual Reality,
- Motion sensing of Human Activity,
- Sports,
- Ergonomics,
- and Locomotion and Interactions between walkers.

7.2. Animation, Autonomous Characters and Digital Storytelling

MimeTIC main research path consists in associating motion analysis and synthesis to enhance the naturalness in computer animation, with applications in movie previsualisation, and autonomous virtual character control. Thus, we pushed example-based techniques in order to reach a good tradeoff between simulation efficiency and naturalness of the results. In 2019, to achieve this goal, MimeTIC continued to explore the use of perceptual studies and model-based approaches, but also began to investigate deep learning, for example to control cameras in Movie previsualization.

7.2.1. VR as a Content Creation Tool for Movie Previsualisation

Participants: Marc Christie [contact], Quentin Galvane.

This work proposes a VR authoring system which provides intuitive ways of crafting visual sequences in 3D environments, both for expert animators and expert creatives. It is designed in mind to be applied animation and film industries, but can find broader applications (eg. in multimedia content creation). Creatives in animation and film productions have forever been exploring the use of new means to prototype their visual sequences before realizing them, by relying on hand-drawn storyboards, physical mockups or more recently 3D modelling and animation tools. However these 3D tools are designed in mind for dedicated animators rather than creatives such as film directors or directors of photography and remain complex to control and master. The proposed system is designed to reflect the traditional process through (i) a storyboarding mode that enables rapid creation of annotated still images, (ii) a previsualisation mode that enables the animation of the characters, objects and cameras, and (iii) a technical mode that enables the placement and animation of complex camera rigs (such as cameras cranes) and light rigs. Our methodology strongly relies on the benefits of VR manipulations to re-think how content creation can be performed in this specific context, typically how to animate contents in space and time. As a result, the proposed system is complimentary to existing tools, and provides a seamless back-and-forth process between all stages of previsualisation. We evaluated the tool with professional users to gather experts' perspectives on the specific benefits of VR in 3D content creation [36].

7.2.2. Deep Learning Techniques for Camera Trajectories

Participant: Marc Christie [contact].

Designing a camera motion controller which places and moves virtual cameras in relation with contents in a cinematographic way is a complex and challenging task. Many cinematographic rules exist, yet practice shows there are significant stylistic variations in how these can be applied. While contributions have attempted to encode rules by hand, this work is the very first to propose an end-to-end framework that automatically learns from real and synthetic movie sequences how the camera behaves in relation with contents. Our deep-learning framework extracts cinematic features of movies through a novel feature estimator trained on synthetic data, and learns camera behaviors from those extracted features, through the design of a Recurrent Neural Network (RNN) with a Mixture of Experts (MoE) gating mechanism. This cascaded network is designed to capture important variations in camera behaviors while ensuring the generalization capacity in the learning of similar behaviors. We demonstrate the features of our framework through experiments that highlight (i) the quality of our cinematic feature extractor (ii) the capacity to learn ranges of behaviors through the gating mechanism, and (iii) the ability to analyse the camera behaviors from a given input sequence, and automatically re-apply these behaviors on new virtual contents, offering exciting new possibilities towards a deeper understanding of cinematographic style and enhanced possibilities in transferring style from real to virtual. The work is a collaboration with the Beijing Film Academy in China.

7.2.3. Efficient Visibility Computation for Camera Control

Participants: Marc Christie [contact], Ludovic Burg.

Efficient visibility computation is a prominent requirement when designing automated camera control techniques for dynamic 3D environments; computer games, interactive storytelling or 3D media applications all need to track 3D entities while ensuring their visibility and delivering a smooth cinematographic experience. Addressing this problem requires to sample a very large set of potential camera positions and estimate visibility for each of them, which in practice is intractable. In this work, we introduce a novel technique to perform efficient visibility computation and anticipate occlusions. We first propose a GPU-rendering technique to sample visibility in Toric Space coordinates – a parametric space designed for camera control. We then rely on this visibility evaluation to compute an anticipation map which predicts the future visibility of a large set of cameras over a specified number of frames. We finally design a camera motion strategy that exploits this anticipation map to maximize the visibility of entities over time. The key features of our approach are demonstrated through comparison with classical ray-casting techniques on benchmark environments, and through an integration in multiple game-like 3D environment with heavy sparse and dense occluders.

7.2.4. Analysing and Predicting Inter-Observer Gaze Congruency

Participant: Marc Christie [contact].

In trying to better understand film media, we have been recently exploring the relation between the distribution of gaze states and the features of images, with the objective of establishing correlations to understand how films manipulate users gaze (and how gaze can be manipulated be re-editing film sequences). According to the literature regarding visual saliency, observers may exhibit considerable variations in their gaze behaviors. These variations are influenced by aspects such as cultural background, age or prior experiences, but also by features in the observed images. The dispersion between the gaze of different observers looking at the same image is commonly referred as inter-observer congruency (IOC). Predicting this congruence can be of great interest when it comes to study the visual perception of an image. We introduce a new method based on deep learning techniques to predict the IOC of an image [31]. This is achieved by first extracting features from an image through a deep convolutional network. We then show that using such features to train a model with a shallow network regression technique significantly improves the precision of the prediction over existing approaches.

7.2.5. Deep Saliency Models: the Quest for the Loss Function

Participant: Marc Christie [contact].

Following our idea of understanding gaze patterns in movie watching, and predicting these gaze patterns on sequences, we have been exploring the influence of loss functions in learning the visual saliency. Indeed, numerous models in the literature present new ways to design neural networks, to arrange gaze pattern data, or to extract as much high and low-level image features as possible in order to create the best saliency representation. However, one key part of a typical deep learning model is often neglected: the choice of the loss function. In this work, we explore some of the most popular loss functions that are used in deep saliency models [49]. We demonstrate that on a fixed network architecture, modifying the loss function can significantly improve (or depreciate) the results, hence emphasizing the importance of the choice of the loss function when designing a model. We also introduce new loss functions that have never been used for saliency prediction to our knowledge. And finally, we show that a linear combination of several well-chosen loss functions leads to significant improvements in performances on different datasets as well as on a different network architecture, hence demonstrating the robustness of a combined metric.

7.2.6. Contact Preserving Shape Transfer For Rigging-Free Motion Retargeting

Participants: Franck Multon [contact], Jean Basset.

In 2018, we introduced the idea of context graph to capture the relationship between body parts surfaces and enhance the quality of the motion retargetting problem. Hence, it becomes possible to retarget the motion of a source character to a target one while preserving the topological relationship between body parts surfaces. However this approach implies to strictly satisfy distance constraints between body parts, whereas some of them could be relaxed to preserve naturalness. In 2019, we introduced a new paradigm based on transfering the shape instead of encoding the pose constraints to tackle this problem [29].

Hence, retargeting a motion from a source to a target character is an important problem in computer animation, as it allows to reuse existing rigged databases or transfer motion capture to virtual characters. Surface based pose transfer is a promising approach to avoid the trial-and-error process when controlling the joint angles. The main contribution of this work is to investigate whether shape transfer instead of pose transfer would better preserve the original contextual meaning of the source pose. To this end, we propose an optimization-based method to deform the source shape+pose using three main energy functions: similarity to the target shape, body part volume preservation, and collision management (preserve existing contacts and prevent penetrations). The results show that our method is able to retarget complex poses, including several contacts, to very different morphologies. In particular, we introduce new contacts that are linked to the change in morphology, and which would be difficult to obtain with previous works based on pose transfer that aim at distance preservation between body parts. These preliminary results are encouraging and open several perspectives, such as decreasing computation time, and better understanding how to model pose and shape constraints.

7.2.7. The Influence of Step Length to Step Frequency Ratio on the Perception of Virtual Walking Motions

Participants: Ludovic Hoyet [contact], Benjamin Niay, Anne-Hélène Olivier.

Synthesizing walking motions that look realistic and diverse is a challenging task in animation, and even more when the target is to create realistic motions for large group of characters. Indeed, in order to keep a good trade-off between computational costs and realism, biomechanical constraints of human walk are not always fulfilled. In pilot experiments [38], [46], we have therefore started to investigate the ability of viewers to identify an invariant parameter of human walking named the walk ratio, representing the ratio between step length and step frequency of an individual, when applied to virtual humans. To this end, we recorded 4 actors (2 males, 2 females) walking at different freely chosen speeds, as well as at different combinations of step frequency and step length. We then performed pilot perceptual studies to identify the ability of viewers to detect the range of walk ratios considered as natural and compared it to the walk ratio freely chosen by the actor when performing walks at the same speeds. Our results will provide new considerations to drive the animation of walking virtual characters using the walk ratio as a parameter, which we believe could enable animators to control the speed of characters through simple parameters while retaining the naturalness of the locomotion.

7.3. Fidelity of Virtual Reality

MimeTIC wishes to promote the use of Virtual Reality to analyze and train human motor performance. It raises the fundamental question of the transfer of knowledge and skills acquired in VR to real life. In 2019, we put efforts in better understanding the potential fidelity of Virtual Reality experiences compared to real life experiences. It has been applied to various aspects of the interaction between pedestrians, but also the biomechanical fidelity of using haptic devices in highly constrainted conditions, such as hammering tasks.

7.3.1. Influence of Motion Speed on the Perception of Latency in Avatar Control

Participants: Ludovic Hoyet [contact], Richard Kulpa, Anthony Sorel, Franck Multon.

With the dissemination of Head Mounted Display devices in which users cannot see their body, simulating plausible avatars has become a key challenge. For fullbody interaction, avatar simulation and control involves several steps, such as capturing and processing the motion (or intentions) of the user using input interfaces, providing the resulting user state information to the simulation platform, computing a plausible adaptation of the virtual world, rendering the scene, and displaying the multisensory feedback to the user through output interfaces. All these steps imply that the displayed avatar motion appears to users with a delay (or latency) compared to their actual performance. Previous works have shown an impact of this delay on the perception-action loop, with possible impact on Presence and embodiment. We have explored [37] how the speed of the motion performed when controlling a fullbody avatar can impact the way people perceive and react to such a delay. We conducted an experiment where users were asked to follow a moving object with their finger, while embodied in a realistic avatar. We artificially increased the latency by introducing different levels of delays (up to 300ms) and measured their performance in the mentioned task, as well as their feeling about the perceived latency. Our results show that motion speed influenced the perception of latency: we found critical latencies of 80ms for medium and fast motion speeds, while the critical latency reached 120ms for a slow motion speed. We also noticed that performance is affected by both latency and motion speed, with higher speeds leading to decreased performance. Interestingly, we also found that performance was affected by latency before the critical latency for medium and fast speeds, but not for a slower speed. These findings could help to design immersive environments to minimize the effect of latency on the performance of the user, with potential impacts on Presence and embodiment.

7.3.2. Influence of Personality Traits and Body Awareness on the Sense of Embodiment in Virtual Reality

Participants: Ludovic Hoyet [contact], Rebecca Fribourg, Diane Dewez.

With the increasing use of avatars (i.e. the virtual representation of the user in a virtual environment) in virtual reality, it is important to identify the factors eliciting the sense of embodiment or the factors that can disrupt this feeling. This paper [35] reports an exploratory study aiming at identifying internal factors (personality traits and body awareness) that might cause either a resistance or a predisposition to feel a sense of embodiment towards a virtual avatar. To this purpose, we conducted an experiment (n=123) in which participants were immersed in a virtual environment and embodied in a gender-matched generic virtual avatar through a head-mounted display. After an exposure phase in which they had to perform a number of visuomotor tasks (during 2 minutes) a virtual character entered the virtual scene and stabbed the participants' virtual hand with a knife. The participants' sense of embodiment was measured, as well as several personality traits (Big Five traits and locus of control) and body awareness, to evaluate the influence of participants' personality on the acceptance of the virtual body. The major finding of the experiment is that the locus of control is linked to several components of embodiment: the sense of agency is positively correlated with an internal locus of control and the sense of body ownership is positively correlated with an external locus of control. Interestingly, both components are not influenced by the same traits, which confirms that they can appear independently. Taken together our results suggest that the locus of control could be a good predictor of the sense of embodiment when the user embodies an avatar with a similar physical appearance.

7.3.3. Gaze Behaviour During Person-Person Interaction in VR

Participants: Ludovic Hoyet, Anne-Hélène Olivier [contact], Florian Berton.

Figure 4. Conditions used in this work to understand the effect of VR setup on gaze beaviour in a collision avoidance task.

Simulating realistic interactions between virtual characters has been of interest to research communities for years, and is particularly important to automatically populate virtual environments. This problem requires to accurately understand and model how humans interact, which can be difficult to assess. In this context, Virtual Reality (VR) is a powerful tool to study human behaviour, especially as it allows assessing conditions which are both ecological and controlled. While VR was shown to allow realistic collision avoidance adaptations, in the frame of the ecological theory of perception and action, interactions between walkers can not solely be characterized through motion adaptations but also through the perception processes involved in such interactions. The objective of this study [30] is therefore to evaluate how different VR setups influence gaze behaviour during collision avoidance tasks between walkers. In collaboration with Julien Pettré in Rainbow team, we designed an experiment involving a collision avoidance task between a participant and another walker (real confederate or virtual character). During this task, we compared both the participant's locomotion and gaze behaviour in a real environment and the same situation in different VR setups (including a CAVE, a screen and a Head-Mounted Display) as illustrated on Figure4 . Our results show that even if some quantitative differences exist, gaze behaviour is qualitatively similar between VR and real conditions. Especially, gaze behaviour in VR setups including a HMD is more in line with the real situation than the other setups. Furthermore, the outcome on motion adaptations confirms previous work, where collision avoidance behaviour is qualitatively similar in VR and real conditions. In conclusion, our results show that VR has potential for qualitative analysis of locomotion and gaze behaviour during collision avoidance. This opens perspectives in the design of new experiments to better understand human behaviour, in order to design more realistic virtual humans.

7.3.4. Gaze Anticipation in Curved Path in VR

Participants: Anne-Hélène Olivier [contact], Hugo Brument.

This work was performed in collaboration with Ferran Argelaguet-Sanz and Maud Marchal from Hybrid team [32]. We investigated whether the body anticipation synergies in real environments (REs) are preserved during navigation in virtual environments (VEs). Experimental studies related to the control of human locomotion in REs during curved trajectories report a top-down reorientation strategy with the reorientation of the gaze anticipating the reorientation of head, the shoulders and finally the global body motion. This anticipation behavior provides a stable reference frame to the walker to control and reorient whole-body according to the future direction. To assess body anticipation during navigation in VEs, we conducted an experiment where participants, wearing a head-mounted display, were asked to perform a lemniscate trajectory in a virtual environment (VE) using five different navigation techniques, including walking, virtual steering (hand, head or torso steering) and passive navigation. For the purpose of this experiment, we designed a new control law based on the powerlaw relation between speed and curvature during human walking. Taken together our results showed that a similar ordered top-down sequence of reorientation of the gaze, head and shoulders during curved trajectories between walking in REs and in VEs (for all the evaluated techniques). However, this anticipation mechanism significantly differs between physical walking in VE, where the anticipation is higher, and the other virtual navigation techniques. The results presented in this paper pave the way to the better

understanding of the underlying mechanisms of human navigation in VEs and to the design of navigation techniques more adapted to humans

7.3.5. Validity of VR to Study Social Norms During Person-Person Interaction

Participants: Anne-Hélène Olivier [contact], Ludovic Hoyet, Florian Berton.

The modelling of virtual crowds for major events, such as the Oympics in Paris in 2024, takes into account the global proxemics standards of individuals without questioning the possible variability of these standards according to the space in which the interactions are performed. We know that body interactions (Goffman, 1974) are subject to rules whose variability is, at least in part, cultural (Hall, 1971). Obviously, these proxemics standards also address practical issues such as available space and space occupancy density. Our objective in this study was to understand the conditions which can explain that the discomfort felt and the adaptive behaviour performed differ when the interaction takes place in the same city and in spaces with identical occupancy densities. Especially, we focused on the effect of the social context of the environement. We aim at estimating the extent to which the prospect of attending a sports performance alters sensitivity to the transgression of proxemics norms. An additional objective was to evaluate whether virtual reality can help us to provide new insights in such a social context, where objective measures out-of-the lab are complex to perform. To answer this question, we designed in collaboration with Julien Pettré (Rainbow team) and colleagues in the field of sociology François Le Yondre, Théo Rougant and Tristan Duverne (Univ Rennes II) an experiment (in real context and then in virtual reality) in two different locations: a train station and the surroundings of a stadium before a league 1 football match) but with similar densities. The task performed by a confederate was to walk and stand excessively close to men aged 20 to 40. The individual's behaviour (not conscious of being a subject of the experiment) was observed by ethnography and explanatory interviews were conducted immediately afterwards. This same experiment was carried out in virtual reality conditions on the same type of population, modelling the two spaces and making it possible to acquire more precise and quantifiable data than in real conditions such as distances, travel time and eye fixations. The results show that the discomfort shown is much higher in the train station. The sporting context seems to participate in a form of relaxation of the norms of bodily interaction. Such a gap is not observable in virtual reality. From a methodological point of view, explicit interviews make it possible to usefully identify the reasons why virtual reality does not generate the same reactions, although it sometimes provokes the same sensitivity. Future work is needed to evaluate the effect of an increased immersion on such Social Science studies.

7.4. Motion Sensing of Human Activity

MimeTIC has a long experience in motion analysis in laboratory condition. In the MimeTIC project, we proposed to explore how these approaches could be transferred to ecological situations, with a lack of control on the experimental conditions. In the continuation of 2018, we have proposed to explore the use of cheap depth cameras solution for on-site motion analysis in ergonomics.

7.4.1. Motion Analysis of Work Conditions Using Commercial Depth Cameras in Real Industrial Conditions

Participant: Franck Multon [contact].

Based on a former PhD thesis (of Pierre Plantard) we have demonstrated the use of depth sensors in industry to assess risks of musculoskeletal disorders ar work. It has leaded to the creation of the the KIMEA software and of the Moovency start-up company in November 2018. In 2019 we published a synthesis work with new results [48] to demonstrate that such an approach can actually support the work of ergonomists in their goal to enhance the quality of life of workers in industry.

Hence, measuring human motion activity in real work condition is challenging as the environment is not controlled, while the worker should perform his/her task without perturbation. Since the early 2010s, affordable and easy-to-use depth cameras, such as the Microsoft Kinect system, have been applied for in-home entertainment for the general public. In this work, we evaluated such a system for the use in motion analysis in work conditions and propose software algorithms to enhance the tracking accuracy. Firstly, we highlighted the

high performance of the system when used under the recommended setup without occlusions. However, when the position/orientation of the sensor changes, occlusions may occur and the performance of the system may decrease, making it difficult to be used in real work conditions. Secondly, we propose a software algorithm to adapt the system to challenging conditions with occlusions to enhance the robustness and accuracy. Thirdly, we show that real work condition assessment using such an adapted system leads to similar results comparing with those performed manually by ergonomists. These results show that such adapted systems could be used to support the ergonomists work by providing them with reproducible and objective information about the human movement. It consequently saves ergonomists time and effort and allows them to focus on high-level analysis and actions.

7.5. Sports

MimeTIC promotes the idea of coupling motion analysis and synthesis in various domains, especially sports. More specifically, we have a long experience and international leadership in using Virtual Reality for analyzing and training sports performance. In 2019, we continued to explore 1) how enhancing on-site sports motion analysis using models inspired from motion simulation techniques, and 2) how Virtual Reality could be used to analyze and train motor and perceptual skills in sports.

7.5.1. Analysis of Fencing Lunge Accuracy and Response Time in Uncertain Conditions With an Innovative Simulator

Participants: Anthony Sorel [contact], Richard Kulpa, Nicolas Bideau, Charles Pontonnier.

We conducted a study evaluating the motor control strategies implied by the introduction of uncertainty in the realization of lunge motions [27]. Lunge motion is one of the fundamental attacks used in modern fencing, asking for a high level of coordination, speed and accuracy to be efficient. The aim of the current paper was the assessment of fencer's performance and response time in lunge attacks under uncertain conditions. For this study, an innovative fencing lunge simulator was designed. The performance of 11 regional to national-level fencers performing lunges in Fixed, Moving and Uncertain conditions was assessed. The results highlighted notably that i) Accuracy and success decreased significantly in Moving and Uncertain conditions with regard to Fixed ones ii) Movement and Reaction times were also affected by the experimental conditions iii) Different fencer profiles were distinguishable among subjects. In conclusion, the hypothesis that fencers may privilege an adaptation to the attack conditions and preserve accuracy instead of privileging quickness was supported by the results. Such simulators may be further used to analyze in more detail the motor control strategies of fencers through the measure and processing of biomechanical quantities and a wider range of fencing levels. It has also a great potential to be used as training device to improve fencer's performance to adapt his attack to controlled opponent's motion.

7.5.2. Enactive Approach to Assess Perceived Speed Error during Walking and Running in Virtual Reality

Participants: Théo Perrin, Richard Kulpa [contact], Charles Faure, Anthony Sorel, Benoit Bideau.

The recent development of virtual reality (VR) devices such as head mounted displays (HMDs) increases opportunities for applications at the confluence of physical activity and gaming. Recently, the fields of sport and fitness have turned to VR, including for locomotor activities, to enhance motor and energetic resources, as well as motivation and adherence. For example, VR can provide visual feedbacks during treadmill running, thereby reducing monotony and increasing the feeling of movement and engagement with the activity. However, the relevance of using VR tools during locomotion depends on the ability of these systems to provide natural immersive feelings, specifically a coherent perception of speed. The objective of this study is to estimate the error between actual and perceived locomotor speed in VE using an enactive approach, i.e. allowing an active control of the environment. Sixteen healthy individuals participated in the experiment, which consisted in walking and running on a motorized treadmill at speeds ranging from 3 to 11 km/h with 0.5 km/h increments, in a randomized order while wearing a HMD device (HTC Vive) displaying a virtual racetrack. Participants were instructed to match VE speed with what they perceived was their actual

locomotion speed (LS), using a handheld Vive controller. They were able to modify the optic flow speed (OFS) with a 0.02 km/h increment/decrement accuracy. An optic flow multiplier (OFM) was computed based on the error between OFS and LS. It represents the gain that exists between the visually perceived speed and the real locomotion speed experienced by participants for each trial. For all conditions, the average of OFM was 1.00 $\pm$.25 to best match LS. This finding is at odds with previous works reporting an underestimation of speed perception in VR. It could be explained by the use of an enactive approach allowing an active and accurate matching of visually and proprioceptively perceived speeds by participants. But above all, our study showed that the perception of speed in VR is strongly individual, with some participants always overestimating and others constantly underestimating. Therefore, a general OFM should not be used to correct speed in VE to ensure congruence in speed perception, and we propose the use of individual models as recommendations for setting up locomotion-based VR applications.

7.5.3. Acting Together, Acting Stronger? Interference Between Participants During Face-to-Face Cooperative Interception Task

Participants: Charles Faure, Théo Perrin, Richard Kulpa [contact], Anthony Sorel, Anabelle Limballe, Benoit Bideau.

People generally coordinate their action to be more effective. However, in some cases, interference between them occur, resulting in an inefficient collaboration. The main goal of this study [16], [16] is to explore the way two persons regulate their actions when performing a cooperative task of ball interception, and how interference between them may occur. Starting face to face, twenty-four participants (twelve teams of two) had to physically intercept balls moving down from the roof to the floor in a virtual room. To this end, they controlled a virtual paddle attached to their hand moving along the anterior-posterior axis, and were not allowed to communicate. Results globally showed participants were often able to intercept balls without collision by dividing the interception space in two equivalent parts. However, an area of uncertainty (where many trials were not intercepted) appeared in the center of the scene highlighting the presence of interference between participants. The width of this area increased when situation became more complex and when less information was available. Moreover, participants often interpreted balls starting above them as balls they should intercept, even when these balls were in fine intercepted by their partner. Overall, results showed that team coordination emerges from between-participants interactions in this ball interception task and that interference between them depends on task complexity (uncertainty on partner's action and visual information available).

7.5.4. Detection of Deceptive Motions in Rugby from Visual Motion Cues

Participants: Richard Kulpa [contact], Anne-Hélène Olivier, Benoit Bideau.

Figure 5. Illustration of a rugby player interacting in VR with 3 representations of a virtual attacker.

Frequently, in rugby, players incorporate deceptive motions (e.g., a side-step) in order to pass their opponent. Previous works showed that expert defenders are more efficient in detecting deceptive motions. Performance was shown to be correlated with the evolution of the center of gravity of the attacker, suggesting that experts may rely on global motion cues. This study [19] aims at investigating whether a representation of center of gravity can be useful for training purposes, by using this representation alone or by combining it with the local motion cues given by body parts. We designed an experiment in virtual reality to control the motion cues available to the defenders. Sixteen healthy participants (seven experts and nine novices) acted as defenders while a virtual attacker approached. Participants completed two separate tasks. The first was a time occlusion perception task, occlusion after 100ms, 200ms or 300ms after the initial change in direction, thereafter participants indicated the passing direction of the attacker. The second was a perception-action task Figure 5 , participants were instructed to intercept the oncoming attacker by displacing medio-laterally. The attacker performed either a non-deceptive motion, directly toward the final passing direction or a deceptive motion, initially toward a false direction before quickly reorienting to the true direction. There was a main effect of expertise, appearance, cut off times and motion on correct responses during both tasks. There was an interaction between visual appearance and expertise, and between motion type and expertise during the perception task, however, this interaction was not present during the perception-action task. We observed that experts maintained superiority in the perception of deceptive motion; however when the visual appearance is reduced to global motion alone the difference between novices and experts is reduced. We further explore the interactions and discuss the effects observed for the visual appearance and expertise.

7.5.5. IMU-based Motion Capture for Cycling Performance

Participants: Nicolas Bideau [contact], Guillaume Nicolas, Benoit Bideau, Sebastien Cordillet, Erwan Delhaye.

The quantification of 3D kinematical parameters such as body segment orientations and joint angles is important in the monitoring of cycling to provide relevant biomechanical parameters associated with performance optimization and/or injury prevention. Numerous experiments based on optoelectronic motion capture have been conducted in the laboratory to analyze kinematical variables (e.g., joint angles) during cycling. However, the assessment of kinematics in real conditions during training or competition is a challenging task, especially since conventional optoelectronic motion capture systems suffer from major drawbacks (restricted fields of view, cumbersome and time consuming) in this regard. To overcome these limitations, inertial measurement units (IMU) is a relevant solution for in situ cycling analysis as they allow a continuous data acquisition process throughout a cycling exercise. Beyond the common problem of the drift related to the integration of gyroscope data, one of the major issues in joint kinematics assessment using IMU devices lies is the misalignment of sensor axes with the anatomical body segment axis, which is not straightforward. Thus, we developed a novel sensor-to-segment calibration procedure for inertial sensor-based knee joint kinematics analysis during cycling. This procedure was designed to be feasible in-field, autonomously, and without any external operator or device. It combines a static standing up posture and a pedaling task. In comparison with conventional calibration methods commonly employed in gait analysis, the new method we proposed significantly improved the accuracy of 3D knee joint angle measurement when applied to cycling analysis [14]. As a second step related to the in-field application to track cycling, we estimated lower limb joint angles during a time trial on a velodrome. This integrative measurement exhibited the evolution of kinematic parameters in relation with distance but also with the track curvature [43].

7.6. Ergonomics

Ergonomics has become an important application domains in MimeTIC: being able to capture, analyze, and model human performance at work. In this domain, key challenge consists in using limited equipment to capture the physical activity of workers in real conditions. Hence, in 2019, we have designed a new approach to predict external forces using mainly motion capture data, and to personnalize the biomechanical capabilities (maximum feasible force/torque) of specific population.

7.6.1. Motion-based Prediction of External Forces

Participants: Charles Pontonnier [contact], Georges Dumont, Claire Livet, Anthony Sorel, Nicolas Bideau.

We proposed [21] a method to predict the external efforts exerted on a subject during handling tasks, only with a measure of his motion. These efforts are the contacts forces and moments on the ground and on the load carried by the subject. The method is based on a contact model initially developed to predict the ground reaction forces and moments. Discrete contact points are defined on the biomechanical model at the feet and the hands. An optimization technique computes the minimal forces at each of these points satisfying the dynamic equations of the biomechanical model and the load. The method was tested on a set of asymmetric handling tasks performed by 13 subjects and validated using force platforms and an instrumented load. For each task, predictions of the vertical forces obtained a RMSE of about 0.25 N/kg for the feet contacts and below 1 N/kg for the hands contacts. This method enables to quantitatively assess asymmetric handling tasks on the basis of kinetics variables without additional instrumentation such as force sensors and thus improve the ecological aspect of the studied tasks. We evaluated this method [23] on manual material handling (MMH) tasks. From a set of hypothesized contact points between the subject and the environment (ground and load), external forces were calculated as the minimal forces at each contact point while ensuring the dynamics equilibrium. Ground reaction forces and moments (GRF&M) and load contact forces and moments (LCF&M) were computed from motion data alone. With an inverse dynamics method, the predicted data were then used to compute kinetic variables such as back loading. On a cohort of 65 subjects performing MMH tasks, the mean correlation coefficients between predicted and experimentally measured GRF for the vertical, antero-posterior and medio-lateral components were 0.91 (0.08), 0.95 (0.03) and 0.94 (0.08), respectively. The associated RMSE were 0.51 N/kg, 0.22 N/kg and 0.19 N/kg. The correlation coefficient between L5/S1 joint moments computed from predicted and measured data was 0.95 with a RMSE of 14 Nm for the flexion / extension component. This method thus allows the assessment of MMH tasks without force platforms, which increases the ecological aspect of the tasks studied and enables performance of dynamic analyses in real settings outside the laboratory.

This method was successfully applied [24] on lunge motion that is a fundamental attack of modern fencing, asking for a high level of coordination, speed and accuracy. It consists in an explosive extension of the front leg accompanying an extension of the sword arm. In such motions, the direction of action and the way feet are oriented – guard position - are particularly challenging for a GRF&M prediction method. These methods are avalaible in CusToM software [22].

7.6.2. Biomechanics for Motion Analysis-Synthesis and Analysis of Torque Generation Capacities

Participants: Charles Pontonnier [contact], Georges Dumont, Nicolas Bideau, Guillaume Nicolas, Pierre Puchaud.

Characterization of muscle mechanism through the torque-angle and torque-velocity relationships [17] is critical for human movement evaluation and simulation. In-vivo determination of these relationships through dynamometric measurements and modelling is based on physiological and mathematical aspects. However, no investigation regarding the effects of the mathematical model and the physiological parameters underneath these models was found. The purpose of the current study was to compare the capacity of various torque-angle and torque-velocity models to fit experimental dynamometric measurement of the elbow and provide meaningful mechanical and physiological information. Therefore, varying mathematical function and physiological muscle parameters from the literature were tested. While a quadratic torque-angle model seemed to increase predicted to measured elbow torque fitting, a new power-based torque-velocity parametric model gave meaningful physiological values with similar fitting results to a classical torque-velocity model. This model is of interest to extract modelling and clinical knowledge characterizing the mechanical behavior the joint. Based on the same kind of methods, we proposed [25] to analyse torque generation capacities of a human knee. The torque generation capacities are often assessed for human performance, as well as for prediction of internal forces through musculoskeletal modelling. Scaling individual strength generation capacities is challenging but can provide physiologically meaningful perspectives. We propose to fit the models to isokinetic measurements of joint torques in different angle and angular velocity conditions. Assuming muscles are viscoelastic actuators, their entire architectures contribute to Joint Torque-Angle and Torque-Velocity Relationships (JTAR and JTVR respectively, and their coupling JTAVR) at the joint level. Experimental observation at different scales (muscle sarcomere, muscle fibre and joint) resulted in various JTAR models available in the literature. On

the other side, JVTR models are often modelled without obvious physiological consistency. The above mentionned JTVR model was shown to increase physiological transparency of the elbow JTAVR. As those results might be joint-specific, we extended it to evaluate five JTAR and two JTVR models on the knee flexion and extension.

7.7. Locomotion and Interactions between Walkers

MimeTIC is a leader in the study and modeling of walkers' visuo-motor strategies. This implies to understand how humans generate their walking trajectories within an environment. This year, one main focus was to consider how the interaction models change with specific populations (including kids, older adults, concussed athletes or person on a wheelchair) as well as in specific environment (including narrow sidewalk, or environment with varying social context).

7.7.1. Effect of Foot Stimulation on Locomotion

Participants: Anne-Hélène Olivier, Armel Crétual [contact], Carole Puil.

Medio-Intern Element (EMI®) is a thin plantar insert used by podiatrists to treat postural deficiency. It was shown an influence of a 3 mm high EMI on Medio-Lateral (ML) displacement of the Centre of Pressure (CoP) of healthy participants in quasi-static standing. Recently it has been demonstrated that EMI has an impact on eyes vergence, and especially in population with plantar postural dysfunction. These effects were weakly assessed however and only using static tasks. Therefore, the objective of this work [53], [52], [41], was to evaluate the effect of the EMI while performing a locomotor task. We expected a controlateral deviation of the trajectory when this insert was located under one foot. Indeed, in previous studies dealing with bottom-up control of locomotion, it was shown that a 30 min podokinetic stimulation leads to a ML deviation of the trajectory when participants were asked to walk in a straight line with eyes closed. 20 healthy participants volunteered for this study. They participated into 3 different sessions in random order: either without EMI, with EMI under the right foot or under the left foot. Each session involved first, static tasks (with and without vision) to compare with previous work, then, dynamic locomotor tasks with 6 different conditions mixing trajectory (straight walking, 90° left or right turn) and vision (with and without vision) in random order. In static conditions, we computed the average ML position of the CoP. In dynamic conditions, we analyzed the difference in the final orientation of the locomotor trajectory with and without vision with an EMI with respect to this difference without the EMI. No significant effect of the EMI was observed for either static or dynamic conditions. Our results do not confirm the previous work in static conditions. Future work is needed to better understand the effect of this insert. In particular, our participants were healthy and it could be interesting to evaluate this effect in participants with postural deficiencies. These results would have an application in the design of new clinical tests.

7.7.2. Collision Avoidance between Walkers on a Curvilinear Path

Participants: Anne-Hélène Olivier [contact], Armel Crétual, Richard Kulpa, Anthony Sorel.

Crowded public spaces require humans to interact with what the environment affords to regulate interpersonal distance to avoid collisions. In the case of rectilinear trajectories, the collision avoidance behaviours have been extensively studied. It has been shown that the perceived action-opportunities of the walkers might be afforded based on a future distance of closest approach (also coined 'Minimal Predicted Distance', MPD). However, typical daily interactions do not always follow rectilinear but also curvilinear trajectories. In that context, it has been shown that a ball following a curvilinear trajectory can be successfully intercepted. However, it remains unclear whether the collision avoidance strategies in the well-studied linear trajectories can be transferred to curvilinear trajectories. Therefore, the aim of this work [44] was to examine collision avoidance behaviours when interacting with walkers following curvilinear trajectories. An experiment was designed using virtual reality in which 22 participants navigated toward a goal in a virtual environment with a joystick. A Virtual Human (VH) crossed the path of the participant from left and right with varying risks of collision. The VH followed either a curvilinear path with a fixed radius of 5 m or 10 m, approaching from in-front of and behind the participant, or a control rectilinear path. The final crossing distance, the number of collisions and inversions

of initial crossing order were analysed to determine the success of the task. Further, MPD evolution over time and specific timing events was analysed across conditions. For a curvilinear path with a 5 m radius there were significantly more collisions when the VH approached from behind the participant, and significantly more inversions of the initial crossing order when the VH approached from in-front than the control rectilinear path. Final crossing distance was shorter when the VH followed a path with a 5 m radius from behind the participant. Finally, the evolution of the MPD over time was similar for paths with a 10 m radius when compared to the control rectilinear path, whereas the 5 m curvilinear paths had significant differences during the interaction. Overall, with few collisions and few inversions of crossing order we can conclude that participants were capable of interacting with virtual walkers on curvilinear trajectories. Further, the task was solved with similar avoidance adaptations to those observed for rectilinear interactions. However, paths with a smaller radius had more reported collisions and inversions. Future work should address how a curved trajectory during collision avoidance is perceived.

7.7.3. Collision Avoidance in Person-Specific Populations

Participants: Anne-Hélène Olivier [contact], Armel Crétual.

In the frame of the Inria BEAR associate team, we have used our 90° crossing paradigm to understand visuo-motor coordination in specific population. This is important, not only from a theoretical point of view but also to design more individual model of human locomotion in a dynamic environment.

Figure 6. Illustration of person-person interaction experiments in a) kids, b) older adults, c) previously concussed athletes, d) a person on an electric powered wheelchair

We first investigated the effect of age on visuo-motor coordination by considering a collision avoidance task in kids (8-12 years) and older adults (65-74 years) as illustrated on Figure 6 a,b. On one hand, middle-aged children have been shown to have poor perception-action coupling during static and dynamic collision avoidance tasks. Research has yet to examine whether perception-action coupling deficits persist in a dynamic collision avoidance task involving a child and another walker. In this work [26], [54], we invesigated whether the metric MPD(t) be used to examine collision avoidance strategies between children and adults. To this end, eighteen children (age: 10 $\pm$ 1.5 years) and eighteen adults (34 $\pm$ 9.6 years) walked while avoiding another participant (child or adult). Groups of three children and three adults were recruited per session. The results demonstrated that (1) MPD(t) can be used to predict future collisions in children, (2) MPD(t) is an absolute measure that is consistently lower when a child is involved compared to two adult walkers, (3) the individual passing second, even when it is a child, contributes more to MPD(t) than the walker passing first. It then appears that children have developed adult-like strategies during a collision avoidance task involving two walkers. Body anthropometrics should be considered when determining collision avoidance strategies between children and adults. On the other hand, every year, 1 in 3 older adults are likely to fall at least once and many falls occurs while walking where an individual needs to adapt to environmental hazards. Studies with older adults interacting within an environment showed difficulties in estimating time to arrival of vehicles, larger critical ratio and more variability in door aperture task as well as larger clearance distance when avoiding a moving object. The current study [51] aims to identify whether differences in collision

avoidance behaviours of older adults during a person-person collision avoidance task are the result of age-related visuomotor processing deficits. Results showed that no collision occurred, where older and younger adults were able to act appropriately. However, larger threshold were needed to trigger avoidance when an older adult is second in crossing order, possibly due to visuomotor delays. Moreover, we observed more crossing inversions with older adults, which may suggest a poor visumotor processing. Finally, the clearance distance was samller when to older adults interact with each others, resulting in “risky” behaviours. Interestingly, social factors seems to be involved since when a young and an older adults interact, the young adult contribute more to solve the collision avoidance task.

In close relation with the Application Domain "Sports", we also investigated visuo-motor coordination during locomotion in previously concussed rugby-players (Figure 6 c). Despite adherence to return-to-play guidelines, athletes with previous concussion exhibit persistent visuomotor deficits during static balance and visuomotor integration tasks such as collision avoidance months after returning to sport. Previous research in collision avoidance was done in a static setting, however less is known about visuomotor strategies utilized in dynamic scenarios, such as person-person interactions. In this context, during a collision avoidance locomotor task, individuals make adjustments to their path and/or their velocity in response to a risk of collision. These adjustments ensure that the clearance distance would be large enough such that no collision occurs. However, athletes with previous concussion may demonstrate impaired performance during a collision avoidance task requiring path adjustments based on visual information. The purpose of this study [55] was to investigate collision avoidance strategies when avoiding another walker between previously concussed athletes and healthy athletes. We hypothesized that previously concussed athletes would demonstrate altered trajectory adaptation and changes in individual contribution to the avoidance compared to healthy athletes. Preliminary results show that individuals with previous concussion demonstrated trajectory adaptation behaviours consistent with healthy athletes and young adults. However, previously concussed athletes passed with a reduced distance between themselves and the other walker when they are second in passage order at the crossing point. Athletes who have sustained a previous concussion show decreased collision avoidance behaviour. This behaviour results in a higher risk of a collision occurring, as individuals showed reduced contributions (i.e. creating physical space) to the avoidance of the collision. This change in typical behaviour on a visuomotor task may indicate a persistent deficit in perceptual abilities following concussion. Although trajectory adaptations were consistent with healthy athletes, these results suggest that athletes with previous concussion remain at an elevated risk of collision and possible injury following concussion recovery. This study provides novel insights and additional evidence that visuomotor and perceptual impairments persist following return to play in previously concussed athletes. Additionally, this protocol has important implications for the assessment and rehabilitation of visuomotor processes that are affected following a concussion. Future research could further develop this protocol to be used in sideline assessment, and guide treatment of concussions past clinical recovery.

7.7.4. Collision Avoidance between a Walker and an Electric Powered Wheelchair: Towards Smart Wheelchair

Participants: Anne-Hélène Olivier [contact], Armel Crétual.

In collaboration with Marie Babel and Julien Pettré from Inria Rainbow team, we are interested in the development of smart electric powered wheelchairs (EPW), which provide driver assistance. Developing smart assistance requires to better understand interactions between walkers and such vehicles. We focus on collision avoidance task between an EPW (fully operated by a human) and a walker, where the difference in the nature of the agents (weight, maximal speed, acceleration profiles) results into asymmetrical physical risk in case of a collision, for example due to the protection EPW provides to its driver, or the higher energy transferred to the walker during head-on collision. In this work [39], [47], our goal is to demonstrate that this physical risk asymmetry results into differences in the walker’s behavior during collision avoidance in comparison to human-human situations. 20 participants (15 walkers and 5 EPW drivers) volunteered to this study. The experiment was performed in a 30mx20m gymnasium. We designed a collision avoidance task, where an EPW and a human walker moved towards a goal with orthogonal crossing trajectories (Figure6 d). We recorded their trajectory among 246 trials (each trial being 1 collision avoidance). We compared the predicted passage order

when they can first see each other with the one observed at the crossing point to identify if inversions occur during the interaction. Note that during walker-walker interactions it was shown that the initial passage order is almost systematically preserved all along the interaction up to the crossing point. We also computed the shape-to-shape clearance distance. We observed 23.7% of passage order inversion, specifically in 20.8% of trials where walkers were supposed to cross first, they crossed second. This means that walkers were more likely to pass behind the EPW than in front. On average, human walkers crossed first when having sufficient advance on the wheelchair to reach the crossing point. We estimated this advance up to 0.91m. The shape-to-shape clearance distance was influenced by the passage order at the crossing point, with larger distance when the walker cross first (M=0.78m) than second (M=0.34m). Results show that walkers set more conservative strategies when interacting with an EPW. By passing more frequently behind the EPW, they avoid risks of collisions that would lead to high energy transfer. Also, when they pass in front, they significantly increase the clearance distance, compared to cases where they pass behind. These results can then be linked to the difference in the physical characteristics of the walkers and EPW where asymmetry in the physical risks raised by collisions influence the strategies performed by the walkers in comparison with a similar walker-walker situation. This gives interesting insights in the task of modeling such interactions, indicating that geometrical terms are not sufficient to explain behaviours, physical terms linked to collision momentum should also be considered.

7.7.5. Collision Avoidance on a Narrow Sidewalk

Participant: Anne-Hélène Olivier [contact].

In the context of transportation research and a collaboration with the colleagues of Ifsttar (LEPSIS, LESCOT), we investigate person-person interaction when walking on a narrow sidewalk [34]. Narrow sidewalks are not the result of imagination nor a heritage of the former urban planning in the oldest cities. They exist in many modern cities, a simple web query provides a lot of examples in the world. In most cases, two pedestrians walking in opposite way cannot stay both on the sidewalk when they cross: one has to give a free way on the curb by stepping down on the road, which can generate risky situations for pedestrians. These situations are nowadays underestimated and so are the associated risk. In this context, driving simulators and walking simulators are useful tools to conduct studies in a safe environment with controlled conditions. Therefore, they can allow improving our knowledge on the way pedestrians interact on a narrow sidewalk and how drivers can react when facing this situation. This contribution aims to model the behaviours of simulated pedestrians, Non Player Characters (NPC). Using an interdisciplinary framework, we first identified from the literature psychosocial factors that should be involved in such interactions. Then, we designed a questionnaire to evaluate the impact of these factors on the perception of these interaction. Based on the main factors, we developed a perception model, and we modified the ORCA model, which is one of the most used for pedestrian collision avoidance simulation. Finally, we assessed the consistency of all our simulated interactions with a user study.

7.7.6. Shared Effort Model During Collision Avoidance

Participants: Anne-Hélène Olivier [contact], Armel Crétual.

In collaboration with Jose Grimaldo da Silva and Thierry Fraichard (Inria Grenoble), we finally designed a shared-effort model during interaction between a moving robot and a human relying on walker-walker collision avoidance data. [33]. Recent works in the domain of Human-Robot Motion (HRM) attempted to plan collision avoidance behavior that accounts for cooperation between agents. Cooperative collision avoidance between humans and robots should be conducted under several factors such as speed, heading and also human attention and intention. Based on some of these factors, people decide their crossing order during collision avoidance. However, whenever situations arise in which the choice crossing order is not consistent for people, the robot is forced to account for the possibility that both agents will assume the same role, a decision detrimental to collision avoidance. In our work we evaluate the boundary that separates the decision to avoid collision as first or last crosser. Approximating the uncertainty around this boundary allows our collision avoidance strategy to address this problem based on the insight that the robot should plan its collision avoidance motion in such a way that, even if agents, at first, incorrectly choose the same crossing order, they would be able to unambiguously perceive their crossing order on their following collision avoidance action.

POTIOC Project-Team

7. New Results

7.1. Mixed-Reality System for Collaborative Learning at School

Participants: Philippe Giraudeau, Théo Segonds, Solène Lambert, Martin Hachet

External collaborators: Université de Lorraine

Traditional computer systems based on the WIMP paradigm (Window, Icon, Menu, Pointer) have shown potential benefits at school (e.g. for web browsing). On the other hand, they are not well suited as soon as hands-on and collaborative activities are targeted. To face this problem, we have designed and developed CARDS, a Mixed-Reality system that combines together physical and digital objects in a seamless workspace to foster active and collaborative learning (Figure 3). In [23], we describe the design process based on a participatory approach with researchers, teachers, and pupils. We then present and discuss the results of a user study that tends to show that CARDS has a good educational potential for the targeted activities.

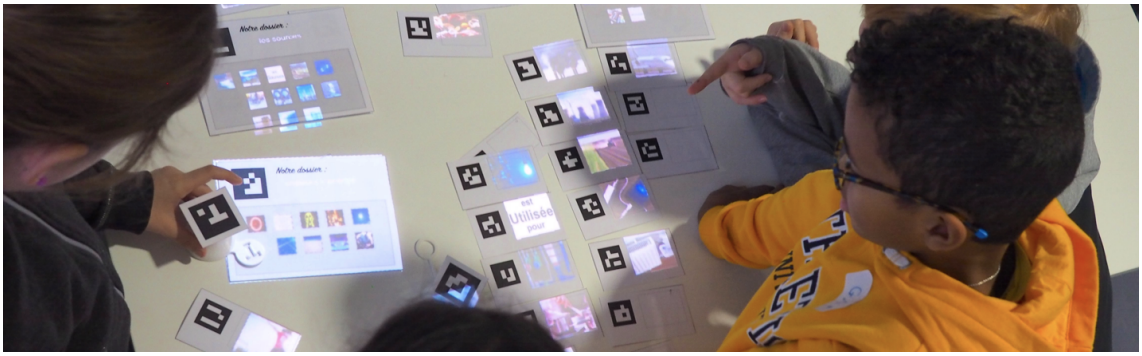


Figure 3. CARDS: Collaborative Activities based on the Real and the Digital Superimposition.

7.2. DroneSAR: Extending Physical Spaces in Spatial Augmented Reality using Projection on a Drone

Participants: Rajkumar Darbar, Joan Sol Roo, Thibaut Lainé, Martin Hachet

Spatial Augmented Reality (SAR) transforms real-world objects into interactive displays by projecting digital content using video projectors. SAR enables co-located collaboration immediately between multiple viewers without the need to wear any special glasses. Unfortunately, one major limitation of SAR is that visual content can only be projected onto its physical supports. As a result, displaying User Interfaces (UI) widgets such as menus and pop-up windows in SAR is very challenging. We are trying to address this limitation by extending SAR space in mid-air. We propose Drone-SAR, which extends the physical space of SAR by projecting digital information dynamically on the tracked panels mounted on a drone (see Figure 4 . DroneSAR is a proof of concept of novel SAR User Interface (UI) , which provides support for 2D widgets (i.e., label, menu, interactive tools, etc.) to enrich SAR interactive experience. We describe this concept, as well as implementation details of our proposed approach in [22].

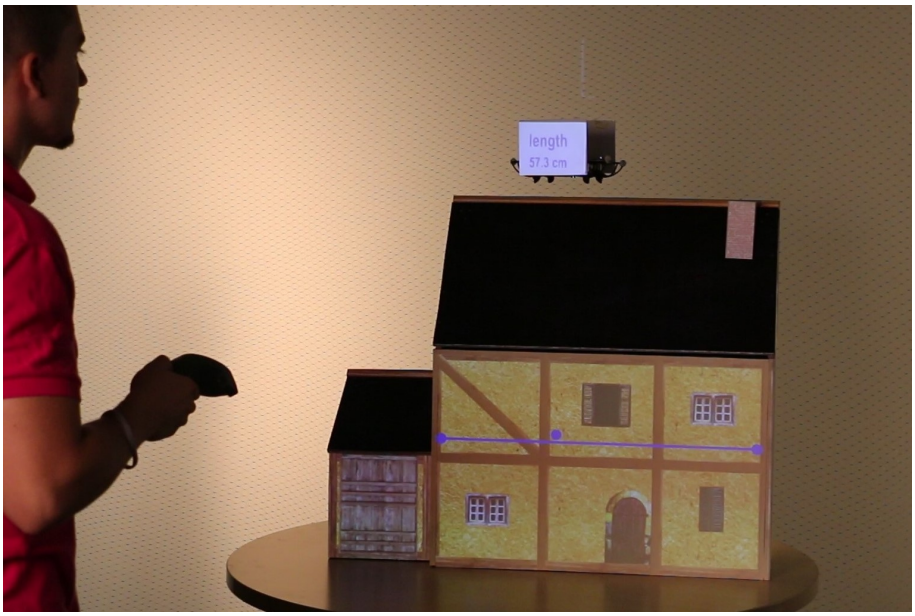


Figure 4. DroneSAR: Projection on a drone allows us to extend the physical space for interacting spatial augmented reality.

7.3. Tangible and modular devices for supporting communication

Participants: Joan Sol Roo, Pierre-Antoine Cinquin, Martin Hachet

External collaborators: Ullo

Our physiological activity reflects our inner workings. However, we are not always aware of it in full detail. Physiological devices allow us to monitor and create adaptive systems and support introspection. Given that these devices have access to sensitive data, it is vital that users have a clear understanding of the internal mechanisms (extrospection), yet the underlying processes are hard to understand and control, resulting in a loss of agency. In this work, we focus on bringing the agency back to the user, by using design guidelines based on principles of honest communication and driven by positive activities. To this end, we conceived a tangible, modular approach for the construction of physiological interfaces (see Figure 5). We are exploring the potential of such an approach with a set of examples, supporting introspection, dialog, music creation, and play.

7.4. Accessible Interactive Audio-Tactile Drawings

Participants: Lauren Thevin, Anke Brock, Martin Hachet

External collaborators: CNRS, Univesité Paul Sabatier, ENAC

Interactive tactile graphics have shown a true potential for people with visual impairments, for instance for acquiring spatial knowledge. Until today, however, they are not well adopted in real-life settings (e.g. special education schools). One obstacle consists in the creation of these media, which requires specific skills, such as the use of vector-graphic software for drawing and inserting interactive zones, which is challenging for stakeholders (social workers, teachers, families of people with visual impairments, etc.). We explored how a Spatial Augmented Reality approach can enhance the creation of interactive tactile graphics by sighted users. We developed the system using a participatory design method. A user study showed that the augmented

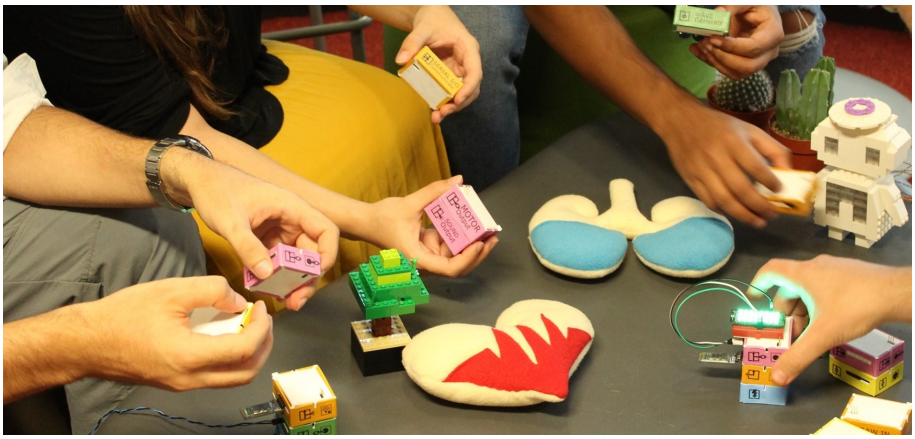


Figure 5. modular bricks that support the easy creation and interfacing with physiological applications.

reality device allowed stakeholders (N=28) to create interactive tactile graphics more efficiently than with a regular vector-drawing software (baseline), independently of their technical background. This work illustrated in Figure 6 is described in [29].

Following the same approach, we are currently exploring how physical board games can be moved into accessible ones for people with visual impairments.

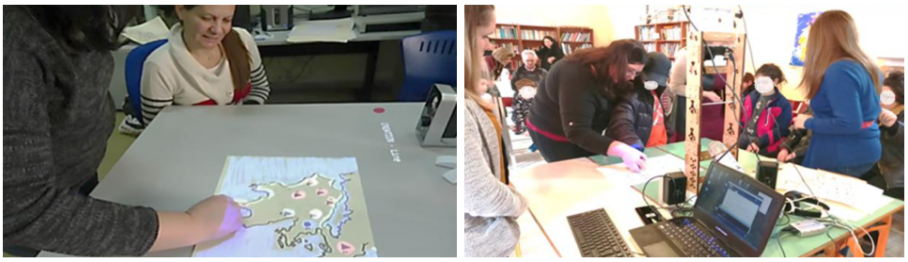


Figure 6. Combining touch and audio for accessible drawings at school.

7.5. Accessibility of e-learning systems

Participants: Pierre-Antoine Cinquin, Damien Caselli and Pascal Guitton

External collaborators: Hélène Sauzéon

In 2019, we continued to work on new digital teaching systems such as MOOCs. Unfortunately, accessibility for people with disabilities is often forgotten, which excludes them, particularly those with cognitive impairments for whom accessibility standards are fare from being established. We have shown in [12] that very few research activities deal with this issue.

In past years, we have proposed new design principles based on knowledge in the areas of accessibility (Ability-based Design and Universal Design), digital pedagogy (Instruction Design with functionalities that reduce the cognitive load : navigation by concept, slowing of the flow…), specialized pedagogy (Universal Design for Learning, eg, automatic note-taking, and Self Determination Theory, e.g., configuration of the interface according to users needs and preferences) and psychoducational interventions (eg, support the joint teacher-learner attention), but also through a participatory design approach involving students with disabilities and experts in the field of disability (Figure 7). From these framework, we have designed interaction features which have been implemented in a specific MOOC player called Aïana. Moreover, we have produced a MOOC on digital accessibility which is published on the national MOOC platform (FUN) using Aïana (4 sessions since 2016 with more than 11 000 registered participants). https://mooc-francophone.com/cours/mooc-accessibilite-numerique/. Our first field studies demonstrate the benefits of using Aïana for disabled participants [32].

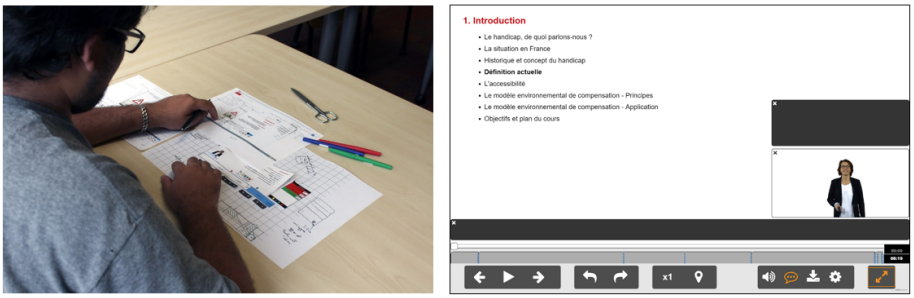


Figure 7. Design of Aïana MOOC player.

7.6. A hybrid setup for an artistic experience

Participants: Vincent Da Silva Pinto, Martin Hachet

External collaborators: Léna d'azy

In 2019, we have worked with the stenographer Cécile Léna to conceive and build a hybrid setup that combines a real physical mock-up and a virtual environments. This allows the participants to explore the sky by pointing at planets with a telescope in miniature, and observing a virtual view of the pointed planet by looking through an immersive stereoscopic installation (Figure 8).

7.7. Mental state decoding from EEG signals using robust machine learning

Participants: Aurélien Appriou, Smeethy Pramij, Khadijeh Sadatnejad, Aline Roc, Léa Pillette, Thibaut Monseigne, Fabien Lotte

External collaborators: Andrzej Cichocki, Pierre-Yves Oudeyer, Edith Law, Jessie Ceha, Frédéric Dehais, Alban Duprès, Sarah Blum, Nicolas Drougard, Sébastien Scannella, Raphaëlle N. Roy

Modern machine learning algorithms to classify cognitive and affective states from electroencephalography signals: Estimating cognitive or affective states from brain signals is a key but challenging step in the creation of passive brain-computer interface (BCI) applications. So far, estimating mental workload or emotions from EEG signals is only feasible with modest classification accuracies, thus leading to unreliable neuroadaptive applications. However, recent machine learning algorithms, notably Riemannian geometry based classifiers (RGC) and convolutional neural networks (CNN), have shown to be promising for other BCI systems, e.g., motor imagery-BCIs. However, they have not been formally studied and compared together for

Figure 8. Echelles célestes allows the participant to explore the sky.

cognitive or affective states classification. We have thus explored such machine learning algorithms, proposed new variants of them, and benchmarked them with classical methods to estimate both mental workload and affective states (Valence/Arousal) from EEG signals. We studied these approaches with both subject-specific and subject-independent calibration, to go towards calibration-free systems. Our results suggested that a CNN obtained the highest mean accuracy, although not significantly so, in both conditions for the mental workload study, followed by RGCs. However, this same CNN underperformed in both conditions for the emotion data set, a data set with little training data. On the contrary, RGCs proved to have the highest mean accuracy with the Filter Bank Tangent Space classifier (FBTSC) we introduced in this paper. Our results thus contributed to improve the reliability of cognitive and affective states classification from EEG. They also provide guidelines about when to use which machine learning algorithm. This work was just accepted for publication in the IEEE System Man and Cybernetics magazine.

Towards decoding curiosity from Brain and physiological signals: The neurophysiological mechanisms underlying curiosity and intrinsic motivation are currently not well understood. However, being able to identify objectively, from neurophysiological signals, the curiosity level of a user, would bring a very useful tool both to neuroscientists and psychologists, to understand curiosity deeper, as well as to designers of human-computer interaction, in order to trigger curiosity or to adapt an interaction to the curiosity levels of its users. A first step to do that, is to collect neurophysiological signals during known states of curiosity, in order to develop signal processing/machine learning tools to recognize those states from such signals. We designed and ran an experimental protocol to measure both brain activity through Electroencephalography (EEG) and physiological responses (heart rate, skin conductance, Electrocardiogram) when subjects were induced into different states of curiosity. During the experiment, fun facts were presented to subjects to induce different levels of curiosity. We obtained those fun facts using the Google functionality "I'm feeling curious" as well as crowdsourcing. A subject could choose a fun fact that made him curious, and push forward with a 4-to-10 questions chain on this theme. For each question on a given theme, a subject could choose to reveal the answer (interpreted as a curious state) or to skip it (interpreted as a non-curious state). Skipping an answer will automatically break the chain and will point the subject to the next fun fact. Neurophysiological signals were collected from 28 subjects, between a question and the choice of revealing the answer. Then those subjects graded the question on a 1-to-7 curiosity level scale. We are currently working on finding biological markers of curiosity by analyzing the collected signals using machine learning.

Channel Selection over Riemannian manifold with non-stationarity consideration for Brain-Computer interface applications: EEG signals are essentially non-stationary. Such non-stationarities, including cross-trial, cross-session, and cross-subject non-stationarities, are the result of various neurophysiological and extra-physiological causes. Such non-stationarities lead to variations in BCI users' performance. To handle this problem, we designed and compared multiple criteria for selecting EEG channels over the Riemannian manifold, for EEG classification. These criteria aim to promote EEG covariance matrix classifiers to generalize well by considering EEG data non-stationarity. Our approach consists of both increasing the discriminative information between classes over the manifold and reducing the dispersion within classes. We also reduce the influence of outliers in both discriminative and dispersion measures. The criteria were evaluated on EEG signals recorded from a tetraplegic subject and dataset IVa from BCI competition III. Experimental evidences confirm that considering the dispersion within each class as a measure for quantifying the effects of non-stationarity and removing the most affected channels can improve BCI performance (see Figure 9). This work was submitted to ICASSP 2020.

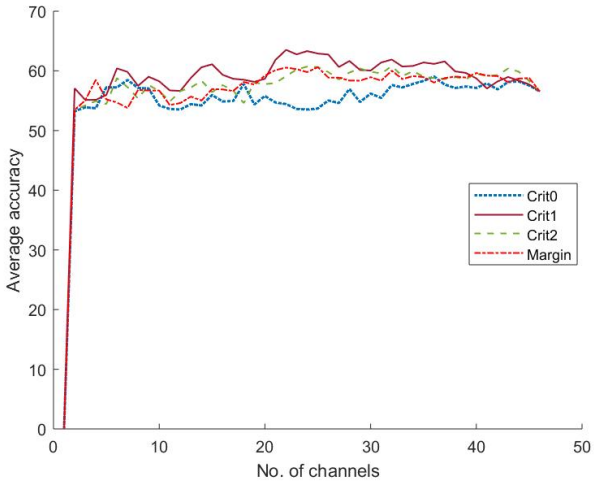


Figure 9. Results of BCI classification accuracy for various number of selected channels, for different channel selection algorithms. Our proposed algorithms - termed here Crit1, Crit2 and Margin - all improved upon the state-of-the-art (Crit0).

Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions: Recent technological progress has allowed the development of low-cost and highly portable brain sensors such as pre-amplified dry-electrodes to measure cognitive activity out of the laboratory. This technology opens promising perspectives to monitor the "brain at work" in complex real-life situations such as while operating aircraft. However, there is a need to benchmark these sensors in real operational conditions. We therefore designed a scenario in which twenty-two pilots equipped with a six-dry-electrode EEG system had to perform one low load and one high load traffic pattern along with a passive auditory oddball. In the low load condition, the participants were monitoring the flight handled by a flight instructor, whereas they were flying the aircraft in the high load condition. At the group level, statistical analyses disclosed higher P300 amplitude for the auditory target (Pz, P4 and Oz electrodes) along with higher alpha band power (Pz electrode), and higher theta band power (Oz electrode) in the low load condition as compared to the high load one. Single trial classification accuracy using both event-related potentials and event-related frequency features at the same time did not exceed chance level to discriminate the two load conditions. However, when considering only the frequency features computed over the continuous signal,

classification accuracy reached around 70% on average. This study demonstrates the potential of dry-EEG to monitor cognition in a highly ecological and noisy environment, but also reveals that hardware improvement is still needed before it can be used for everyday flight operations. This work was published in the journal Sensors in [13].

7.8. Understand and modeling Mental-Imagery BCI user training

Participants: Camille Benaroch, Aline Roc, Léa Pillette, Fabien Lotte

External collaborators: Camille Jeunet, Bernard N'Kaoua

Computational models of performance: Mental-Imagery based Brain-Computer Interfaces (MI-BCIs) make use of brain signals produced during mental imagery tasks to control a computerised system. The currently low reliability of MI-BCIs could be due, at least in part, to the use of inappropriate user-training procedures. In order to improve these procedures, it is necessary first to understand the mechanisms underlying MI-BCI user-training, notably through the identification of the factors influencing it. Thus, we first aimed at creating a statistical model that could explain/predict the performances and the progression of MI-BCI users using their traits (e.g., personality). We used the data of 42 participants (i.e., 180 MI-BCI sessions in total) collected from three different studies that were based on the same MI-BCI paradigm. We used machine learning regressions with a leave-one-subject-out cross validation to build different models. A first results showed that using the users' traits only may enable the prediction of performances for a single multiple-session experiment, but might not be sufficient to reliably predict MI-BCI performances across different experiments. A second result showed that using the users' traits and the users' past performances may enable the prediction of the progression of one user as reliable models were found for two of the three studies. Part of this work was published at the International Graz BCI conference in [21].

Would Motor-Imagery based BCI user training benefit from more women experimenters?: Throughout MI-BCI use, human supervision (e.g., experimenter or caregiver) plays a central role. While providing emotional and social feedback, people present BCIs to users and ensure smooth users' progress with BCI use. Though, very little is known about the influence experimenters might have on the results obtained. Such influence is to be expected as social and emotional feedback were shown to influence MI-BCI performances. Furthermore, literature from different fields showed an experimenter effect, and specifically of their gender, on experimental outcome. We assessed the impact of the interaction between experimenter and participant gender on MI-BCI performances and progress throughout a session. Our results revealed an interaction between participants gender, experimenter gender and progress over runs. It seems to suggest that women experimenters may positively influence participants' progress compared to men experimenters. This work was published at the International Graz BCI conference in [28].

7.9. Redefining and optimizing BCI user training tasks, stimulations and feedback

Participants: Jelena Mladenovic, Smeethy Pramij, Léa Pillette, Romain Sabau, Fabien Lotte

External collaborators: Jérémy Frey, Jérémie Mattout, Matteus Joffily, Emmanuel Maby, Bertrand Glize, Bernard N'Kaoua, Pierre-Alain Joseph, Camille Jeunet, Roger N'Kambou, Boris Mansencal

Active inference as a unifying, generic and adaptive framework for a P300-based BCI: We proposed the use of a generic, computational framework – Active (Bayesian) Inference to automatically lead the adaptation process in a P300 speller BCI. It adapts through the use of a probabilistic model of the user built upon user's reactions to flashing/spelled letters. Using such observations, at each iteration it updates its beliefs about user intentions, and converges towards a predefined goal, i.e. correctly spelled letters. Active Inference is a recent computational neuroscience approach that models learning and decision making of the brain. As such, by endowing such model to the BCI machine, it enables the machine to adapt in a similar fashion as the brain would. We demonstrate an implementation of Active Inference on a simulated P300-Speller BCI, with real EEG data from 18 subjects. Results demonstrate the ability of Active Inference to yield a significant increase

in bit rate (17%) over state-of-the-art approaches. This work was published in Journal of Neural Engineering in [18].

Towards adaptive and adapted difficulty for MI-BCI user training: We investigated the relationship between the human factors and BCI performance during MI-BCI training. Additionally, we investigated the influence of user personality traits and states on learning the MI skill, i.e., evolution of performance over a session. We conducted a MI experiment in which we influence the user through task difficulty. We acquire data to build a predictive model that could unveil which kind of task is optimal for what kind of user. Moreover, depending on what we set to be predicted, be it a flow state or performance, it can serve as a guide for overall adaptation, i.e., it can serve as an optimization criteria to wager between user experience and system accuracy for instance. We then used priors on user traits and states acquired from the prediction models to perform a simple adaptive method which provides optimal task difficulty to each user. To demonstrate the usefulness of the model for maximizing performance, we perform a simulation using real data from the MI-BCI experiment mentioned above. This work was presented in the PhD thesis of Jelena Mladenovic, that was successfully defended on September 10th, 2019.

Impact of MI-BCI feedback for post-stroke and neurotypical people: We investigated how the modality of the feedback could be adapted to the learners. First, based on a review of the literature, we argued that somatosensory abilities of post-stroke patients have not, but should be, taken into account for BCI-based motor therapies. Indeed, somatosensory abilities play an important role in motor rehabilitation in general, and in BCI-based therapies in particular. It is assumed that during BCI based therapies the co-activation of ascending (i.e., somatosensory) and descending (i.e., sensorimotor) networks enables significant functional motor improvement, together with significant sensorimotor-related neurophysiological changes. Somatosensory abilities seem essential for the patients to benefit from the feedback provided by the BCI system. Yet, around half of post-stroke patients suffer from somatosensory deficits. We hypothesize that these deficits alter their ability to benefit from BCI-based therapies. Our review of the literature on BCI-based motor rehabilitation post-stroke of 14 randomized clinical trials indicates that somatosensory abilities were rarely considered and/or reported. Only two studies over the fourteen reported using them as inclusion/exclusion criteria. Though, none of these two studies reported how they assess the somatosensory abilities. We argue that assessing the somatosensory abilities of the patients is necessary to avoid any bias and enable reliable comparison between-subject and between-study. It could also be leveraged to improve our understanding of the underlying mechanisms of motor recovery and adapt the therapy to the patients' abilities.

Our review of the literature also informed us that a multimodal feedback composed of both somatosensory and visual feedback enables better performances than an unimodal visual feedback, at least in the short term. Though, the long term influence of such feedback remained unknown. Therefore, we assessed the long term effects of a multimodal feedback composed of both vibrotactile and realistic visual stimulations (presented in [43], see also Figure 10), and a unimodal feedback with only realistic visual stimulations. We found that the beneficial impact of a multimodal feedback composed of both visual and somatosensory stimulation compared to a visual feedback alone remains true even for long term training, which had not been tested before. Also, the order of presentation of the different modalities of feedback might have an influence. Using an unimodal visual feedback only seems to be better suited for untrained participants. We hypothesis that integrating information arising from two modalities of feedback while performing the task could be particularly challenging for a novice learner. Both these works were presented in the PhD thesis of Léa Pillette, that was successfully defended on December 16th, 2019.

A physical learning companion for Mental-Imagery BCI User Training: We continued our work on PEANUT, that we designed, implemented and tested, and which is the first learning companion dedicated to providing social presence and emotional feedback during MI-BCI user training. PEANUT provided social presence and emotional support, depending on the performance and progress of the user, through interventions combining both pronounced sentences and facial expressions. It was designed based on the literature, data analyses and user-studies. We notably conducted several online user surveys to identify the desired characteristics of our learning companion in terms of appearance and supporting speech content. From the results of these surveys we notably deduced which should be the characteristics (personal/non-personal, exclamatory/declarative) of the sentences to be used depending on the performance and progression of a

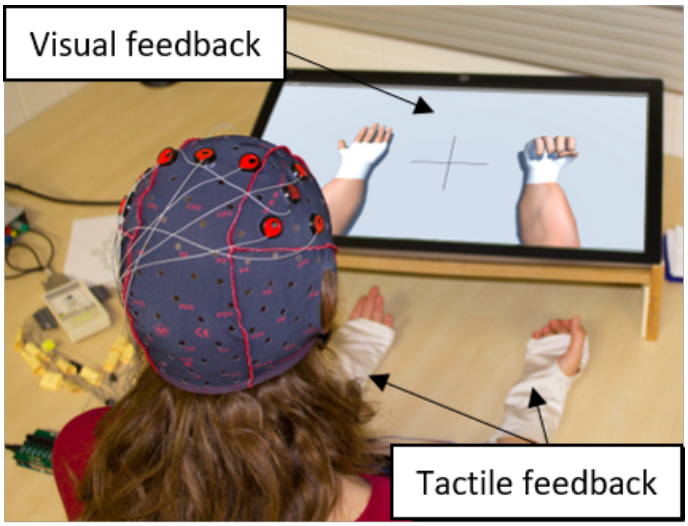


Figure 10. Our multimodal (realistic visual + vibrotactile) feedback for BCI training.

learner. We also found that eyebrows could increase expressiveness of cartoon-like faces. Then, once this companion was implemented, we evaluated it during real online MI-BCI use. We found that non-autonomous people, who are more inclined to work in a group and are usually disadvantaged when using MI-BCI, were advantaged compared to autonomous people when PEANUT was present with an increase of 3.9% of peak performances. Furthermore, in terms of user experience, PEANUT seems to have improved how people felt about their ability to learn and memorize how to use an MI-BCI by 7.4%, which is a dimension of the user experience we assessed. This work was published in the International Journal of Human-Computer Studies in [19].

Long-term mental imagery BCI training of a tetraplegic user: We participated to the Cybathlon BCI series 2019 competition in Graz (https://www.tugraz.at/institutes/ine/graz-bci-conferences/8th-graz-bci-conference-2019/cybathlon-bci-series-2019/), as team NITRO (Neurotechnology Inria Team Racing Odyssey), during which we trained a tetraplegic user over several months, with up to 3 training sessions per week, to learn to control a 4-class and self-paced mental imagery BCI connected to a racing video game (see Figure 11). This training and the resulting BCI design used several of our recent research and development works, notably new OpenViBE development on the feedback, progressive user training and adaptive Riemannian EEG classifiers [41].

7.10. Turning negative into positives! Exploiting “negative” results in Brain-Computer Interface research

Participants: Fabien Lotte

External collaborators: Laurent Bougrain, Ricardo Chavarriaga, Camille Jeunet, Karen Dijkstra, Andrea Kübler, Reinhold Scherer, Moritz Grosse-Wentrup, Natalie Dayan, Dave Thompson, Md Rakibul Mowla

Results that do not confirm expectations are generally referred to as “negative” results. While essential for scientific progress, they are too rarely reported in the literature - BCI research is no exception. This led us to organize a workshop on BCI negative results during the 2018 International BCI meeting. First, we demonstrated why (valid) negative results are useful, and even necessary for BCIs. These results can be used to confirm or disprove current BCI knowledge, or to refine current theories. Second, we provided concrete

Figure 11. BCI-based control of a racing video game by a tetraplegic user during the Cybathlon BCI series in Graz, Austria.

examples of such useful negative results, including the limits in BCI-control for complete locked-in users and predictors of motor imagery BCI performances. Finally, we suggested levers to promote the diffusion of (valid) BCI negative results, e.g., promoting hypothesis-driven research using valid statistical tools, organizing special issues dedicated to BCI negative results, or convincing institutions and editors that negative results are valuable. This work was published in the Brain-Computer Interface journal, in [16].

7.11. Speed of rapid serial visual presentation of pictures, numbers and words affects event-related potential-based detection accuracy

Participants: Fabien Lotte

External collaborators: Stephanie Lees, Paul McCullagh, Liam Maguire, Damien Coyle

Rapid serial visual presentation (RSVP) based brain-computer interfaces (BCIs) can detect target images among a continuous stream of rapidly presented images, by classifying a viewer's event related potentials (ERPs) associated with the target and non-targets images. Whilst the majority of RSVP-BCI studies to date have concentrated on the identification of a single type of image, namely pictures, here we studied the capability of RSVP-BCI to detect three different target image types: pictures, numbers and words. The impact of presentation duration (speed) i.e., 100-200ms (5-10Hz), 200-300ms (3.3-5Hz) or 300-400ms (2.5-3.3Hz), was also investigated. 2-way repeated measure ANOVA on accuracies of detecting targets from non-target stimuli (ratio 1:9) measured via area under the receiver operator characteristics curve (AUC) for N=15 subjects revealed a significant effect of factor Stimulus-Type (pictures, numbers, words) ($F(2,28) = 7.243$, $p = 0.003$) and for Stimulus-Duration ($F(2,28) = 5.591$, $p = 0.011$). Furthermore, there was an interaction between stimulus type and duration: $F(4,56) = 4.419$, $p = 0.004$). The results indicated that when designing RSVP-BCI paradigms, the content of the images and the rate at which images are presented impact on the accuracy of detection and hence these parameters are key experimental variables in protocol design and applications, which apply RSVP for multimodal image datasets. This work was published in IEEE Transactions on Neural Systems and Rehabilitation Engineering, in [15].

7.12. Design and preliminary study of a neurofeedback protocol to self-regulate an EEG marker of drowsiness

Participants: Thibaut Monseigne, Fabien Lotte

External collaborators: Stéphanie Bioulac, Pierre Philip, Jean-Arthur Micoulaud-Franchi

Neurofeedback (NF) consists in using EEG measurements to guide users to perform a cognitive learning using information coming from their own brain activity, by means of a real-time sensory feedback (e.g., visual or auditory). Many NF approaches have been studied to improve attentional abilities, notably for attention deficit hyper activity disorder. However, to our knowledge, no NF solution has been proposed to specifically reduce drowsiness. Thus, we propose an EEG-NF solution to train users to self-regulate an EEG marker of drowsiness, and evaluate it with a preliminary study. Results with five healthy subjects showed that three of them could learn to self-regulate this EEG marker with a relatively short number of NF sessions (up to 8 sessions of 40 min). This work was published at the International Graz BCI conference in [27].

TITANE Project-Team

7. New Results

7.1. Analysis

7.1.1. Pyramid scene parsing network in 3D: improving semantic segmentation of point clouds with multi-scale contextual information

Participants: Hao Fang, Florent Lafarge.

Analyzing and extracting geometric features from 3D data is a fundamental step in 3D scene understanding. Recent works demonstrated that deep learning architectures can operate directly on raw point clouds, i.e. without the use of intermediate grid-like structures. These architectures are however not designed to encode contextual information in-between objects efficiently. Inspired by a global feature aggregation algorithm designed for images, we propose a 3D pyramid module to enrich pointwise features with multi-scale contextual information. Our module can be easily coupled with 3D semantic segmantation methods operating on 3D point clouds. We evaluated our method on three large scale datasets with four baseline models. Experimental results show that the use of enriched features brings significant improvements to the semantic segmentation of indoor and outdoor scenes (See Figure 1). This work was published in the ISPRS journal of Remote Sensing and Photogrammetry [6].

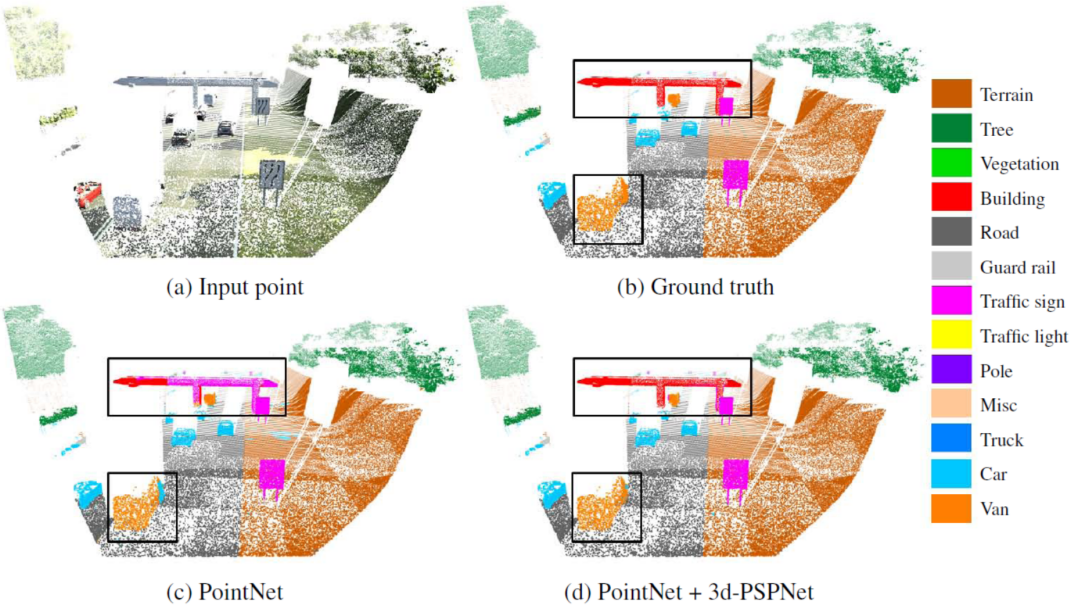


Figure 1. Semantic segmentation of a point cloud with and without our 3d-PSPNet module. Given an input point cloud (a), PointNet fails to predict correct labels for points describing large-scale objects (see rectangles in (c)). PointNet equipped with our 3d-PSPNet module gives better prediction results by enriching global contextual information (d).

7.1.2. Low-power neural networks for semantic segmentation of satellite images

Participants: Gaetan Bahl, Florent Lafarge.

In collaboration with Lionel Daniel and Matthieu Moretti (IRT Saint-Exupéry).

Semantic segmentation methods have made impressive progress with deep learning. However, while achieving higher and higher accuracy, state-of-the-art neural networks overlook the complexity of architectures, which typically feature dozens of millions of trainable parameters. As a result, these networks requires high computational resources and are mostly not suited to perform on edge devices with tight resource constraints, such as phones, drones, or satellites. In this work, we propose two highly-compact neural network architectures for semantic segmentation of images, which are up to 100 000 times less complex than state-of-the-art architectures while approaching their accuracy. To decrease the complexity of existing networks, our main ideas consist in exploiting lightweight encoders and decoders with depth-wise separable convolutions and decreasing memory usage with the removal of skip connections between encoder and decoder. Our architectures are designed to be implemented on a basic FPGA such as the one featured on the Intel Altera Cyclone V family. We demonstrate the potential of our solutions in the case of binary segmentation of remote sensing images, in particular for extracting clouds and trees from RGB satellite images. This work was published in the Low-Power Computer Vision ICCV workshop [13].

7.1.3. A learning approach to evaluate the quality of 3D city models

Participants: Oussama Ennafii, Florent Lafarge.

In collaboration with Arnaud Le Bris and Clément Mallet (IGN).

The automatic generation of 3D building models from geospatial data is now a standard procedure. An abundant literature covers the last two decades and several softwares are now available. However, urban areas are very complex environments. Inevitably, practitioners still have to visually assess, at city-scale, the correctness of these models and detect frequent reconstruction errors. Such a process relies on experts, and is highly time-consuming with approximately two hours per square kilometer for one expert. This work proposes an approach for automatically evaluating the quality of 3D building models. Potential errors are compiled in a novel hierarchical and versatile taxonomy. This allows, for the first time, to disentangle fidelity and modeling errors, whatever the level of detail of the modeled buildings. The quality of models is predicted using the geometric properties of buildings and, when available, Very High Resolution images and Digital Surface Models. A baseline of handcrafted, yet generic, features is fed into a Random Forest classifier. Both multi-class and multi-label cases are considered: due to the interdependence between classes of errors, it is possible to retrieve all errors at the same time while simply predicting correct and erroneous buildings. The proposed framework was tested on three distinct urban areas in France with more than 3,000 buildings. $80\% - 99\%$ F-score values are attained for the most frequent errors. For scalability purposes, the impact of the urban area composition on the error prediction was also studied, in terms of transferability, generalization and representativeness of the classifiers. It shows the necessity of multimodal remote sensing data and mixing training samples from various cities to ensure stability of the detection ratios, even with very limited training set sizes. This work was presented at the IGARSS conference [16] and published in the PE&RS journal [5].

7.1.4. Robust joint image reconstruction from color and monochrome cameras

Participant: Muxingzi Li.

In collaboration with Peihan Tu (Uni. of Maryland) and Wolfgang Heidrich (KAUST).

Recent years have seen an explosion of the number of camera modules integrated into individual consumer mobile devices, including configurations that contain multiple different types of image sensors. One popular configuration is to combine an RGB camera for color imaging with a monochrome camera that has improved performance in low-light settings, as well as some sensitivity in the infrared. In this work we introduce a method to combine simultaneously captured images from such a two-camera stereo system to generate a high-quality, noise reduced color image. To do so, pixel-to-pixel alignment has to be constructed between the two captured monochrome and color images, which however, is prone to artifacts due to parallax. The joint image

reconstruction is made robust by introducing a novel artifact-robust optimization formulation. We provide extensive experimental results based on the two-camera configuration of a commercially available cell phone. This work was presented at the BMVC conference [18].

7.1.5. Noisy supervision for correcting misaligned cadaster maps without perfect Ground Truth data

Participants: Nicolas Girard, Yuliya Tarabalka.

In collaboration with Guillaume Charpiat (Tau Inria project-team).

In machine learning the best performance on a certain task is achieved by fully supervised methods when perfect ground truth labels are available. However, labels are often noisy, especially in remote sensing where manually curated public datasets are rare. We study the multi-modal cadaster map alignment problem for which available annotations are misaligned polygons, resulting in noisy supervision. We subsequently set up a multiple-rounds training scheme which corrects the ground truth annotations at each round to better train the model at the next round. We show that it is possible to reduce the noise of the dataset by iteratively training a better alignment model to correct the annotation alignment. This work was presented at the IGARSS conference [10].

7.1.6. Incremental Learning for Semantic Segmentation of Large-Scale Remote Sensing Data

Participants: Onur Tasar, Pierre Alliez, Yuliya Tarabalka.

In spite of remarkable success of the convolutional neural networks on semantic segmentation, they suffer from catastrophic shortcomings: a significant performance drop for the already learned classes when new classes are added on the data having no annotations for the old classes. We propose an incremental learning methodology, enabling to learn segmenting new classes without hindering dense labeling abilities for the previous classes, although the entire previous data are not accessible. The key points of the proposed approach are adapting the network to learn new as well as old classes on the new training data, and allowing it to remember the previously learned information for the old classes. For adaptation, we keep a frozen copy of the previously trained network, which is used as a memory for the updated network in absence of annotations for the former classes. The updated network minimizes a loss function, which balances the discrepancy between outputs for the previous classes from the memory and updated networks, and the mis-classification rate between outputs for the new classes from the updated network and the new ground-truth. We either regularly feed samples from the stored, small fraction of the previous data or use the memory network, depending on whether the new data are collected from completely different geographic areas or from the same city (see Figure 2). Our experimental results prove that it is possible to add new classes to the network, while maintaining its performance for the previous classes, despite the whole previous training data are not available. This work was published in the IEEE journal of Selected Topics in Applied Earth Observations and Remote Sensing [9].

7.1.7. Multi-Task Deep Learning for Satellite Image Pansharpening and Segmentation

Participants: Onur Tasar, Yuliya Tarabalka.

In collaboration with Andrew Khalel (Cairo University), Guillaume Charpiat (Inria, TAU team)

In this work, we propose a novel multi-task framework, to learn satellite image pansharpening and segmentation jointly (Figure 3). Our framework is based on the encoder-decoder architecture, where both tasks share the same encoder but each one has its own decoder. We compare our framework against single-task models with different architectures. Results show that our framework outperforms all other approaches in both tasks. This work was presented at the IGARSS conference [11].

7.1.8. A Generic Framework for Combining Multiple Segmentations in Geographic Object-Based Image Analysis

Participant: Onur Tasar.

In collaboration with Sébastien Lefèvre (Université Bretagne Sud, IRISA) and David Sheeren (DYNAFOR, University of Toulouse, INRA)

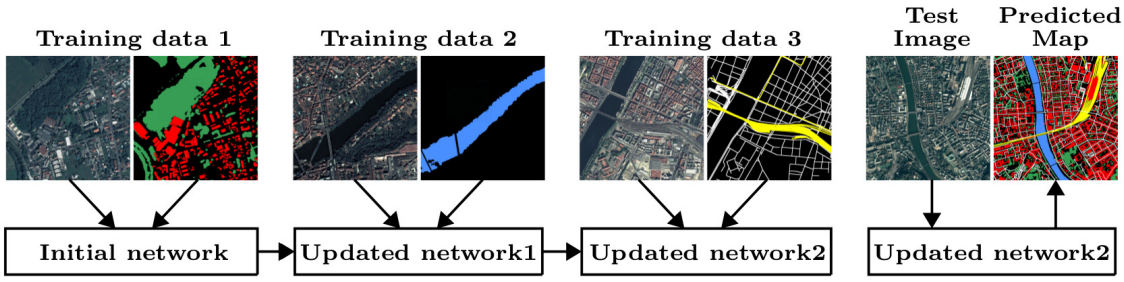


Figure 2. An example of an incremental learning scenario. Firstly, satellite images as well as their label maps for building and high vegetation classes are fed to the network. Then, from the second training data, the network learns the water class without forgetting building and high vegetation classes. Finally, road and railway classes are taught to the network. Whenever new training data are obtained, we store only a small part of the previous ones for the network to remember. When a new test image is provided, the network is able to detect all the classes.

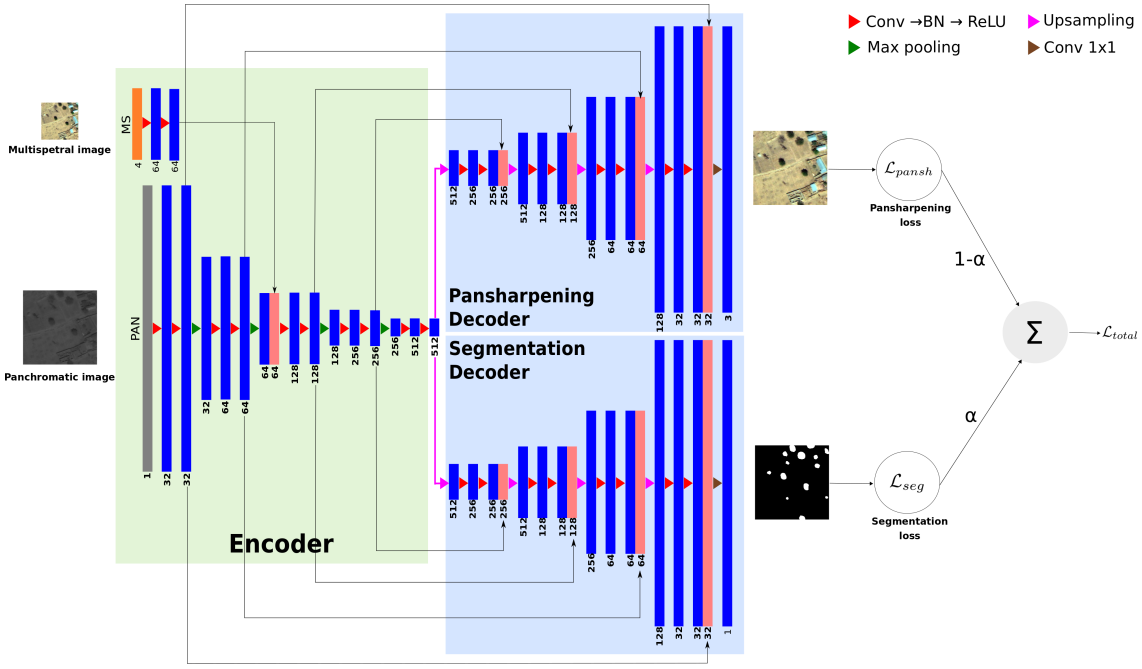


Figure 3. The overall pansharpening and segmentation framework.

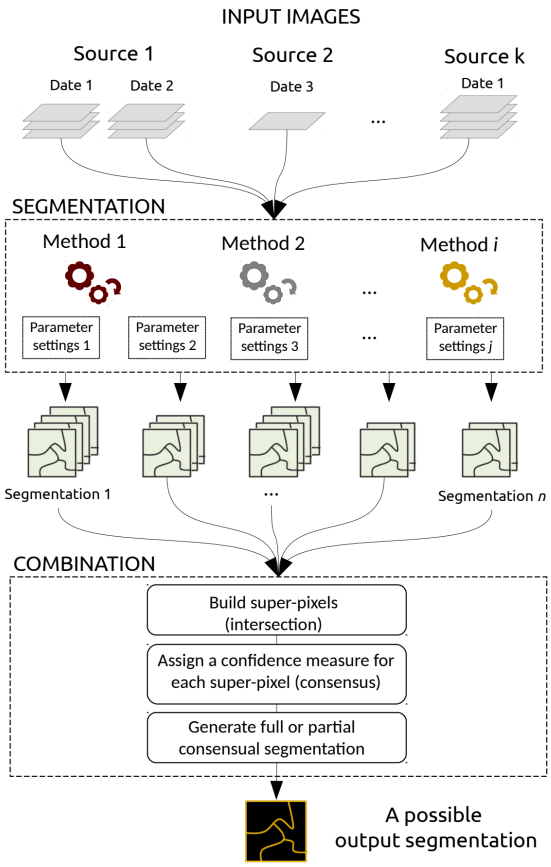


Figure 4. Our generic framework to combine multiple segmentations in the GEOBIA paradigm. Segmentations can come from different data sources (e.g., optical and radar sensors) and be acquired at different dates. They may also be produced using different methods (e.g., region-based or edge-based) relying on different parameter values.

The Geographic Object-Based Image Analysis (GEOBIA) paradigm relies strongly on the segmentation concept, i.e., partitioning of an image into regions or objects that are then further analyzed. Segmentation is a critical step, for which a wide range of methods, parameters and input data are available. To reduce the sensitivity of the GEOBIA process to the segmentation step, here we consider that a set of segmentation maps can be derived from remote sensing data. Inspired by the ensemble paradigm that combines multiple weak classifiers to build a strong one, we propose a novel framework for combining multiple segmentation maps (Figure 4). The combination leads to a fine-grained partition of segments (super-pixels) that is built by intersecting individual input partitions, and each segment is assigned a segmentation confidence score that relates directly to the local consensus between the different segmentation maps. Furthermore, each input segmentation can be assigned some local or global quality score based on expert assessment or automatic analysis. These scores are then taken into account when computing the confidence map that results from the combination of the segmentation processes. This means the process is less affected by incorrect segmentation inputs either at the local scale of a region, or at the global scale of a map. In contrast to related works, the proposed framework is fully generic and does not rely on specific input data to drive the combination process. We assess its relevance through experiments conducted on ISPRS 2D Semantic Labeling. Results show that the confidence map provides valuable information that can be produced when combining segmentations, and fusion at the object level is competitive w.r.t. fusion at the pixel or decision level. This work was published in the ISPRS journal of Geo-Information [8].

7.2. Reconstruction

7.2.1. City Reconstruction from Airborne Lidar: A Computational Geometry Approach

Participants: Jean-Philippe Bauchet, Florent Lafarge.

We introduce a pipeline that reconstructs buildings of urban environments as concise polygonal meshes from airborne LiDAR scans. It consists of three main steps: classification, building contouring, and building reconstruction, the two last steps being achieved using computational geometry tools. Our algorithm demonstrates its robustness, flexibility and scalability by producing accurate and compact 3D models over large and varied urban areas in a few minutes only (See Figure 5). This work was published in the ISPRS international conference 3D GeoInfo [14].

7.2.2. Extracting geometric structures in images with Delaunay point processes

Participant: Florent Lafarge.

In collaboration with Jean-Dominique Favreau (Ekinnox), Adrien Bousseau (GraphDeco Inria team) and Alex Auvolat (Wide Inria team).

We introduce Delaunay Point Processes, a framework for the extraction of geometric structures from images. Our approach simultaneously locates and groups geometric primitives (line segments, triangles) to form extended structures (line networks, polygons) for a variety of image analysis tasks. Similarly to traditional point processes, our approach uses Markov Chain Monte Carlo to minimize an energy that balances fidelity to the input image data with geometric priors on the output structures. However, while existing point processes struggle to model structures composed of interconnected components, we propose to embed the point process into a Delaunay triangulation, which provides high-quality connectivity by construction. We further leverage key properties of the Delaunay triangulation to devise a fast Markov Chain Monte Carlo sampler. We demonstrate the flexibility of our approach on a variety of applications, including line network extraction, object contouring, and mesh-based image compression (See Figure 6). This work was published in the IEEE journal TPAMI [7].

7.3. Approximation

7.3.1. Cost-driven framework for progressive compression of textured meshes

Participants: Cédric Portaneri, Pierre Alliez.

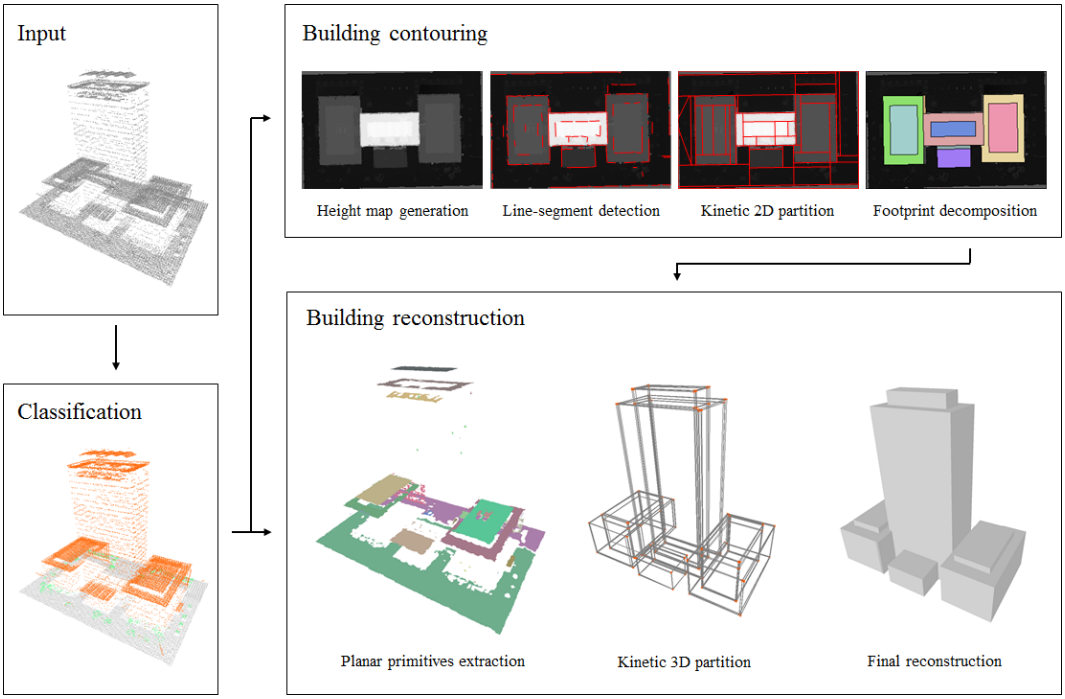


Figure 5. City Reconstruction from Airborne Lidar. Our method consists of three main steps. We first label points of the LiDAR scan as ground, vegetation or roof. Then, we apply a contouring algorithm to the height map, revealing the facades initially absent in the point set. Finally, we extract and propagate planar primitives from the point cloud, dividing the space into polyhedra that are labeled to obtain a 3D reconstruction of buildings.

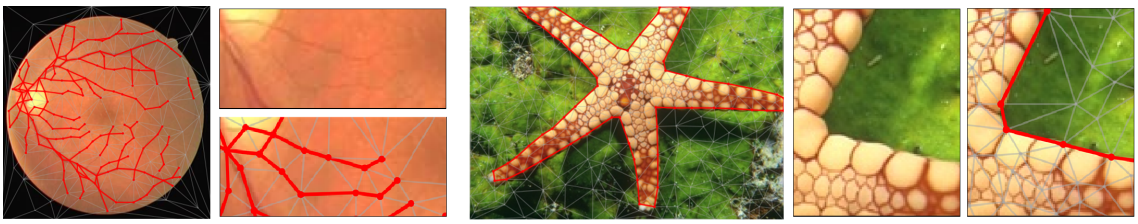


Figure 6. Example applications of Delaunay Point Processes to extract planar graphs representing blood vessels in retina images (left), and complex polygons representing object silhouettes (right). The point distribution creates a dynamic Delaunay triangulation while edge and facet labels specify the geometric structure (see red edges on close-ups).

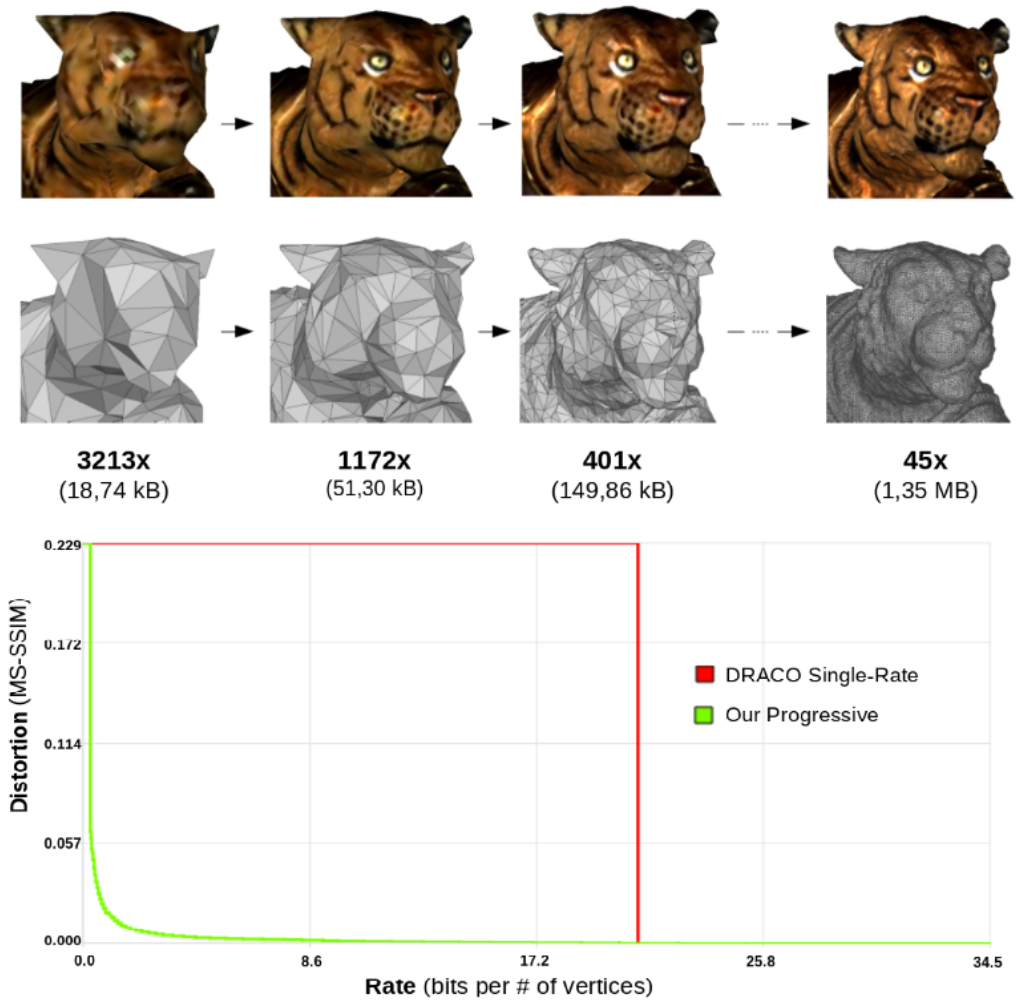


Figure 7. Progressive decompression of a textured surface triangle mesh. Top: key levels of detail with their size and compression rate compared to the raw obj file (texture data not included). Bottom: Distortion against the bit consumption, in bits per vertex, where the number of vertices refers to the input mesh. Green is our progressive approach, red is the state-of-the-art single-rate DRACO encoder.

In collaboration with Michael Hemmer (Google X), Lukas Birklein and Elmar Schoemer (Uni. of Mainz).

Recent advances in digitization of geometry and radiometry generate in routine massive amounts of surface meshes with texture or color attributes. This large amount of data can be compressed using a progressive approach which provides at decoding low complexity levels of details (LoDs) that are continuously refined until retrieving the original model. The goal of such a progressive mesh compression algorithm is to improve the overall quality of the transmission for the user, by optimizing the rate-distortion trade-off. In this paper, we introduce a novel meaningful measure for the cost of a progressive transmission of a textured mesh by observing that the rate-distortion curve is in fact a staircase, which enables an effective comparison and optimization of progressive transmissions in the first place. We contribute a novel generic framework which utilizes the cost function to encode triangle surface meshes via multiplexing several geometry reduction steps (mesh decimation via half-edge or full-edge collapse operators, xyz quantization reduction and uv quantization reduction). This framework can also deal with textures by multiplexing an additional texture reduction step. We also design a texture atlas that enables us to preserve texture seams during decimation while not impairing the quality of resulting LODs. For encoding the inverse mesh decimation steps we further contribute a significant improvement over the state-of-the-art in terms of rate-distortion performance and yields a compression-rate of 22:1, on average. Finally, we propose a unique single-rate alternative solution using a selection scheme of a subset among LODs, optimized for our cost function, and provided with our atlas that enables interleaved progressive texture refinements (see Figure 7). This work was presented at the ACM Multimedia Systems conference [19] and obtained the best paper award.

7.3.2. Selective padding for Polycube-based hexahedral meshing

Participant: Pierre Alliez.

In collaboration with Gianmarco Cherchi and Riccardo Scateni from University of Cagliari (Sardinia), Max Lyon from University of Aachen and David Bommes from University of Bern.

Hexahedral meshes generated from polycube mapping often exhibit a low number of singularities but also poor quality elements located near the surface. It is thus necessary to improve the overall mesh quality, in terms of the minimum Scaled Jacobian (MSJ) or average Scaled Jacobian (ASJ). Improving the quality may be obtained via global padding (or pillowing), which pushes the singularities inside by adding an extra layer of hexahedra on the entire domain boundary. Such a global padding operation suffers from a large increase of complexity, with unnecessary hexahedra added. In addition, the quality of elements near the boundary may decrease. We propose a novel optimization method which inserts sheets of hexahedra so as to perform selective padding, where it is most needed for improving the mesh quality. A sheet can pad part of the domain boundary, traverse the domain and form singularities. Our global formulation, based on solving a binary problem, enables us to control the balance between quality improvement, increase of complexity and number of singularities. We show in a series of experiments that our approach increases the MSJ value and preserves (or even improves) the ASJ, while adding fewer hexahedra than global padding. (See Figure 8). This work was published in an international journal and was presented at the EUROGRAPHICS conference [4].

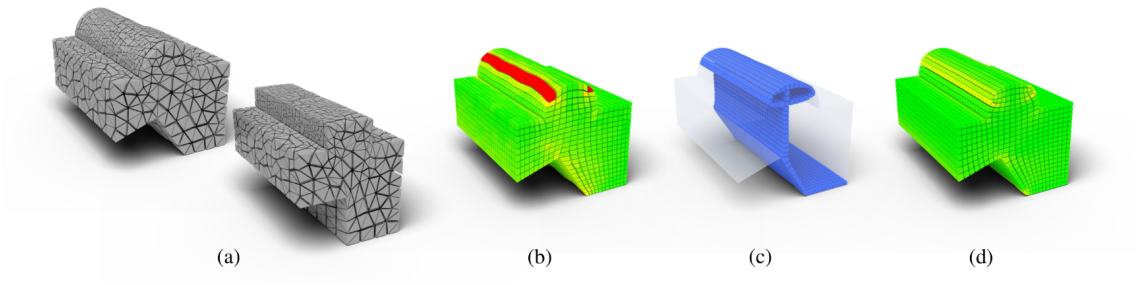


Figure 8. Polycube-based hexahedral meshing. Our pipeline takes as input a model and its polycube mapping (a); we compute the relative hex-mesh and locate the surface areas in need of padding analyzing the mapping quality (b); we set and solve a binary problem to find a set of facets to extrude in order to create a selective padding layer (c); we compute and analyze the mapping with the new hex-mesh structure (d).

ALMANACH Project-Team

7. New Results

7.1. New results on text simplification

Participants: Benoît Sagot, Éric Villemonte de La Clergerie, Louis Martin.

Text simplification (TS) aims at making a text easier to read and understand by simplifying grammar and structure while keeping the underlying meaning and information identical. It is therefore an instance of language variation, based on language complexity. It can benefit numerous audiences, such as people with disabilities, language learners or even everyone, for instance when dealing with intrinsically complex texts such as legal documents.

We have initiated in 2017 a collaboration with the Facebook Artificial Intelligence Research (FAIR) lab in Paris and with the UNAPEI, the federation of French associations helping people with mental disabilities and their families. The objective of this collaboration is to develop tools for helping the simplification of texts aimed at mentally disabled people. More precisely, the is to develop a computer-assisted text simplification platform (as opposed to an automatic TS system). In this context, a CIFRE PhD thesis has started in collaboration with the FAIR on the TS task. We have first dedicated important efforts to the problem of the evaluation of TS systems, which remains an open challenge. As the task has common points with machine translation (MT), TS is often evaluated using MT metrics such as BLEU. However, such metrics require high quality reference data, which is rarely available for TS. TS has the advantage over MT of being a monolingual task, which allows for direct comparisons to be made between the simplified text and its original version. We compared multiple approaches to reference-less quality estimation of sentence-level TS systems, based on the dataset used for the QATS 2016 shared task. We distinguished three different dimensions: grammaticality, meaning preservation and simplicity. We have shown that n-gram-based MT metrics such as BLEU and METEOR correlate the most with human judgement of grammaticality and meaning preservation, whereas simplicity is best evaluated by basic length-based metrics [87]. Our implementations of several metrics have been made this year easily accessible and described in a demo paper in collaboration with the University of Sheffield [16].

In 2019, we have also investigated an important issue inherent to the TS task. Although it is often considered an all-purpose generic task where the same simplification is suitable for all, multiple audiences can benefit from simplified text in different ways. We have therefore introduced a discrete parametrisation mechanism that provides explicit control on TS systems based on Seq2Seq neural models. As a result, users can condition the simplifications returned by a model on parameters such as length and lexical complexity. We also show that carefully chosen values of these parameters allow out-of-the-box Seq2Seq neural models to outperform their standard counterparts on simplification benchmarks. Our best parametrised model improves over the previous state of the art performance [61].

Finally, we are involved in the development of a new text simplification corpus. In order to simplify a sentence, human editors perform multiple rewriting transformations: splitting it into several shorter sentences, paraphrasing (i.e. replacing complex words or phrases by simpler synonyms), reordering components, and/or deleting information deemed unnecessary. Despite the vast range of possible text alterations, current models for automatic sentence simplification are evaluated using datasets that are focused on single transformations, such as paraphrasing or splitting. This makes it impossible to understand the ability of simplification models in more abstractive and realistic settings. This is what motivated the development of ASSET, a new dataset for assessing sentence simplification in English, in collaboration with the University of Sheffield (United Kingdom). ASSET is a crowdsourced multi-reference corpus where each simplification was produced by executing several rewriting transformations. Through quantitative and qualitative experiments, we have shown that simplifications in ASSET are better at capturing characteristics of simplicity when compared to other standard evaluation datasets for the task. Furthermore, we have motivated the need for developing better methods for automatic evaluation using ASSET, since we show that current popular metrics may not be suitable for assessment when multiple simplification transformations were performed.

7.2. NLP and computational neurolinguistics

Participants: Éric Villemonte de La Clergerie, Murielle Fabre.

In the context of the CRCNS international network, the ANR-NSF NCM-ML project (dubbed "*Petit Prince* project") aims to discover and explore correlations between features (or predictors) provided by NLP tools such as parsers, and brain imagery (fMRI) data resulting from listening of the novel Le Petit Prince. Following the availability of an increasing amount of fMRI datasets in French and English, the project has investigated the correlations between fMRI observations and an increasing number of parser-based features based on several parsers representing a number of architecture types (LSTM, RNN, Dyalog-SR [statistical], FRMG [hybrid symbolic/statistical]) [20].

While pursuing the purely computation goal of developing a method of variable beam size inference for Recurrent Neural Network Grammar (RNNG) the project investigated how different beam search methods can show different goodness of fit with fMRI signal recorded during naturalistic story listening [58]. This approach is part of a new trend that is now emerging under the name of cognitively inspired NLP, where the effort to leverage from what we know of human cognition to increase machine processing of language data. Drawing inspiration from sequential Monte-Carlo methods such as particle filtering, we illustrated the relevance of our new method for speeding up the computations of direct generative parsing for RNNG, and revealing the potential cognitive interpretation of the underlying representations built by the search method and its beam activity through the analysis of neuro-imaging signal.

A second focus of the project is on compositionality, memory retrieval and syntactic composition during language comprehension. By using quantifications of these hypothesised processes as obtained from computational linguistics we seek to highlight their neural substrates and better understand or model human cognitition.

While linguistic expressions have been binarised as compositional and non-compositional given the lack of compositional linguistic analysis, the so-called Multi-word Expressions (MWEs) demonstrate finer-grained degrees of conventionalisation and predictability in psycho-linguisitcs, which can be quantified through computational Association Measures, like Point-wise Mutual Information and Dice's Coefficient [57]. An fMRI analysis was conducted to investigate to what extent these computational measures and the underlying cognitive processes they reflect are observable during on-line naturalistic sentence processing. Our results show that predictability, as quantified through Dice's Coefficent, is a better predictor of neural activation for processing MWEs and the more cognitively plausible computational metric. Computational results (1348) were obtained on MWE identification in French based on new method searching for frequent dependency-patterns [13]. These identifications in the Little Prince are contrasted with the ones published for English [69] and will yield an fMRI analysis comparing the two languages and the possible typological differences that the two languages may reflect in terms of morphological strategies to achieve lexical conventionalisation.

7.3. Large-scale raw corpus development

Participants: Benoît Sagot, Éric Villemonte de La Clergerie, Laurent Romary, Pedro Ortiz Suárez, Murielle Fabre, Louis Martin, Benjamin Muller, Yoann Dupont.

In order to be in phase (and comparable) with the US partners of the "Petit-Prince" ANR project, Murielle Fabre assembled two French corpora:

- a small corpus for domain adaptation to children's books: it will permit the fine tuning of the different parsers to a great amount of dialogues and Q&A present in *Le Petit Prince*.
- a large corpus of Contemporary French oral transcriptions and texts to calculate lexical association measures (AM) like PMI (Point-wise Mutual information) or Dice scores on the MWEs found in *Le Petit Prince*. This corpus of approx. 600 millions words, called CaBerNET, represents a balanced counterpart to the American COCA corpus. [0]

[0] https://corpus.byu.edu/coca/

We have also developed a general, highly parallel, multi-threaded pipeline to clean and classify Common Crawl by language. Common Crawl is a huge (over 20TB), heterogeneous multilingual corpus comprised of documents crawled from the internet, not sorted per language. We designed our pipeline, called goclassy, so that it runs efficiently on medium to low resource infrastructures where I/O speeds are the main constraint. We have created and we distribute a 6.3TB version of Common Crawl, called OSCAR, which is filtered, classified by language, shuffled at line level in order to avoid copyright issues, and ready to be used for NLP applications [29]. OSCAR corpora served as input data to train a variety of neural language models, including the French BERT model CamemBERT (see relevant module for more information). Bridging corpus development, NLP and computational neurolinguistics on of our next step is to train BERT model with the above cited French balanced corpus CaBerNet to create CaBERTnet and extract form it parsing metrics that will be correlated with brain activity as measured by French fMRI recording while listening *Le Petit Prince* in French.

7.4. Neural language modelling

Participants: Benoît Sagot, Djamé Seddah, Éric Villemonte de La Clergerie, Laurent Romary, Louis Martin, Benjamin Muller, Pedro Ortiz Suárez, Yoann Dupont, Ganesh Jawahar.

Pretrained language models are now ubiquitous in Natural Language Processing. Despite their success, most available models have either been trained on English data or on the concatenation of data in multiple languages. This makes practical use of such models—in all languages except English—very limited. In 2019, one of the most visible achievements of the ALMAnaCH team was the training and release of CamemBERT, a BERT-like [75] (rather, RoBERTa-like) neural language model for French trained on the French section of our large-scale web-based OSCAR corpus, together with CamemBERT variants [60]. Our goal was to investigate the feasibility of training monolingual Transformer-based language models for other languages, taking French as an example and evaluating our language models on part-of-speech tagging, dependency parsing, named entity recognition and natural language inference tasks. We have shown that the use of web-crawled data such as found in OSCAR to train such language models is preferable to the use of Wikipedia data, because of the homogeneity of Wikipedia data. More surprisingly, we have also shown that a relatively small web crawled dataset (4GB randomly extracted from the French section of OSCAR) leads to results that are as good as those obtained using larger datasets (130+GB, i.e. the whole French section of OSCAR). CamemBERT allowed us to reach or improve the state of the art in all four downstream tasks.

Beyond training neural language models, we have reinforced the exploration of an active question, that of their interpretability. With the emergence of contextual vector representations of words, such as the ELMo [89] and BERT language models and word embeddings, the interpretability of neural models becomes a key research topic. It is a way to understand what such neural networks actually learn in an unsupervised way from (huge amounts of) textual data, and in which circumstances they manage to do so. The work carried out in the team this year to identify where morphological vs. syntactic vs. semantic information is stored in a BERT language model [26] was part of a more general trend (see for example [78]). And our work on training ELMo models for five mid-resourced languages has shown that such LSTM-based models, when trained on large scale although non edited dataset such as our web-based corpora OSCAR, can lead to outperforming state-of-the-art performance on an number of downstream tasks such as part-of-speech tagging and parsing. Finally, we have carried out comparative evaluations of the performance of CamemBERT and of ELMo models trained on the same French section of OSCAR on a number of downstream task, with an emphasis on named-entity recognition—a work that led us to publish a new version of the named-entity-annotated version of the French TreeBank [67] that we published in 2012 [99].

We have also investigated how word embeddings can capture the evolution of word usage and meaning over time, at a fine-grained scale. As part of the ANR SoSweet and the PHC Maimonide projects (in collaboration with Bar Ilan University for the latter), ALMAnaCH has invested a lot of efforts since 2018 into studying language variation within user-generated content (UGC), taking into account two main interrelated dimensions: how language variation is related to socio-demographic and dynamic network variables, and how UGC language evolves over time. Taking advantage of the SoSweet corpus (600 millions tweet) and of the Bar Ilan Hebrew Tweets (180M tweets) both collected over the last 5 years, we have been addressing the problem

of studying semantic changes via the use of dynamic word embeddings, that is embeddings evolving over time. We devised a novel attention model, based on Bernouilli word embeddings, that are conditioned on contextual extra-linguistic features such as network, spatial and socio-economic variables, which can be inferred from Twitter users metadata, as well as topic-based features. We posit that these social features provide an inductive bias that is susceptible to helping our model to overcome the narrow time-span regime problem. Our extensive experiments reveal that, as a result of being less biased towards frequency cues, our proposed model was able to capture subtle semantic shifts and therefore benefits from the inclusion of a reduced set of contextual features. Our model thus fit the data better than current state-of-the-art dynamic word embedding models and therefore is a promising tool to study diachronic semantic changes over small time periods. We published these ideas and results in [41].

A deep understanding of what is learned, and, beyond that, of how it is learned by neural language models, both synchronic and diachronic, will be a crucial step towards the improvement of such architectures (e.g. targeting low-resource languages or scenarios) and the design and deployment of new generations of neural networks for NLP. Particularly important is to assess the role of the training corpus size and heterogeneity, as well as the impact of the properties of the language at hand (e.g. morphological richness, token-type ratio, etc.). This line of research will also have an impact on our understanding of language variation and on our ability to improve the robustness of neural-network-based NLP tools to such variation.

7.5. Processing non-standard language: user-generated content and code-mixed language

Participants: Djamé Seddah, Benoît Sagot, Éric Villemonte de La Clergerie, Benjamin Muller, Ganesh Jawahar, Abhishek Srivastava, Jose Rosales Nuñez, Hafida Le Cloirec, Farah Essaidi, Matthieu Futeral.

In 2019, we have resumed our long-lasting efforts towards increasing the robustness of our language analysis tools to the variation found in user-generated content (UGC). We have done this in two directions, in the context of the SoSweet and Parsiti projects.

Firstly, we have investigated how our state-of-the-art hybrid (symbolic and statistical) parsing architecture for French, based on SxPipe, FRMG and the Lefff, behaves on French UGC data, namely on 20 millions tweets from the SoSweet corpus. A first observation was that the current level of pre-parsing normalization was not sufficient to ensure a good parsing coverage with FRMG (around 67%, to be compared with around 93% on journalistic texts such as the French TreeBank), also leading to high parsing times because of correction strategies. However, we applied our error mining strategy [6] to identify a first set of easy errors. Clustering and word embedding were also tried for lemmas relying on the dependency parse trees, again leading to semi-successful results due to the poor quality of the pre-parsing phases.

Secondly, we have investigated the normalisation task, whose goal is to transform possibly noisy UGC into less noisy inputs that are more adapted to our standard neural analysis models (e.g. taggers and parsers). More precisely, we have investigated how useful a language model such as BERT [75], trained on standard data, can be in handling non-canonical text. We study the ability of BERT to perform lexical normalisation in a realistic, and therefore low-resource, English UGC scenario [28]. By framing lexical normalisation as a token prediction task, by enhancing its architecture and by carefully fine-tuning it, we have shown that BERT can be a competitive lexical normalisation model without the need of any UGC resources aside from 3,000 training sentences. To the best of our knowledge, it is the first work done in adapting and analysing the ability of this model to handle noisy UGC data.

Thirdly, we have compared the performances achieved by Phrase-Based Statistical Machine Translation systems (PBSMT) and attention-based Neural Machine Translation systems (NMT) when translating UGC from French to English [44]. We have shown that, contrary to what could have been expected, PBSMT outperforms NMT when translating non-canonical inputs. Our error analysis uncovers the specificities of UGC that are problematic for sequential NMT architectures and suggests new avenue for improving NMT models.

Finally, building natural language processing systems for highly variable and low resource languages is a hard challenge. The recent success of large-scale multilingual pretrained neural language models (including our CamemBERT language model for French) provides us with new modeling tools to tackle it. We have studied the ability of the multilingual version of BERT to model an unseen dialect, namely the Latin-script user-generated North African Arabic dialect called Arabizi. We have shown in different scenarios that multilingual language models are able to transfer to such an unseen dialect, specifically in two extreme cases: across script (Arabic to Latin) and from Maltese, a related language written in the Arabic script, unseen during pretraining. Preliminary results have already been published [66].

7.6. Long-range diachronic variation

Participants: Benoît Sagot, Laurent Romary, Éric Villemonte de La Clergerie, Clémentine Fourrier, Gaël Guibon, Mathilde Regnault, Kim Gerdes.

ALMAnaCH members have resumed their work on longer-range diachronic variation, in two distinct directions:

- Firstly, we have been working on resources and tools for Old French, using contemporary French as a starting point for which resources and tools are available. This work is carried out within the ANR project "Profiterole", whose goal is to automatically annotate a large corpus of medieval French (9th-15th centuries) in dependency syntax and to provide a methodology for dealing with heterogeneous data as found in such a corpus. Indeed, Old French does not only involve diachronic variation when contrasted with contemporary French. It also involve large internal variation, notably because of diachronic (within Old French), dialectal, geographic, stylistic and genre-based variation. We have carried out experiments on morphosyntactic tagging by trying to determine which parameters and which training sets are the best ones to use when annotating a new text. We explored two approaches for parsing. On the one hand, an ongoing thesis aims at adapting the FRMG metagrammar to medieval French, notably by changing the constraints on certain syntactic phenomena and relaxing the order of words [31], [30]. This work relies on the new morphological and syntactic lexicon for Old French, OFrLex, developed at ALMAnaCH [34]. On the other hand, we conducted parsing experiments with neural models (DyALog's SRNN models).
- Secondly, we have started experiments to investigate whether and under which conditions neural networks can be used for learning sound correspondences between two related languages, i.e. for predicting cognates of source language words in a related target language. In order to obtain suitably large homogeneously phonetised data, we extracted bilingual lexicons and cognate sets from available resources, including our EtymDB etymological database, of which a new, extended version was created in 2019. This data was then used to train and evaluate several neural architectures (seq2seq, Siamese). Preliminary results are promising, but further investigation is required.

These two research directions will find a common ground now that we have begun to investigate, in the context of the Profiterole ANR project, how we can model the diachronic evolution of the lexicon from Old French to contemporary French. Moreover, our work on Basnage's 1701 *Dictionnaire Universel*, in the context of the BASNUM ANR project might draw some inspiration from the Profiterole project. But since 1700's French is much closer from contemporary French than Old French, another source of inspiration for BASNAGE might come from our work on sociolinguistic variation in contemporary French and more generally on our work on User-Generated Content (UCG).

7.7. Syntax and treebanking

Participants: Djamé Seddah, Benoît Sagot, Kim Gerdes, Benjamin Muller, Pedro Ortiz Suárez, Marine Courtin.

In 2019 we have introduced the first treebank for a romanized user-generated content of Algerian, a North-African Arabic dialect called Arabizi. It contains 1500 sentences, fully annotated in morpho-syntax and universal dependencies, and is freely available. We complement it with 50k unlabeled sentences that were collected using intensive data-mining techniques from Common Crawl and web-crawled data. Preliminary results show its usefulness for POS tagging and dependency parsing.

We have also developed the first syntactic treebank for spoken Naija, an English pidgincreole, which is rapidly spreading across Nigeria. The syntactic annotation is developed in the Surface-Syntactic Universal Dependency annotation scheme (SUD) [77] and automatically converted into Universal Dependencies (UD). A crucial step in the syntactic analysis of a spoken language consists in manually adding a markup onto the transcription, indicating the segmentation into major syntactic units and their internal structure. We have shown that this so-called "macrosyntactic" markup improves parsing results. We have also studied some iconic syntactic phenomena that clearly distinguish Naija from English. This work is published in [36].

We have carried out two pilot studies in empirical syntax based on UD treebanks. In a first study [38], we investigate the relationship between dependency distance and frequency based on the analysis of an English dependency treebank. The preliminary result shows that there is a non-linear relation between dependency distance and frequency. This relation between them can be further formalised as a power law function which can be used to predict the distribution of dependency distance in a treebank. In a second study [40], we discussed an empirical refoundation of selected Greenbergian word order univer-sals based on a data analysis of the Universal Dependencies project. The nature of the data we worked on allows us to extract rich details for testing well-known typological universals and constitutes therefore a valuable basis for validating Greenberg's universals. Our results show that we can refine some Greenbergian universals in a more empirical and accurate way by means of a data-driven typological analysis.

Finally, we have introduced a new schema to annotate Chinese Treebanks on the character level. The original UD and SUD projects provide token-level resources with rich morphosyntactic language details. However, without any commonly accepted word definition for Chinese, the dependency parsing always faces the dilemma of word segmentation. Therefore we have presented a character-level annotation schema integrated into the existing Universal Dependencies schema as an extension [39]. The different SUD projects were also presented at the Journées scientifiques "Linguistique informatique, formelle et de terrain" (LIFT 2019), Nov 28-29, 2019 at the University of Orléans.

7.8. Analysing and enriching legacy dictionaries

Participants: Laurent Romary, Benoît Sagot, Mohamed Khemakhem, Pedro Ortiz Suárez, Achraf Azhar.

2019 has been a year of deployment and large-scale experiment of the work initiated in 2016 on the analysis and enrichment of legacy dictionaries and implemented in the GROBID-dictionary framework [84]. GROBID-dictionary is an extension of the generic GROBID Suite [95] and implements an architecture of cascading CRF models with the purpose to parse and categorize components of a pdf documents, whether born-digital or resulting from an OCR. It is developped as part of the doctoral work of Mohamed Khemakhem. GROBID dictionaries produces an output that is conformant to the Text Encoding Initiative guideline and thus easy to distribute and further process in an open science context. We have had the opportunity the show the performances and robustness of the architecture on a variety of dictionaries and contexts resulting both from internal and external collaborations:

- In the context of the language documentation project of Jack Bowers dealing with Mixtepec-Mixtec (ISO 639-3: mix, [72], we have been successful in completely parsing a new edition of an historical lexical resource of Colonial Mixtec 'Voces del Dzaha Dzahui' published by the Dominican fray Francisco Alvarado in the year 1593, published by Jansen and Perez Jiménez (2009). The result is now integrated into the reference lexical description maintained by Jack). See [18];
- Within the Nénufar project, a collaboration with the Praxiling laboratory in Montpellier, we have been contributing to the analyses and encoding of several editions of the Petit Larousse Illustré, a central legacy publication for the French language. [17], [27];
- For the ANR funded project BASNUM, we are deeply involved in understanding how a complex, semi-structured dictionary, for which we do not necessarily have a high quality digitized primary source, can be properly segmented in lexical entries and subfields from which we expect being able to extract fine-grained linguistic content (e.g. named entities for literary sources). In [42], we have shown for instant how the GROBID-dictionary framework could be robust to variations in scanning and thus OCR quality;

- In the same context of the BASNUM project, we have also started to explore the possibility of deploying deep learning components. As shown in [43], the main challenges is the lack of available annotated data in order to train machine learning models, decreased accuracy when using modern pre-trained models due to the differences between present-day and 18th century French, and even unreliable or low quality OCRisation;
- These various experiments have been accompanied by an intense training and hand-on activity in the context in particular of the Lexical Data Master Class and collaboration within the ELEXIS project, which has opted for using the system for building a dictionary matrix from legacy dictionaries [0].Further alignments with the ongoing standardisation activities around TEI Lex0 and ISO 24613 (LMF) has been carried out to ensure a proper standards compliance of the generated output.
- Finally, and as a nice example of the kind of DH collaborations that our researches can lead to, we should mention here the targeted experiments that we carried out on extending the GROBID-dictionary framework to deal with objects which, although analogous with dictionary entries from a distance, appear to have a highly specific structure. This is the case Manuscript Sales Catalogues, which are highly important for authenticating documents and studying the reception of authors. Their regular publication throughout Europe since the beginning of the 19th c. has raised the interest around scaling up the means for automatically structuring their contents. [33] presents the results of advanced tests of the system's capacity to handle a large corpus with MSC of different dealers, and therefore multiple layouts.

7.9. Coreference resolution

Participants: Loïc Grobol, Éric Villemonte de La Clergerie.

In 2019 we have resumed our work on coreference resolution for French with the release in [25] of the first end-to-end automatic coreference resolution system for spoken French by adapting state-of-the art neural network system to the case of noisy non-standard inputs.

This first release uses no external knowledge beyond pretrained non-contextual word embeddings, making it suitable for applications to languages with less pre-existing resources. We also investigated the integration of further knowledge, both in the form of contextual embedding techniques such as CamemBERT and syntactic parsers developped at ALMAnaCH (works to be published in 2020).

[0] https://grobid.elex.is

COML Team

6. New Results

6.1. Unsupervised learning

Huamns learn to speak and to perceive the world in a largely self-supervised fashion. Yet, most of machine learning is still devoted to supervised algorithms that rely on abundant quantities of human labelled data. We have used humans as sources of inspiration for developing 3 novel machine learning benchmarks in order to push the field towards self-supervised learning.

- In the Zero Resource Speech Challenge 2019 [19], presented as a special session at Interspeech 2019, we propose to build a speech synthesizer without any text or phonetic labels: hence, TTS without T (text-to-speech without text). We provide raw audio for a target voice in an unknown language (the Voice dataset), but no alignment, text or labels. Participants must discover subword units in an unsupervised way (using the Unit Discovery dataset) and align them to the voice recordings in a way that works best for the purpose of synthesizing novel utterances from novel speakers, similar to the target speaker's voice. We describe the metrics used for evaluation, a baseline system consisting of unsupervised subword unit discovery plus a standard TTS system, and a topline TTS using gold phoneme transcriptions. We present an overview of the 19 submitted systems from 11 teams and discuss the main results.
- In [27], we introduce a new collection of spoken English audio suitable for training speech recognition systems under limited or no supervision. It is derived from open-source audio books from the LibriVox project. It contains over 60K hours of audio, which is, to our knowledge, the largest freely-available corpus of speech. The audio has been segmented using voice activity detection and is tagged with SNR, speaker ID and genre descriptions. Additionally, we provide baseline systems and evaluation metrics working under three settings: (1) the zero resource/unsupervised setting (ABX), (2) the semi-supervised setting (PER, CER) and (3) the distant supervision setting (WER). Settings (2) and (3) use limited textual resources (10 minutes to 10 hours) aligned with the speech. Setting (3) uses large amounts of unaligned text. They are evaluated on the standard LibriSpeech dev and test sets for comparison with the supervised state-of-the-art.
- In order to reach human performance on complex visual tasks, artificial systems need to incorporate a significant amount of understanding of the world in terms of macroscopic objects, movements, forces, etc. Inspired by work on intuitive physics in infants, we propose in [28] an evaluation framework which diagnoses how much a given system understands about physics by testing whether it can tell apart well matched videos of possible versus impossible events. The test requires systems to compute a physical plausibility score over an entire video. It is free of bias and can test a range of specific physical reasoning skills. We then describe the first release of a benchmark dataset aimed at learning intuitive physics in an unsupervised way, using videos constructed with a game engine. We describe two Deep Neural Network baseline systems trained with a future frame prediction objective and tested on the possible versus impossible discrimination task. The analysis of their results compared to human data gives novel insights in the potentials and limitations of next frame prediction architectures. This benchmark is currently being used in the DARPA project Machine Common Sense.

6.2. Language emergence in communicative agents

In this relatively new research topic, which is currently the focus of Rahma Chaabouni's PhD thesis, we study the inductive biases of neural systems by presenting them with few or no data.

- In [18], we study LSTMs' biases with respect to "natural" word-order constraints. To this end, we train them to communicate about trajectories in a grid world, using an artificial language that reflect or violate various natural language trends, such as the tendency to avoid redundancy or to minimize long-distance dependencies. We measure the speed of individual learning and the generational stability of language patterns in an iterative learning setting. Our results show a mixed picture. If LSTMs are affected by some "natural" word-order constraints, such as a preference for iconic orders and short-distance constructions, they have a preference toward redundant languages.
- In [25], we ask whether LSTMs have least-effort constraints and how this can affect their language. We let the neural systems develop their own language, to study a fundamental characteristic of natural language; Zipf's Law of Abbreviation (ZLA). In other words, we inverstigate if, even with the lack of the least-effort, LSTMs would produce a ZLA-like distribution like what we observe in natural language. Surprisingly, we find that networks develop an anti-efficient encoding scheme, in which the most frequent inputs are associated to the longest messages, and messages in general are skewed towards the maximum length threshold. This anti-efficient code appears easier to discriminate for the listener, and, unlike in human communication, the speaker does not impose a contrasting least-effort pressure towards brevity, as observed in [18]. Indeed, when the cost function includes a penalty for longer messages, the resulting message distribution starts respecting (ZLA). Our analysis stresses the importance of studying the basic features of emergent communication in a highly controlled setup, to ensure the latter will not strand too far from human language. Moreover, we present a concrete illustration of how different functional pressures can lead to successful communication codes that lack basic properties of human language, thus highlighting the role such pressures play in the latter.
- There is renewed interest in simulating language emergence among deep neural agents that communicate to jointly solve a task, spurred by the practical aim to develop language-enabled interactive AIs, and by theoretical questions about the evolution of human language. However, optimizing deep architectures connected by a discrete communication channel (such as that in which language emerges) is technically challenging. In [21], we introduce EGG, a toolkit that greatly simplifies the implementation of emergent-language communication experiments. EGG's modular design provides a set of building blocks that the user can combine to create new communication games, easily navigating the optimization and architecture space. We hope that the tool will lower the technical barrier, and encourage researchers from various backgrounds to do original work in this exciting area/

6.3. Evaluation of AI algorithms

Machine learning algorithms are typically evaluated in terms of end-to-end tasks, but it is very often difficult to get a grasp of how they achieve these tasks, what could be their break point, and more generally, how they would compare to the algorithms used by humans to do the same tasks. This is especially true of Deep Learning systems which are particularly opaque. The team develops evaluation methods based on psycholinguistic/linguistic criteria, and deploy them for systematic comparison of systems.

- Recurrent neural networks (RNNs) can learn continuous vector representations of symbolic structures such as sequences and sentences; these representations often exhibit linear regularities (analogies). Such regularities motivate our hypothesis that RNNs that show such regularities implicitly compile symbolic structures into tensor product representations (TPRs; Smolensky, 1990), which additively combine tensor products of vectors representing roles (e.g., sequence positions) and vectors representing fillers (e.g., particular words). To test this hypothesis, we introduce Tensor Product Decomposition Networks (TPDNs), which use TPRs to approximate existing vector representations. We demonstrate using synthetic data that TPDNs can successfully approximate linear and tree-based RNN autoencoder representations, suggesting that these representations exhibit interpretable compositional structure; we explore the settings that lead RNNs to induce such structure-sensitive representations. By contrast, further TPDN experiments show that the representations of four models trained to encode naturally-occurring sentences can be largely approximated with a bag of words,

with only marginal improvements from more sophisticated structures. We conclude that TPDNs provide a powerful method for interpreting vector representations, and that standard RNNs can induce compositional sequence representations that are remarkably well approximated by TPRs; at the same time, existing training tasks for sentence representation learning may not be sufficient for inducing robust structural representations.

- LSTMs have proven very successful at language modeling. However, it remains unclear to what extent they are able to capture complex morphosyntactic structures. In [29], we examine whether LSTMs are sensitive to verb argument structures. We introduce a German grammaticality dataset in which ungrammatical sentences are constructed by manipulating case assignments (eg substituting nominative by accusative or dative). We find that LSTMs are better than chance in detecting incorrect argument structures and slightly worse than humans tested on the same dataset. Surprisingly, LSTMs are contaminated by heuristics not found in humans like a preference toward nominative noun phrases. In other respects they show human-similar results like biases for particular orders of case assignments.
- Pater (2019) proposes to use neural networks to model learning within existing grammatical frameworks. In [16] we argue that there is a fundamental gap to be bridged that does not receive enough attention : how can we use neural networks to examine whether it is possible to learn some linguistic representation (a tree, for example) when, after learning is finished, we cannot even tell if this is the type of representation that has been learned (all we see is a sequence of numbers)? Drawing a correspondence between an abstract linguistic representational system and an opaque parameter vector that can (or perhaps cannot) be seen as an instance of such a representation is an implementational mapping problem. Rather than relying on existing frameworks that propose partial solutions to this problem, such as harmonic grammar, we suggest that fusional research of the kind proposed needs to directly address how to 'find' linguistic representations in neural network representations.

6.4. Learnability relevant descriptions of linguistic corpora

Evidently, infants are acquiring their language based on whatever linguistic input is available around them. The extent of variation that can be found across languages, cultures and socio-economic background provides strong constraints (lower bounds on data, higher bounds on noise, and variation and ambiguity) for language learning algorithms.

- Previous computational modeling suggests it is much easier to segment words from child-directed (CDS) than adult-directed speech (ADS). However, this conclusion is based on data collected in the laboratory, with CDS from play sessions and ADS between a parent and an experimenter, which may not be representative of ecologically-collected CDS and ADS. In [15], fully naturalistic ADS and CDS collected with a non-intrusive recording device as the child went about her day were analyzed with a diverse set of algorithms. The difference between registers was small compared to differences between algorithms, it reduced when corpora were matched, and it even reversed under some conditions. These results highlight the interest of studying learnability using naturalistic corpora and diverse algorithmic definitions.
- A number of unsupervised learning algorithms have been proposed in the last 20 years for modeling early word learning, some of which have been implemented computationally, but whose results remain difficult to compare across papers. In [14], we created a tool that is open source, enables reproducible results, and encourages cumulative science in this domain. WordSeg has a modular architecture: It combines a set of corpora description routines, multiple algorithms varying in complexity and cognitive assumptions (including several that were not publicly available, or insufficiently documented), and a rich evaluation package. In the paper, we illustrate the use of this package by analyzing a corpus of child-directed speech in various ways, which further allows us to make recommendations for experimental design of follow-up work. Supplementary materials allow readers to reproduce every result in this paper, and detailed online instructions further enable them to go

beyond what we have done. Moreover, the system can be installed within container software that ensures a stable and reliable environment. Finally, by virtue of its modular architecture and transparency, WordSeg can work as an open-source platform, to which other researchers can add their own segmentation algorithms.

6.5. Test of the psychological validity of AI algorithms.

In this section, we focus on the utilisation of machine learning algorithms of speech and language processing to derive testable quantitative predictions in humans (adults or infants).

- In [24], we compare the performance of humans (English and French listeners) versus an unsupervised speech model in a perception experiment (ABX discrimination task). Although the ABX task has been used for acoustic model evaluation in previous research, the results have not, until now, been compared directly with human behaviour in an experiment. We show that a standard, well-performing model (DPGMM) has better accuracy at predicting human responses than the acoustic baseline. The model also shows a native language effect, better resembling native listeners of the language on which it was trained. However, the native language effect shown by the models is different than the one shown by the human listeners, and, notably , the models do not show the same overall patterns of vowel confusions.
- Word learning relies on the ability to master the sound con-trasts that are phonemic (i.e., signal meaning difference) in a given language. Though the timeline of phoneme development has been studied extensively over the past few decades, the mechanism of this development is poorly understood. In [20], we take inspiration from computational modeling work in language grounding where phonetic and visual information is learned jointly. In this study, we varied the taxonomic distance of pairs of objects and tested how adult learners judged the phonemic status of the sound contrast associated with each of these pairs. We found that judgments were sensitive to gradients in the taxonomic structure, suggesting that learners use probabilistic information at the semantic level to optimize the accuracy of their judgements at the phonological level. The findings provide evidence for an interaction between phonological learning and meaning generalization in human learning.

6.6. Applications and tools for researchers

Some of CoMLs' activity is to produce speech and language technology tools that facilitate research into language development or clinical applications.

- Speech classifiers of paralinguistic traits traditionally learn from diverse hand-crafted low-level features, by selecting the relevant information for the task at hand. We explore an alternative to this selection, by learning jointly the classifier, and the feature extraction. Recent work on speech recognition has shown improved performance over speech features by learning from the waveform. In [24], we extend this approach to paralinguistic classification and propose a neural network that can learn a filterbank, a normalization factor and a compression power from the raw speech, jointly with the rest of the architecture. We apply this model to dysarthria detection from sentence-level audio recordings. Starting from a strong attention-based baseline on which mel-filterbanks out-perform standard low-level descriptors, we show that learning the filters or the normalization and compression improves over fixed features by 10% absolute accuracy. We also observe a gain over OpenSmile features by learning jointly the feature extraction, the normalization, and the compression factor with the architecture. This constitutes a first attempt at learning jointly all these operations from raw audio for a speech classification task.
- This paper [23] presents the problems and solutions addressed at the JSALT workshop when using a single microphone for speaker detection in adverse scenarios. The main focus was to tackle a wide range of conditions that go from meetings to wild speech. We describe the research threads we explored and a set of modules that was successful for these scenarios. The ultimate goal was to explore speaker detection; but our first finding was that an effective diarization improves detection,

and not having a diarization stage impoverishes the performance. All the different configurations of our research agree on this fact and follow a main backbone that includes diarization as a previous stage. With this backbone, we analyzed the following problems: voice activity detection, how to deal with noisy signals, domain mismatch, how to improve the clustering; and the overall impact of previous stages in the final speaker detection. In this paper, we show partial results for speaker diarizarion to have a better understanding of the problem and we present the final results for speaker detection.

MULTISPEECH Project-Team

7. New Results

7.1. Beyond black-box supervised learning

Participants: Emmanuel Vincent, Denis Jouvet, Antoine Deleforge, Vincent Colotte, Irène Illina, Romain Serizel, Imran Sheikh, Pierre Champion, Adrien Dufraux, Ajinkya Kulkarni, Manuel Pariente, Georgios Zervakis, Zaineb Chelly Dagdia, Mehmet Ali Tugtekin Turan, Brij Mohan Lal Srivastava.

This year marked a significant increase in our research activities on domain-agnostic challenges relating to deep learning, such as the integration of domain knowledge, data efficiency, or privacy preservation. Our vision was illustrated by a keynote [18] and several talks [19], [17] on the key challenges and solutions.

7.1.1. Integrating domain knowledge

7.1.1.1. Integration of signal processing knowledge

State-of-the-art methods for single-channel speech enhancement or separation are based on end-to-end neural networks including learned real-valued filterbanks. We tackled two limitations of this approach. First, to ensure that the representation properly encodes phase properties as the short time Fourier transform and other conventional time-frequency transforms, we designed complex-valued analytic learned filterbanks and defined corresponding representations and masking strategies which outperformed the popular ConvTasNet algorithm [59]. Second, in order to allow generalization to mixtures of sources not seen together in training, we explored the modeling of speech spectra by variational autoencoders (VAEs), which are a variant of the probabilistic generative models classically used in source separation before the deep learning era. The VAEs are trained separately for each source and used to infer the source signals underlying a given mixture. Compared with existing iterative inference algorithms involving Gibbs sampling or gradient descent, we proposed a computationally efficient variational inference method based on an analytical derivation in which the encoder of the pre-learned VAE can be used to estimate the variational approximation of the true posterior [42], [55].

7.1.2. Learning from little/no labeled data

7.1.2.1. Learning from noisy labels

ASR systems are typically trained in a supervised fashion using manually labeled data. This labeling process incurs a high cost. Classical semi-supervised learning and transfer learning approaches to reduce the transcription cost achieve limited performance because the amount of knowledge that can be inferred from unlabeled data is intrinsically lower. We explored the middle ground where the training data are neither accurately labeled nor unlabeled but a not-so-expensive "noisy" transcription is available instead. We proposed a method to learn an end-to-end ASR model given a noise model and a single noisy transcription per utterance by adapting the auto segmentation criterion (ASG) loss to account for several possible transcriptions. Because the computation of this loss is intractable, we used a differentiable beam search algorithm that samples only the best alignments of the best transcriptions [32].

7.1.2.2. Transfer learning

We worked on the disentanglement of speaker, emotion and content in the acoustic domain for transferring expressivity information from one speaker to another one, particularly when only neutral speech data is available for the latter. In [36], we proposed to transfer the expressive characteristics through layer adaptation during the learning step. The obtained results highlighted that there is a difficult trade-off between speaker's identity to remove and the expressivity to transfer. We are now working on an approach relying on multiclass N-pair based deep metric learning in recurrent conditional variational autoencoder (RCVAE) for implementing a multispeaker expressive text-to-speech (TTS) system. The proposed approach conditions the text-to-speech system on speaker embeddings, and leads to a clustering with respect to emotion in a latent space. The deep

metric learning helps to reduce the intra-class variance and increase the inter-class variance. We transfer the expressivity by using the latent variables for each emotion to generate expressive speech in the voice of a different speaker for which no expressive speech is available. The performance measured shows the model's capability to transfer the expressivity while preserving the speaker's voice in synthesized speech.

7.1.3. Preserving privacy

Speech signals involve a lot of private information. With a few minutes of data, the speaker identity can be modeled for malicious purposes like voice cloning, spoofing, etc. To reduce this risk, we investigated speaker anonymization strategies based on voice conversion. In contrast to prior evaluations, we argue that different types of attackers can be defined depending on the extent of their knowledge. We compared three conversion methods in three attack scenarios, and showed that these methods fail to protect against an attacker that has extensive knowledge of the type of conversion and how it has been applied, but may provide some protection against less knowledgeable attackers [64]. As an alternative, we proposed an adversarial approach to learn representations that perform well for ASR while hiding speaker identity. Our results demonstrate that adversarial training dramatically reduces the closed-set speaker classification accuracy, but this does not translate into increased open-set speaker verification error [45]. We are currently organizing the 1st Voice Privacy Challenge in which these and other approaches will be further assessed and compared.

7.2. Speech production and perception

7.2.1. Articulatory modeling

Participants: Denis Jouvet, Anne Bonneau, Dominique Fohr, Yves Laprie, Vincent Colotte, Slim Ouni, Agnes Piquard-Kipffer, Elodie Gauthier, Manfred Pastatter, Théo Biasutto-Lervat, Sara Dahmani, Ioannis Douros, Amal Houidhek, Lou Lee, Shakeel Ahmad Sheikh, Anastasiia Tsukanova, Louis Delebecque, Valérian Girard, Thomas Girod, Seyed Ahmad Hosseini, Mathieu Hu, Leon Rohrbacher, Imene Zangar.

7.2.1.1. Articulatory speech synthesis

A number of simplifying assumptions have to be made in articulatory synthesis to enable the speech signal to be generated in a reasonable time. They mainly consist of approximating the propagation of the sound in the vocal tract as a plane wave and approximating the 3D vocal tract shape from the mid-sagittal shape [30], and also simplifying the vocal tract topology by removing small cavities [29]. The posture of the subject in the MRI machine was also investigated [31]. Vocal tract resonances were evaluated from the 3D acoustic simulation computed with the K-wave Matlab package from the complete 3D vocal tract shape recovered from MRI and compared to those of real speech [27].

We also developed an approach for using articulatory features for speech synthesis. The approach is based on a deep feed-forward neural network-based speech synthesizer trained with the standard recipe of Merlin on the audio recorded during real-time MRI (RT-MRI) acquisitions: denoised (and yet containing a residual noise of the MRI machine) speech in French and force-aligned state labels encoding phonetic and linguistic information [26]. The synthesizer was augmented with eight parameters representing articulatory information (lips opening and protrusion, distances between the tongue and the velum, between the velum and the pharyngeal wall, and between the tongue and the pharyngeal wall) that were automatically extracted from the captures and aligned with the audio signal and the linguistic specification.

7.2.1.2. Dynamics of vocal tract and glottal opening

The problem of creating a 3D dynamic atlas of the vocal tract that captures the dynamics of the articulators in all three dimensions has been addressed [28]. The core steps of the method are using 2D real time MRI in several sagittal planes and, after temporal alignment, combine them using adaptive kernel regression. As a preprocessing step, a reference space was created to be used in order to remove anatomical information of the speakers and keep only the variability in speech production for the construction of the atlas. Using adaptive kernel regression makes the choice of atlas time points independent of the time points of the frames that are used as an input for the atlas construction.

We started the development of a database of realistic glottal gestures which will be used to design the glottal opening dynamics in articulatory synthesis paradigms. Experimental measurements of glottal opening dynamics in VCV and VCCV sequences uttered by real subjects have been achieved thanks to a specifically designed external photoglottographic device (ePGG) [33]. The existence of different patterns of glottal opening is evidenced according to the class of the consonant articulated.

7.2.1.3. Multimodal coarticulation modeling

We have investigated labial coarticulation to animate a virtual face from speech. We experimented a sequential deep learning model, bidirectional gated recurrent networks, that have been used successfully in addressing the articulatory inversion problem. We have used phonetic information as input to ensure speaker independence. The initialization of the last layers of the network has greatly eased the training and helped to handle coarticulation. It relies on dimensionality reduction strategies, allowing injecting knowledge of useful latent representation of the visual data into the network. We have trained and evaluated the model with a corpus consisting of 4 hours of French speech, and we got a good average RMSE (Root Mean Square Error) close to 1.3 mm [21].

7.2.1.4. Identifying disfluency in stuttered speech

Within the ANR project BENEPHIDIRE, the goal is to automatically identify typical kinds of stuttering disfluency using acoustic and visual cues for their automatic detection. This year, we started analyzing existing stuttering acoustic speech datasets to characterize the kind of data.

7.2.2. Multimodal expressive speech

7.2.2.1. Arabic speech synthesis

We have continued working on Modern Standard Arabic speech synthesis with ENIT (École Nationale d'Ingénieurs de Tunis, Tunisia), using HMM and NN based approaches. This year we investigated the modeling of the fundamental frequency for Arabic speech synthesis with feedforward and recurrent DNN, and using specific linguistic features for Arabic like vowel quantity and gemination [50].

7.2.2.2. Expressive audiovisual synthesis

After acquiring a high quality expressive audio-visual corpus based on fine linguistic analysis, motion capture, and naturalistic acting techniques, we have analyzed, processed, and phonetically aligned it with speech. We used conditional variational autoencoders (CVAE) to generate the duration, acoustic and visual aspects of speech without using emotion labels. Perceptual experiments have confirmed the capacity of our system to generate recognizable emotions. Moreover, the generative nature of the CVAE allowed us to generate well-perceived nuances of the six emotions and to blend different emotions together [23].

7.2.2.3. Lipsync - synchronization of lips movements with speech

In the ATT Dynalips-2, we have developed an English version of the system which allows us having a full multilingual lipsync system. During this ATT, we also worked on the business aspects (business plan, funding, investment, search for clients, etc.) with the goal of creating a startup, spinoff of the laboratory, during 2020.

7.2.3. Categorization of sounds and prosody

7.2.3.1. Non-native speech production

We analysed voicing in sequences of obstruents with French as L1 and German as L2, that is languages characterized by strong differences in the voicing dimension, including assimilation direction. To that purpose, we studied the realizations of two sequences of obstruents, where the first consonant, in final position, was fortis, and the second consonant, in initial position, was either a lenis stop or a lenis fricative. These sequences lead to a possible anticipation of voicing in French, a direction not allowed in German given German phonetics and phonology. Highly variable realizations were observed: progressive and regressive assimilation, and absence of assimilation, often accompanied by an unexpected pause [22].

We also started investigating non-native phoneme productions of French learners of German in comparison to phoneme productions by native German speakers. A set of research questions has been developed for which a customized French/German corpus was designed, and recorded by one reference native speaker of German so far. Based on these initial recordings and according to the targeted research questions, analysis strategies and algorithms have been elaborated and implemented, and are ready to be employed onto a larger data set. By means of these methods we expect to access phonetic and phonological grounds of recurrently occurring mis-pronunciation.

7.2.3.2. Language and reading acquisition by children having some language impairments

We continued examining the schooling experience of 170 children, teenagers and young adults with specific language impairment (dysphasia, dyslexia, dysorthographia) facing severe difficulties in learning to read. The phonemic discrimination, phonological and phonemic analysis difficulties faced in their childhoods had raised reading difficulties, which the pupils did not overcome. With 120 of these young people, we explored the presence of other neuro-developmental disorders. We also studied their reading habits to achieve better understanding of their difficulties.

We continued investigating the acquisition of language by hard-of-hearing children via cued speech (i.e. augmenting the audiovisual speech signal by visualizing the syllables uttered via a code of hand positions). We have used a digital book and a children's picture book with 3 hard-of-hearing children in order to compare scaffolding by the speech therapist or the teacher in these two situations.

We started to examine language difficulties and related problems with children with autism and to work with their parents with a view to creating an environment conducive to their progress [39].

7.2.3.3. Computer assisted language learning

In the METAL project, experiments are planned to investigate the use of speech technologies for foreign language learning and to experiment with middle and high school students learning German. This includes tutoring aspects based on a talking head to show proper articulation of words and sentences; as well as using automatic tools derived from speech recognition technology, for analyzing student pronunciations. The web application is under development, and experiments have continued for analyzing the performance of an automatic detection of mispronunciations made by language learners.

The ALOE project deals with children learning to read. In this project, we are also involved with tutoring aspects based on a talking head, and with grapheme-to-phoneme conversion which is a critical tool for the development of the digitized version of ALOE reading learning tools (tools which were previously developed and offered only in a paper form).

7.2.3.4. Prosody

The keynote [15] summarizes recent research on speech processing and prosody, and presents the extraction of prosodic features, as well as their usage in various tasks. Prosodic correlates of discourse particles have been investigated further. It was found that occurrences of different discourse particles with the same pragmatic value have a great tendency to share the same prosodic pattern; hence, the question of their commutability have been studied [37].

7.3. Speech in its environment

Participants: Denis Jouvet, Antoine Deleforge, Dominique Fohr, Emmanuel Vincent, Md Sahidullah, Irène Illina, Odile Mella, Romain Serizel, Tulika Bose, Guillaume Carbajal, Diego Di Carlo, Sandipana Dowerah, Ashwin Geet Dsa, Adrien Dufraux, Raphaël Duroselle, Mathieu Fontaine, Nicolas Furnon, Mohamed Amine Menacer, Mauricio Michel Olvera Zambrano, Lauréline Perotin, Sunit Sivasankaran, Nicolas Turpault, Nicolas Zampieri, Ismaël Bada, Yassine Boudi, Mathieu Hu, Stephane Level.

7.3.1. Acoustic environment analysis

We are constantly surrounded by ambient sounds and rely heavily on them to obtain important information about our environment. Deep neural networks are useful to learn relevant representations of these sounds. Recent studies have demonstrated the potential of unsupervised representation learning using various flavors of the so-called triplet loss (a triplet is composed of the current sample, a so-called positive sample from the same class, and a negative sample from a different class), and compared it to supervised learning. To address real situations involving both a small labeled dataset and a large unlabeled one, we combined unsupervised and supervised triplet loss based learning into a semi-supervised representation learning approach and compared it with supervised and unsupervised representation learning depending on the ratio between the amount of labeled and unlabeled data [49].

Pursuing our involvement in the community on ambient sound recognition, we co-organized a task on large-scale sound event detection as part of the Detection and Classification of Acoustic Scenes and Events (DCASE) 2019 Challenge [48]. It focused on the problem of learning from audio segments that are either weakly labeled or not labeled, targeting domestic applications. We also published a summary of the outcomes of the DCASE 2017 Challenge, in which we had organized the first version of that task [7] and a detailed analysis of the submissions to that task in 2018 [16] and 2019 [61].

7.3.2. Speech enhancement and noise robustness

7.3.2.1. Sound source localization and counting

In multichannel scenarios, source localization, counting and separation are tightly related tasks. Concerning deep learning based speaker localization, we introduced the real and imaginary parts of the acoustic intensity vector in each time-frequency bin as suitable input features. We analyzed the inner working of the neural network using layerwise relevance propagation [9]. We also defined alternative regression-based approaches for localization and compared them to the usual classification-based approach on a discrete grid [43]. Lauréline Perotin successfully defended her PhD on this topic [2]. In [24], we proposed the first deep-learning based method for blindly estimating early acoustic echoes. We showed how estimates of these echoes enable 2D sound source localization with only two microphones near a reflective surface, a task normally impossible with traditional methods. Finally, we published our former work on motion planning for robot audition [8].

We organized the IEEE Signal Processing Cup 2019, an international competition aimed at teams of undergraduate students [5]. The tasks we proposed were on sound source localization using an array embedded in a flying drone for search and rescue application. Submissions to the first phase of the competition were opened from November 2018 to March 2019, and the final took place on May the 13th at the international conference ICASSP in Brighton. 20 teams of undergraduate students from 18 universities in 11 countries participated, for a total of 132 participants. The drone-embedded sound source localization dataset we recorded for the challenge was made publically available after the competition and has received over 1,000 file downloads as of December 2019.

7.3.2.2. Speech enhancement

We investigated the effect of speaker localization accuracy on deep learning based speech enhancement quality. To do so, we generated a multichannel, multispeaker, reverberated, noisy dataset inspired from the well studied WSJ0-2mix and evaluated enhancement performance in terms of the word error rate. We showed that the signal-to-interference ratio between the speakers has a higher impact on the ASR performance than the angular distance [62]. In addition, we proposed a deflation method which estimates the sources iteratively. At each iteration, we estimate the location of the speaker, derive the corresponding time-frequency mask and remove the estimated source from the mixture before estimating the next one [63].

In parallel, we introduced a method for joint reduction of acoustic echo, reverberation and noise. This method models the target and residual signals after linear echo cancellation and dereverberation using a multichannel Gaussian modeling framework and jointly represents their spectra by means of a neural network. We developed an iterative block-coordinate ascent algorithm to update all the filters. The proposed approach outperforms in terms of overall distortion a cascade of the individual approaches and a joint reduction approach which does not rely on a spectral model of the target and residual signals [53], [57].

In the context of ad-hoc acoustic antennas, we proposed to extend the distributed adaptive node-specific signal estimation approach to a neural networks framework. At each node, a local filtering is performed to send one signal to the other nodes where a mask is estimated by a neural network in order to compute a global multi-channel Wiener filter. In an array of two nodes, we showed that this additional signal can be efficiently taken into account to predict the masks and leads to better speech enhancement performances than when the mask estimation relies only on the local signals [58].

We have been pursuing our work on non-Gaussian heavy-tail models for signal processing, and notably investigated whether such models could be of use to devise new cost functions for the training of deep generative models for source separation [34]. In the case of speech enhancement, it turned out that the related log-likelihood functions could advantageously replace the more constraining squared-error and lead to significant performance gains.

We have also been pursuing our theoretical work on multichannel alpha-stable models, devising two new multichannel filtering methods that are adequate for processing multivariate heavy-tailed vectors. The related work is presented in Mathieu Fontaine's PhD manuscript [1].

7.3.2.3. Robust speech recognition

Achieving robust speech recognition in reverberant, noisy, multi-source conditions requires not only speech enhancement and separation but also robust acoustic modeling. In order to motivate further work by the community, we created the series of CHiME Speech Separation and Recognition Challenges in 2011. We are now organizing the 6th edition of the Challenge, and released the French dataset for ambient assisted living applications previously collected as part of the FUI VOICEHOME project [4].

7.3.2.4. Speaker recognition

Automatic speaker recognition systems give reasonably good recognition accuracy when adequate amount of speech data from clean conditions are used for enrollment and test. However, performance degrades substantially in real-world noisy conditions as well as due to the lack of adequate speech data. Apart from these two practical limitations, speaker recognition performance also degrades in presence of spoofing attacks [51] where playback voice or synthetic speech generated with voice conversion or speech synthesis methods are used by attackers to access a system protected with voice biometrics.

We have explored a new speech quality measure for quality-based fusion of speaker recognition systems. The quality metric is formulated with the zero-order statistics estimated during i-vector extraction. The proposed quality metric is shown to capture the speech duration information, and it has outperformed absolute-duration based quality measures when combining multiple speaker recognition systems. Noticeable improvement over existing methods have been observed specifically for the short-duration conditions [10].

We have also participated in speaker recognition evaluation campaigns NIST SREs and VoxSRC. For the NIST SREs [54], the key problem was to recognize speakers from low-quality telephone conversations. In addition, the language mismatch between system development and data under test made the problem more challenging. In VoxSRC, on the other hand, the main problem was to recognize speakers speaking short sentences of about 10 sec where the speech files are extracted from Youtube video clips. We have explored acoustic feature extraction, domain adaptation, parameter optimization and system fusion for these challenges. For VoxSRC, our system has shown substantial improvement over baseline results.

We also introduced a statistical uncertainty-aware method for robust i-vector based speaker verification in noisy conditions, that is the first one to improve over simple chaining of speech enhancement and speaker verification on the challenging NIST-SRE corpus mixed with real domestic noise and reverberation [44].

Robust speaker recognition is an essential component of speaker diarization systems. We have participated in the second DIHARD challenge where the key problem was the diarization of speech signals collected from diverse real-world conditions. We have explored speech activity detection, domain grouping, acoustic features, and speech enhancement for improved speaker recognition. Our proposed system has shown considerable improvement over the Kaldi-based baseline system provided by the challenge organizer [60].

We have co-organized the ASVspoof 2019 challenge, as an effort to develop next-generation countermeasures for automatic detection of spoofed/fake audio [46]. This involved creating the audio dataset, designing experiments, evaluating and analyzing the results. 154 teams or individuals participated in the challenge. The database is available for research and further exploration from Edinburgh DataShare, and has been downloaded/viewed more than a thousand times so far.

We have also analyzed whether target speaker selection can help in attacking speaker recognition systems with voice impersonation [35]. Our study reveals that impersonators were not successful in attacking the systems, however, the speaker similarity scores transfer well from the attacker's system to the attacked system [12]. Though there were modest changes in F0 and formants, we found that the impersonators were able to considerably change their speaking rates when mimicking targets.

7.3.2.5. Language identification

State-of-the-art spoken language identification systems are constituted of three modules: a frame level feature extractor, a segment level embedding extractor and a classifier. The performance of these systems degrades when facing mismatch between training and testing data. Although most domain adaptation methods focus on adaptation of the classifier, we have developed an unsupervised domain adaptation of the embedding extractor. The proposed approach consists in a modification of the loss of the segment level embedding extractor by adding a regularisation term. Experiments were conducted with respect to transmission channel mismatch between telephone and radio channels using the RATS corpus. The proposed method is superior to adaptation of the classifier and obtain the same performance as published language identification results but without using labelled data from the target domain.

7.3.3. Linguistic and semantic processing

7.3.3.1. Transcription, translation, summarization and comparison of videos

Within the AMIS project, we studied different subjects related to the processing of videos. The first one concerns the machine translation of Arabic-English code-switched documents [41]. Code-switching is defined as the use of more than one language by a speaker within an utterance. The second one deals with the summarization of videos into a target language [11]. This exploits research carried on in several areas including video summarization, speech recognition, machine translation, audio summarization and speech segmentation. One of the big challenges of this work was to conceive a way to evaluate objectively a system composed of several components given that each of them has its limits and that errors propagate through the components. A third aspect was a method for extracting text-based summarization of Arabic videos [40]. The automatic speech recognition system developed to transcribe the videos has been adapted to the Algerian dialect, and additional modules were developed for segmenting the flow of recognized word into sentences, and for summarization. Finally the last aspect concerns the comparison of the opinions of two videos in two different languages [20]. Evaluations have been carried on comparable videos extracted from a corpus of 1503 Arabic and 1874 English videos.

7.3.3.2. Detection of hate speech in social media

The spectacular expansion of the Internet led to the development of a new research problem in natural language processing, the automatic detection of hate speech, since many countries prohibit hate speech in public media. In the context of the M-PHASIS project, we proposed a new approach for the classification of tweets, aiming to predict whether a tweet is abusive, hate or neither. We compare different unsupervised word representations and DNN classifiers, and study the robustness of the proposed approaches to adversarial attacks when adding one (healthy or toxic) word. We are evaluating the proposed methodology on the English Wikipedia Detox corpus and on a Twitter corpus.

7.3.3.3. Introduction of semantic information in an automatic speech recognition system

In current state-of-the-art automatic speech recognition systems, N-gram based models are used to take into account language information. They have a local view and are mainly based on syntax. The introduction of semantic information and longer term information in a recognition system should make it possible to remove some ambiguities and reduce the error rate of the system. Within the MMT project, we are proposing and

evaluating methods for integrating semantic information into our speech recognition system through the use of various word embeddings.

7.3.3.4. Music language modeling

Similarly to speech, language models play a key role in music modeling. We represented the hierarchical structure of a temporal scenario (for instance, a chord progression) via a phrase structure grammar and proposed a method to automatically induce this grammar from a corpus and to exploit it in the context of machine improvisation [6].

PANAMA Project-Team

7. New Results

7.1. Sparse Representations, Inverse Problems, and Dimension Reduction

Sparsity, low-rank, dimension-reduction, inverse problem, sparse recovery, scalability, compressive sensing

The team's activity ranges from theoretical results to algorithmic design and software contributions in the fields of sparse representations, inverse problems and dimension reduction.

7.1.1. Computational Representation Learning: Algorithms and Theory

Participants: Rémi Gribonval, Hakim Hadj Djilani, Cássio Fraga Dantas, Jeremy Cohen.

Main collaborations: Luc Le Magoarou (IRT b-com, Rennes), Nicolas Tremblay (GIPSA-Lab, Grenoble), R. R. Lopes and M. N. Da Costa (DSPCom, Univ. Campinas, Brazil)

An important practical problem in sparse modeling is to choose the adequate dictionary to model a class of signals or images of interest. While diverse heuristic techniques have been proposed in the literature to learn a dictionary from a collection of training samples, classical dictionary learning is limited to small-scale problems. In our work introduced below, by imposing structural constraints on the dictionary and pruning provably unused atoms, we could alleviate the curse of dimensionality.

Multilayer sparse matrix products for faster computations. Inspired by usual fast transforms, we proposed a general dictionary structure (called FAμST for Flexible Approximate Multilayer Sparse Transforms) that allows cheaper manipulation, and an algorithm to learn such dictionaries together with their fast implementation, with reduced sample complexity. A comprehensive journal paper was published in 2016 [75], and we further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms [76], empirically showing that $\mathcal{O}(n \log n)$ approximate implementations of Graph Fourier Transforms are possible for certain families of graphs. This opened the way to substantial accelerations for Fourier Transforms on large graphs. This year we focused on the development of the FAμST software library (see Section 6), providing transparent interfaces of FAμST data-structures with both Matlab and Python.

Kronecker product structure for faster computations. In parallel to the development of FAuST, we proposed another approach to structured dictionary learning that also aims at speeding up both sparse coding and dictionary learning. We used the fact that for tensor data, a natural set of linear operators are those that operate on each dimension separately, which correspond to rank-one multilinear operators. These rank-one operators may be cast as the Kronecker product of several small matrices. Such operators require less memory and are computationally attractive, in particular for performing efficient matrix-matrix and matrix-vector operations. In our proposed approach, dictionaries are constrained to belong to the set of low-rank multilinear operators, that consist of the sum of a few rank-one operators. The general approach, coined HOSUKRO for High Order Sum of Kronecker products, was shown last year to reduce empirically the sample complexity of dictionary learning, as well as theoretical complexity of both the learning and the sparse coding operations [67]. This year we demonstrated its potential for hyperspectral image denoising. A new efficient algorithm with lighter sample complexity requirements and computational burden was proposed and shown to be competitive with the state-of-the-art for hyperspectral image denoising with dedicated adjustements [50], [28], [27].

Combining faster matrix-vector products with screening techniques. We combined accelerated matrix-vector multiplications offered by FAμST / HOSUKRO matrix approximations with dynamic screening [59], that safely eliminates inactive variables to speedup iterative convex sparse recovery algorithms. First, we showed how to obtain safe screening rules for the exact problem while manipulating an approximate dictionary [68]. We then adapted an existing screening rule to this new framework and define a general procedure to leverage the advantages of both strategies. This lead to a journal publication [21] that includes new techniques based on duality gaps to optimally switch from a coarse dictionary approximation to a finer one. Significant complexity reductions were obtained in comparison to screening rules alone.

7.1.2. Generalized matrix inverses and the sparse pseudo-inverse

Participant: Rémi Gribonval.

Main collaboration: Ivan Dokmanic (University of Illinois at Urbana Champaign, USA)

We studied linear generalized inverses that minimize matrix norms. Such generalized inverses are famously represented by the Moore-Penrose pseudoinverse (MPP) which happens to minimize the Frobenius norm. Freeing up the degrees of freedom associated with Frobenius optimality enables us to promote other interesting properties. In a first part of this work [64], we looked at the basic properties of norm-minimizing generalized inverses, especially in terms of uniqueness and relation to the MPP. We first showed that the MPP minimizes many norms beyond those unitarily invariant, thus further bolstering its role as a robust choice in many situations. We then concentrated on some norms which are generally not minimized by the MPP, but whose minimization is relevant for linear inverse problems and sparse representations. In particular, we looked at mixed norms and the induced $\ell^p \to \ell^q$ norms.

An interesting representative is the sparse pseudoinverse which we studied in much more detail in a second part of this work published this year [19], motivated by the idea to replace the Moore-Penrose pseudoinverse by a sparser generalized inverse which is in some sense well-behaved. Sparsity implies that it is faster to apply the resulting matrix; well-behavedness would imply that we do not lose much in stability with respect to the least-squares performance of the MPP. We first addressed questions of uniqueness and non-zero count of (putative) sparse pseudoinverses. We showed that a sparse pseudoinverse is generically unique, and that it indeed reaches optimal sparsity for almost all matrices. We then turned to proving a stability result: finite-size concentration bounds for the Frobenius norm of p-minimal inverses for $1 \leq p \leq 2$. Our proof is based on tools from convex analysis and random matrix theory, in particular the recently developed convex Gaussian min-max theorem. Along the way we proved several results about sparse representations and convex programming that were known folklore, but of which we could find no proof.

7.1.3. Algorithmic exploration of large-scale Compressive Learning via Sketching

Participants: Rémi Gribonval, Antoine Chatalic.

Main collaborations this year: Nicolas Keriven (ENS Paris), Phil Schniter & Evan Byrne (Ohio State University, USA), Laurent Jacques & Vincent Schellekens (Univ Louvain, Belgium), Florimond Houssiau & Y.-A. de Montjoye (Imperial College London, UK)

Sketching for Large-Scale Learning. When learning from voluminous data, memory and computational time can become prohibitive. We proposed during the Ph.D. thesis of Anthony Bourrier [60] and Nicolas Keriven [74] an approach based on sketching. A low-dimensional sketch is computed by averaging (random) features over the training collection. The sketch can be seen as made of a collection of empirical generalized moments of the underlying probability distribution. Leveraging analogies with compressive sensing, we experimentally showed that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough, and released an associated toolbox for reproducible research (see SketchMLBox, Section 6) with the so-called Compressive Learning Orthogonal Matching Pursuit (CL-OMP) algorithm which is inspired by Matching Pursuit. Three unsupervised learning settings have been addressed so far: Gaussian Mixture Modeling, k-means clustering, and principal component analysis. A survey conference paper on sketching for large-scale learning was published this year [25], and an extended journal version of this survey is in preparation.

Efficient algorithms to learn for sketches Last year, we showed that in the high-dimensional setting one can substantially speedup both the sketching stage and the learning stage with CL-OMP by replacing Gaussian random matrices with fast random linear transforms in the sketching procedure [63]. We studied an alternative to CL-OMP for cluster recovery from a sketch, which is based on simplified hybrid generalized approximate message passing (SHyGAMP). Numerical experiments suggest that this approach is more efficient than CL-OMP (in both computational and sample complexity) and more efficient than k-means++ in certain regimes [61]. During his first year of Ph.D., Antoine Chatalic visited the group of Phil Schniter to further investigate this topic, and a journal paper has been published as a result of this collaboration [15].

Privacy-preserving sketches Sketching provides a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum. We established theoretical privacy guarantees (with the *differential privacy* framework) and explored the utility / privacy tradeoffs of Compressive K-means [24]. A journal paper is in preparation where we extend these results to Gaussian mixture modeling and principal component analysis.

Advances in optical-based random projections Random projections are a key ingredient of sketching. Motivated by the recent development of dedicated optics-based hardware for rapid random projections, which leverages the propagation of light in random media, we tackled the problem of recovering the phase of complex linear measurements when only magnitude information is available and we control the input. A signal of interest $\xi \in \mathbb{R}^N$ is mixed by a random scattering medium to compute the projection $y = \mathbf{A}\xi$, with $\mathbf{A} \in \mathbb{C}^{M\times N}$ a realization of a standard complex Gaussian independent and identically distributed (iid) random matrix. Such optics-based matrix multiplications can be much faster and energy-efficient than their CPU or GPU counterparts, yet two difficulties must be resolved: only the intensity $|y|^2$ can be recorded by the camera, and the transmission matrix $\mathbf{A}$ is unknown. We showed that even without knowing $\mathbf{A}$, we can recover the unknown phase of y for some equivalent transmission matrix with the same distribution as $\mathbf{A}$. Our method is based on two observations: first, conjugating or changing the phase of any row of $\mathbf{A}$ does not change its distribution; and second, since we control the input we can interfere ξ with arbitrary reference signals. We showed how to leverage these observations to cast the measurement phase retrieval problem as a Euclidean distance geometry problem. We demonstrated appealing properties of the proposed algorithm in both numerical simulations and real hardware experiments. Not only does our algorithm accurately recover the missing phase, but it mitigates the effects of quantization and the sensitivity threshold, thus improving the measured magnitudes [33].

7.1.4. Theoretical results on Low-dimensional Representations, Inverse problems, and Dimension Reduction

Participants: Rémi Gribonval, Clément Elvira, Jérémy Cohen.

Main collaboration: Nicolas Keriven (ENS Paris), Gilles Blanchard (Univ Postdam, Germany), Cédric Herzet (SIMSMART project-team, IRMAR / Inria Rennes), Charles Soussen (Centrale Supelec, Gif-sur-Yvette), Mila Nikolova (CMLA, Cachan), Nicolas Gillis (UMONS)

Information preservation guarantees with low-dimensional sketches. We established a theoretical framework for sketched learning, encompassing statistical learning guarantees as well as dimension reduction guarantees. The framework provides theoretical grounds supporting the experimental success of our algorithmic approaches to compressive K-means, compressive Gaussian Mixture Modeling, as well as compressive Principal Component Analysis (PCA). A comprehensive preprint is being revised for a journal [71].

Recovery guarantees for algorithms with continuous dictionaries. We established theoretical guarantees on sparse recovery guarantees for a greedy algorithm, orthogonal matching pursuit (OMP), in the context of continuous dictionaries [66], e.g. as appearing in the context of sparse spike deconvolution. Analyses based on discretized dictionary fail to be conclusive when the discretization step tends to zero, as the coherence goes to one. Instead, our analysis is directly conducted in the continuous setting amd exploits specific properties of the positive definite kernel between atom parameters defined by the inner product between the corresponding atoms. For the Laplacian kernel in dimension one, we showed in the noise-free setting that OMP exactly recovers the atom parameters as well as their amplitudes, regardless of the number of distinct atoms [66]. A preprint describing a full class of kernels for which such an analysis holds, in particular for higher dimensional parameters, has been released and submitted to a journal [30], [36], [31], [51].

Identifiability of Complete Dictionary Learning In the era of deep learning, dictionary learning has proven to remain an important and extensively-used data mining and processing tool. Having been studied and used for over twenty years, dictionary learning has well-understood properties. However there was a particular stone missing, which was understanding deterministic conditions for the parameters of dictionary learning to be uniquely retrieved from a training data set. We filled this gap partially by drastically improving on the previously best such conditions in the case of complete dictionaries [16]. Moreover, although algorithms with guaranties to compute the unique best solution do exist, they are seldom used in practice due to their

high computational cost. In subsequent work, we showed that faster algorithms typically used to compute dictionary learning often failed at computing the unique solution (in cases where our previous result guaranties this uniqueness), opening the way to new algorithms that are both fast and guarantied [26].

On Bayesian estimation and proximity operators. There are two major routes to address the ubiquitous family of inverse problems appearing in signal and image processing, such as denoising or deblurring. The first route is Bayesian modeling: prior probabilities are used to model both the distribution of the unknown variables and their statistical dependence with the observed data, and estimation is expressed as the minimization of an expected loss (e.g. minimum mean squared error, or MMSE). The other route is the variational approach, popularized with sparse regularization and compressive sensing. It consists in designing (often convex) optimization problems involving the sum of a data fidelity term and a penalty term promoting certain types of unknowns (e.g., sparsity, promoted through an L1 norm).

Well known relations between these two approaches have lead to some widely spread misconceptions. In particular, while the so-called Maximum A Posterori (MAP) estimate with a Gaussian noise model does lead to an optimization problem with a quadratic data-fidelity term, we disprove through explicit examples the common belief that the converse would be true. In previous work we showed that for denoising in the presence of additive Gaussian noise, for any prior probability on the unknowns, the MMSE is the solution of a penalized least-squares problem, with all the apparent characteristics of a MAP estimation problem with Gaussian noise and a (generally) different prior on the unknowns [72]. In other words, the variational approach is rich enough to build any MMSE estimator associated to additive Gaussian noise via a well chosen penalty.

This year, we achieved generalizations of these results beyond Gaussian denoising and characterized noise models for which the same phenomenon occurs. In particular, we proved that with (a variant of) Poisson noise and any prior probability on the unknowns, MMSE estimation can again be expressed as the solution of a penalized least-squares optimization problem. For additive scalar denoising, the phenomenon holds if and only if the noise distribution is log-concave, resulting in the perhaps surprising fact that scalar Laplacian denoising can be expressed as the solution of a penalized least-squares problem [22] Somewhere in the proofs appears an apparently new characterization of proximity operators of (nonconvex) penalties as subdifferentials of convex potentials [54].

7.1.5. Low-rank approximations: fast constrained algorithms

Participant: Jeremy Cohen.

Main collaborations: Nicolas Gillis (Univ. Mons, Belgium), Andersen Man Shun Ang (Univ. Mons, Belgium), Nicolas Nadisic (Univ. Mons, Belgium).

Low-Rank Approximations (LRA) aim at expressing the content of a multiway array by a sum of simpler separable arrays. Understood as a powerful unsupervised machine learning technique, LRA are most and foremost modern avatars of sparsity that are still not fully understood. In particular, algorithms to compute the parameters of LRA demend a lot of computer ressources and provide sub-optimal results. An important line of work over the last year has been to design efficient algorithms to compute constrained LRA, and in particular constrained low-rank tensor decompositions. This work has been carried out though a collaboration with the ERC project COLORAMAP of Nicolas Gillis (Univ. Mons, Belgium) and his PhD students Nicolas Nadisic (co-supervision) and Andersen Man Shun Ang.

Extrapolated Block-coordinate algorithms for fast tensor decompositions State-of-the-art algorithms for computing tensor decompositions are based on the idea that solving alternatively for smaller blocks of parameters is easier than solving the large problem at once. Despite showing nice convergence speeds, the obtained Block Coordinate Descent algorithms (BCD) are prone to being stuck near saddle points. We have shown in preliminary work, which is still ongoing, that BCD algorithms can be improved using Nesterov extrapolation in-between block updates. This improves empirical convergence speed in constrained and unconstrained tensor decompositions tremendously at almost no additional computation cost, and is therefore bound to have a large impact on the community [37].

Exact sparse nonnegative least-squares solutions to least-squares problems Another important LRA is Nonnegative Matrix factorization, which has found many diverse applications such as in remote sensing or automatic music transcription. Sometimes, imposing sparsity on parameters of NMF is crucial to be able to correctly process and interpret the output of NMF. However, sparse NMF has scarcely been studied, and its computation is challenging. In fact, even only a subproblem in a BCD approach, sparse nonnegative least-squares, is already NP-hard. We proposed to solve this sparse nonnegative least-squares problem exactly using a combinatorial algorithm. To reduce as much as possible the cost of solving this combinatorial problem, a Branch and Bound algorithm was proposed which, on average, reduces the computational complexity drastically. A next step will be to use this branch and bound algorithm as a brick for proposing an efficient algorithm for sparse NMF.

7.1.6. Algorithmic Exploration of Sparse Representations for Neurofeedback

Participant: Rémi Gribonval.

Claire Cury, Pierre Maurel & Christian Barillot (EMPENN Inria project-team, Rennes)

In the context of the HEMISFER (Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation) Comin Labs project (see Section 1), in collaboration with the EMPENN team, we validated a technique to estimate brain neuronal activity by combining EEG and fMRI modalities in a joint framework exploiting sparsity [82]. We then focused on directly estimating neuro-feedback scores rather than brain activity. Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) both allow measurement of brain activity for neuro-feedback (NF), respectively with high temporal resolution for EEG and high spatial resolution for fMRI. Using simultaneously fMRI and EEG for NF training is very promising to devise brain rehabilitation protocols, however performing NF-fMRI is costly, exhausting and time consuming, and cannot be repeated too many times for the same subject. We proposed a technique to predict NF scores from EEG recordings only, using a training phase where both EEG and fMRI NF are available [39]. A journal paper has been submitted.

7.2. Emerging activities on high-dimensional learning with neural networks

Participants: Rémi Gribonval, Himalaya Jain, Pierre Stock.

Main collaborations: Patrick Perez (Technicolor R & I, Rennes), Gitta Kutyniok (TU Berlin, Germany), Morten Nielsen (Aalborg University, Denmark), Felix Voigtlaender (KU Eichstätt, Germany), Herve Jegou and Benjamin Graham (FAIR, Paris)

dictionary learning, large-scale indexing, sparse deep networks, normalization, sinkhorn, regularization

Many of the data analysis and processing pipelines that have been carefully engineered by generations of mathematicians and practitioners can in fact be implemented as deep networks. Allowing the parameters of these networks to be automatically trained (or even randomized) allows to revisit certain classical constructions. Our team has started investigating the potential of such approaches both from an empirical perspective and from the point of view of approximation theory.

Learning compact representations for large-scale image search. The PhD thesis of Himalaya Jain [73], which received the Fondation Rennes 1 PhD prize this year, was dedicated to learning techniques for the design of new efficient methods for large-scale image search and indexing.

Equi-normalization of Neural Networks. Modern neural networks are over-parameterized. In particular, each rectified linear hidden unit can be modified by a multiplicative factor by adjusting input and output weights, without changing the rest of the network. Inspired by the Sinkhorn-Knopp algorithm, we introduced a fast iterative method for minimizing the l2 norm of the weights, equivalently the weight decay regularizer. It provably converges to a unique solution. Interleaving our algorithm with SGD during training improves the test accuracy. For small batches, our approach offers an alternative to batch- and group- normalization on CIFAR-10 and ImageNet with a ResNet-18. This work was presented at ICLR 2019 [41].

Approximation theory with deep networks. We study the expressivity of sparsely connected deep networks. Measuring a network's complexity by its number of connections with nonzero weights, or its number of neurons, we consider the class of functions which error of best approximation with networks of a given complexity decays at a certain rate. Using classical approximation theory, we showed that this class can be endowed with a norm that makes it a nice function space, called approximation space. We established that the presence of certain "skip connections" has no impact of the approximation space, and studied the role of the network's nonlinearity (also known as activation function) on the resulting spaces, as well as the benefits of depth. For the popular ReLU nonlinearity (as well as its powers), we related the newly identified spaces to classical Besov spaces, which have a long history as image models associated to sparse wavelet decompositions. The sharp embeddings that we established highlight how depth enables sparsely connected networks to approximate functions of increased "roughness" (decreased Besov smoothness) compared to shallow networks and wavelets. A preprint has been published and is under review for a journal [23].

7.3. Emerging activities on Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

7.3.1. Audio Inpainting and Denoising

Participants: Rémi Gribonval, Nancy Bertin, Clément Gaultier.

Main collaborations: Srdan Kitic (Orange, Rennes)

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. Building upon our previous pioneering contributions [57], we proposed over the last five years a series of algorithms leveraging the competitive cosparse approach, which offers a very appealing trade-off between reconstruction performance and computational time, and its extensions to the incorporation of the so-called "social" into problems regularized by a cosparse prior. We exhibited a common framework allowing to tackle both denoising and declipping in a unified fashion [69]; these results, together with listening tests results that were specified and prepared in 2019 and will be run soon, will be included in an ongoing journal paper, to be submitted in 2020. This year, following Clément Gaultier Ph.D. defense [12], we progressed towards industrial transfer of these results through informal interaction with a company commercializing audio plugins, in particular with new developments to alleviate some artifacts absent from simulation but arising in real-world use cases.

7.4. Source Localization and Separation

Source separation, sparse representations, probabilistic model, source localization

Acoustic source localization is, in general, the problem of determining the spatial coordinates of one or several sound sources based on microphone recordings. This problem arises in many different fields (speech and sound enhancement, speech recognition, acoustic tomography, robotics, aeroacoustics...) and its resolution, beyond an interest in itself, can also be the key preamble to efficient source separation, which is the task of retrieving the source signals underlying a multichannel mixture signal. Over the last years, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [9], hereafter summarized as the "local Gaussian modeling", and we showed how it could be used to quickly design new models adapted to the data at hand and estimate its parameters via the EM algorithm. This model became the basis of a large number of works in the field, including our own. This accumulated progress lead, in 2015, to two main achievements: a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released [84] and we published an overview paper on recent and going research along the path of *guided* separation in a special issue of IEEE Signal Processing Magazine [11].

From there, our recent work divided into several tracks: maturity work on the concrete use of these tools and principles in real-world scenarios, in particular within the INVATE project and the collaboration with the startup 5^{th} dimension (see Sections 8.1.2 , 8.1.4), on the one hand; on the other hand, an emerging track on audio scene analysis with machine learning, evolved beyond the "localization and separation" paradigm, and is the subject of a more recent axis of research presented in Section 7.5 .

7.4.1. Towards Real-world Localization and Separation

Participants: Nancy Bertin, Frédéric Bimbot, Rémi Gribonval, Ewen Camberlein, Romain Lebarbenchon, Mohammed Hafsati.

Main collaborations: Emmanuel Vincent (MULTISPEECH Inria project-team, Nancy)

Based on the team's accumulated expertise and tools for localization and separation using the local Gaussian model, two real-world applications were addressed in the past year, which in turn gave rise to new research tracks.

First, our work within the voiceHome project (2015-2017), an OSEO-FUI industrial collaboration [0] aiming at developing natural language dialog in home applications, such as control of domotic and multimedia devices, in realistic and challenging situations (very noisy and reverberant environments, distant microphones) found its conclusion with the publication of a journal paper in a special issue of Speech Communication [14].

Accomplished progress and levers of improvements identified thanks to this project resulted in the granting of an Inria ADT (Action de Développement Technologique). This new development phase of the FASST software started in September 2017 and was achieved this year by the release of the third version of the toolbox, with significant progress towards efficient initialization, low latency and reduction of the computational burden.

In addition, evolutions of the MBSSLocate software initiated during this project led to a successful participation in the IEEE-AASP Challenge on Acoustic Source Localization and Tracking (LOCATA) [77], and served as a baseline for the publication of the for the IEEE Signal Processing Cup 2019 [21]. The SP Cup was also fueled by the publicly available DREGON dataset 5 recorded in PANAMA, including noiseless speech and on-flight ego-noise recordings, devoted to source localization from a drone [117].

Finally, these progress also led to a new industrial transfer with the start-up 5^{th} dimension (see Section 8.1.4). During this collaboration aiming at equipping a pair of glasses with an array of microphones and "smart" speech enhancement functionalities, we particularly investigated the impact of obstacles between microphones in the localization and separation performance, the selection of the best subset of microphones in the array for side speakers hidden by the head shadow, and the importance of speaker enrolment (learning spectral dictionaries of target users voices) in this use case.

7.4.2. Separation for Remixing Applications

Participants: Nancy Bertin, Rémi Gribonval, Mohammed Hafsati.

Main collaborations: Nicolas Epain (IRT b<>com, Rennes)

Second, through the Ph.D. of Mohammed Hafsati (in collaboration with the IRT b<>com with the INVATE project, see Section 8.1.2) started in November 2016, we investigated a new application of source separation to sound re-spatialization from Higher Order Ambisonics (HOA) signals [70], in the context of free navigation in 3D audiovisual contents. We studied the applicability conditions of the FASST framework to HOA signals and benchmarked localization and separation methods in this domain. Simulation results showed that separating sources in the HOA domain results in a 5 to 15 dB increase in signal-to-distortion ratio, compared to the microphone domain. These results were accepted for publication in the DAFx international conference [34]. We continued extending our methods following two tracks: hybrid acquisition scenarios, where the separation of HOA signals can be informed by complementary close-up microphonic signals, and the replacement of spectrogram NMF by neural networks for a better spectral adaptation of the models. Future work will include subjective evaluation of the developed workflows.

7.5. Towards comprehensive audio scene analysis

Source localization and separation, machine learning, room geometry, room properties, multichannel audio classification

[0] With partners: onMobile, Delta Dore, eSoftThings, Orange, Technicolor, LOUSTIC, Inria Nancy.

By contrast to the previous lines of work and results on source localization and separation, which are mostly focused on the *sources*, the following emerging activities consider the audio scene and its analysis in a wider sense, including the environment around the sources, and in particular the *room* they are included in, and their properties. This inclusive vision of the audio scene allows in return to revisit classical audio processing tasks, such as localization, separation or classification.

7.5.1. Room Properties: Estimating or Learning Early Echoes

Participants: Nancy Bertin, Diego Di Carlo, Clément Elvira.

Main collaborations: Antoine Deleforge (Inria Nancy – Grand Est), Ivan Dokmanic (University of Illinois at Urbana-Champaign, Coordinated Science Lab, USA), Robin Scheibler (Tokyo Metropolitan University, Tokyo, Japan), Helena Peic-Tukuljac (EPFL, Switzerland).

In [85] we showed that the knowledge of early echoes improved sound source separation performances, which motivates the development of (blind) echo estimation techniques. Echoes are also known to potentially be a key to the room geometry problem [65]. In 2019, two different approaches to this problem were explored.

As a competitive, yet similar approach to our previous work in [83], we proposed a new analytical method for off-the-grid early echoes estimation, based on continuous dictionaries and extensions of sparse recovery methods in this setting. From the well-known *cross-relation* between room impulse responses and signals in a "one source - two microphones" settings, the echo estimation problem can be recast as a Beurling-LASSO problem and solved with algorithms of this kind. This enables near-exact blind and off-grid echo retrieval from discrete-time measurements, and can outperform conventional methods by several orders of magnitude in precision, in an ideal case where the room impulse response is limited to a few weighted Diracs. Future work will include alternative initialization schemes, extensions to sparse-spectrum signals and noisy measurements, and applications to dereverberation and audio-based room shape reconstruction. This work, mostly lead by Clément Elvira, was submitted for publication in Icassp 2020.

On the other hand, the PhD thesis of Diego Di Carlo aims at applying the "Virtual Acoustic Space Traveler" (VAST) framework to the blind estimation of acoustic echoes, or other room properties (such as reverberation time, acoustic properties at the boundaries, etc.) Last year, we focused on identifying promising couples of inputs and outputs for such an approach, especially by leveraging the notions of relative transfer functions between microphones, the room impulse responses, the time-difference-of-arrivals, the angular spectra, and all their mutual relationships. In a simple yet common scenario of 2 microphones close to a reflective surface and one source (which may occur, for instance, when the sensors are placed on a table such as in voice-based assistant devices), we introduced the concept of microphone array augmentation with echoes (MIRAGE) and showed how estimation of early-echo characteristics with a learning-based approach is not only possible but can in fact benefit source localization. In particular, it allows to retrieve 2D direction of arrivals from 2 microphones only, an impossible task in anechoic settings. These first results were published in ICASSP [29]. In 2019, we improved the involved DNN architecture in MIRAGE and worked towards experimental validation of this result, by designing and recording a data set with annotated echoes in different conditions of reverberation. Future work will include extension of this data set, extension to more realistic and more complex scenarios (including more microphones, sources and reflective surfaces) and the estimation of other room properties such as the acoustic absorption at the boundaries, or ultimately, the room geometry. Some of these tracks currently benefit from the visit of Diego di Carlo to Bar-Ilan University (thanks to a MathSTIC doctoral outgoing mobility grant.)

7.5.2. Multichannel Audio Event and Room Classification

Participants: Marie-Anne Lacroix, Nancy Bertin.

Main collaborations: Pascal Scalart, Romuald Rocher (GRANIT Inria project-team, Lannion)

Typically, audio event detection and classification is tackled as a "pure" single-channel signal processing task. By constrast, audio source localization is the perfect example of multi-channel task "by construction". In parallel, the need to classify the type of scene or room has emerged, in particular from the rapid development of wearables, the "Internet of things" and their applications. The PhD of Marie-Anne Lacroix,

started in September 2018, combines these ideas with the aim of developing multi-channel, room-aware or spatially-aware audio classification algorithms for embedded devices. The PhD topic includes low-complexity and low-energy stakes, which will be more specifically tackled thanks to the GRANIT members area of expertise. During the first year of the PhD, we gathered existing data and identified the need for new simulations or recordings, and combined ideas from existing single-channel classification techniques with traditional spatial features in order to design several baseline algorithms for multi-channel joint localization and classification of audio events. The impact of feature quantization on classification performance is also currently under investigation and a participation to the 2020 edition of the IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE) is envisioned.

7.6. Music Content Processing and Information Retrieval

Music structure, music language modeling, System & Contrast model, complexity

Current work developed in our research group in the domain of music content processing and information retrieval explore various information-theoretic frameworks for music structure analysis and description [58], in particular the System & Contrast model [1].

7.6.1. Modeling music by Polytopic Graphs of Latent Relations (PGLR)

Participants: Corentin Louboutin, Frédéric Bimbot.

The musical content observed at a given instant within a music segment obviously tends to share privileged relationships with its immediate past, hence the sequential perception of the music flow. But local music content also relates with distant events which have occurred in the longer term past, especially at instants which are metrically homologous (in previous bars, motifs, phrases, etc.) This is particularly evident in strongly "patterned" music, such as pop music, where recurrence and regularity play a central role in the design of cyclic musical repetitions, anticipations and surprises.

The web of musical elements can be described as a Polytopic Graph of Latent Relations (PGLR) which models relationships developing predominantly between homologous elements within the metrical grid.

For regular segments the PGLR lives on an n-dimensional cube(square, cube, tesseract, etc...), n being the number of scales considered simultaneously in the multiscale model. By extension, the PGLR can be generalized to a more or less regular n-dimensional polytopes.

Each vertex in the polytope corresponds to a low-scale musical element, each edge represents a relationship between two vertices and each face forms an elementary system of relationships.

The estimation of the PGLR structure of a musical segment can be obtained computationally as the joint estimation of the description of the polytope, the nesting configuration of the graph over the polytope (reflecting the flow of dependencies and interactions between the elements within the musical segment) and the set of relations between the nodes of the graph, with potentially multiple possibilities.

If musical elements are chords, relations can be inferred by minimal transport [79] defined as the shortest displacement of notes, in semitones, between a pair of chords. Other chord representations and relations are possible, as studied in [81] where the PGLR approach is presented conceptually and algorithmically, together with an extensive evaluation on a large set of chord sequences from the RWC Pop corpus (100 pop songs).

Specific graph configurations, called Primer Preserving Permutations (PPP) are extensively studied in [80] and are related to 6 main redundant sequences which can be viewed as canonical multiscale structural patterns.

In parallel, recent work has also been dedicated to modeling melodic and rhythmic motifs in order to extend the polytopic model to multiple musical dimensions.

Results obtained in this framework illustrate the efficiency of the proposed model in capturing structural information within musical data and support the view that musical content can be delinearised in order to better describe its structure. Extensive results are included in Corentin Louboutin's PhD [13], defended in March 2019 and which was awarded the Prix Jeune Chercheur Science et Musique, in October.

7.6.2. Exploring Structural Dependencies in Melodic Sequences using Neural Networks

Participants: Nathan Libermann, Frédéric Bimbot.

This work is carried out in the framework of a PhD, co-directed by Emmanuel Vincent (Inria-Nancy).

In order to be able to generate structured melodic phrases and section, we explore various schemes for modeling dependencies between notes within melodies, using deep learning frameworks.

A a first set of experiments, we have considered a GRU-based sequential learning model, studied under different learning scenarios in order to better understand the optimal architectures in this context that can achieve satisfactory results. By this means, we wish to explore different hypotheses relating to temporal non-invariance relationships between notes within a structural segment (motif, phrase, section).

We have defined three types of recursive architectures corresponding to different ways to exploit the local history of a musical note, in terms of information encoding and generalization capabilities.

Initially conducted on the Lakh MIDI dataset, experiments have switched to the Meertens Tune Collections data set (Dutch traditional melodies) and confirm the trends observed in [78], w.r.t. the utility of non-ergodic models for the generation of melodic segments.

Ongoing work is extending these findings to the design of specific NN architectures, which incorporate attention models, to account for this non-invariance of information across musical segments.

7.6.3. Graph Signal Processing for Multiscale Representations of Music Similarity

Participants: Valentin Gillot, Frédéric Bimbot.

"Music Similarity" is a multifaceted concept at the core of Music Information Retrieval (MIR). Among the wide range of possible definitions and approaches to this notion, a popular one is the computation of a so-called content-based similarity matrix (S), in which each coefficient is a similarity measure between descriptors of short time frames at different instants within a music piece or a collection of pieces.

Matrix S can be seen as the adjacency matrix of an underlying graph, embodying the local and non-local similarities between parts of the music material. Considering the nodes of this graph as a new set of indices for the original music frames or pieces opens the door to a "delinearized" representation of music, emphasizing its structure and its semiotic content.

Graph Signal Processing (GSP) is an emerging topic devoted to extend usual signal processing tools (Fourier analysis, filtering, denoising, compression, ...) to signals "living" on graphs rather than on the time line, and to exploit mathematical and algorithmic tools on usual graphs, in order to better represent and manipulate these signals. Toy applications of GSP concepts on music content in music resequencing and music inpainting are illustrating this trend.

From exploratory experiments, first observations point towards the following hypotheses :

- local and non-local structures of a piece are highlighted in the adjacency matrix built from a simple time-frequency representation of the piece,
- the first eigenvectors of the graph Laplacian provide a rough structural segmentation of the piece,
- clusters of frames built from the eigenvectors contain similar, repetitive sound sequences.

The goal of Valentin Gillot's PhD is to consolidate these hypotheses and investigate further the topic of Graph Signal Processing for music, with more powerful conceptual tools and experiments at a larger scale.

The core of the work will consist in designing a methodology and implement an evaluation framework so as to (i) compare different descriptors and similarity measures and their capacity to capture relevant structural information in music pieces or collection of pieces, (ii) explore the structure of musical pieces by refining the frame clustering process, in particular with a multi-resolution approach, (iii) identify salient characteristics of graphs in relation to mid-level structure models and (iv) perform statistics on the typical properties of the similarity graphs on a large corpus of music in relation to music genres and/or composers.

By the end of the PhD, we expect the release of a specific toolbox for music composition, remixing and repurposing using the concepts and algorithms developed during the PhD. First results obtained this year in music recomposition have proven very conclusive [32].

SEMAGRAMME Project-Team

6. New Results

6.1. Syntax-Semantics Interface

Participants: Philippe de Groote, Sylvain Pogodalla, William Babonnaud.

6.1.1. Abstract Categorial Grammars

We have worked on implementing parsing optimization to the Abstract Categorial Grammar tool kit. These optimizations are based on Datalog program rewriting techniques, in particular a general version of Magic Sets [27], [36]. Theses optimizations rely on the tree isomorphism between derivation trees resulting from parsing with a given abstract categorial grammar, and proofs of facts in a corresponding Datalog program. Because magic rewriting breaks the isomorphism, a transformation of proofs back to derivation trees has been proposed.

6.1.2. Lexical Semantics

The lexicon model underlying Montague semantics is an enumerative model that would assign a meaning to each atomic expression. This model does not exhibit any interesting strucuture. In particular, polysemy problems are considered as homonymy phenomena: a word has as many lexical entries as it has senses, and the semantic relations that might exist between the different meanings of a same word are ignored. To overcome these problems, models of generative lexicons have been proposed in the literature. Implementing these generative models in the realm of the typed λ-calculus necessitates a calculus with notions of subtyping and type coercion. In this context, we have investigated several ways of expressing coercion using record types, and intersection types. In addition, William Babonnaud has shown how the structure of a generative lexicon may be formalized in type theory, using the categorical notion of a topos [10].

6.2. Discourse Dynamics

Participants: Maxime Amblard, Clément Beysson, Maria Boritchev, Philippe de Groote, Bruno Guillaume, Pierre Ludmann, Michel Musiol.

6.2.1. Dynamic Logic

We have enriched our type-theoretic dynamic logic in several directions in order to take into account more dynamic phenomena. In particular, we have continued to study the dynamic properties of determiners in order to systematically capture their semantics by defining an appropriate notion of dynamic generalized quantifier. To this end, Clement Beysson has studied several issues raised by the modeling of plural determiners, which necessitates to introduce plural discourse referents that can be formalized as second-order bound variables.

6.2.2. Dialogue Modeling

Maxime Amblard and Maria Boritchev have developed a dynamic model of dialogue. We have focused on the relation between question and answers and on building a resource based on settlers of Catan game records (the DiNG corpus).

We presented in [12] research on a compositional treatment of questions in a neo-Davidsonian event semantics style. [28] presented a dynamic neo-Davidsonian compositional treatment of declarative sentences. Starting from complex formal examples, we enriched Champollion's framework with ways of handling phenomena specific to question-answer pair representation. Maria Boritchev gave two presentations on these issues [16], [21].

In [9], we presented a taxonomy of questions and answers based on real-life data extracted from spontaneous dialogue corpora. This classification allowed us to build a fine-grained annotation scheme, which we applied to several languages: English, French, Italian and Chinese. In [13], we presented an annotation scheme for classifying the content and discourse contribution of question-answer pairs. We proposed detailed guidelines for using the scheme and applied them to dialogues in English, Spanish, and Dutch. Finally, we have reported on initial machine learning experiments for automatic annotation.

In another direction, Maxime Amblard has started a common work with Chloé Braud on Formal and Statistical Modelling of dialogue. To this end, we have started with Chuyuan Li to design a dialogue model to structure the different necessary linguistic informations for interaction. This model will be implemented in a tool that finely manages interaction through formal and learning strategies.

6.2.3. Pathological Discourse Modelling

Michel Musiol has obtained a full-time delegation in the Semagramme team. This proximity makes it possible to set up a more active collaboration on the issue of pathological discourse modeling. He has worked on the development of the possibility of testing his conjectures on the cognitive and psychopathological profile of the interlocutors, in addition to information provided by the model of ruptures and incongruities in pathological discourse. This methodological system makes it possible to discuss, or even evaluate, the heuristic potential of the computational models developed on the basis of empirical facts.

As part of the work carried out in the SLAM project, Maxime Amblard, Michel Musiol and Manuel Rebuschi (*Archives Henri-Poincaré, Université de Lorraine*) continue to work on modelling interactions with schizophrenic patients. We published an article about the corpus [20]. We are writing a book on these issues, in particular, we wrote a long introduction [19]. Maxime Amblard and Michel Musiol were awarded by an Inria Exploratory Action on this issues ODiM. This year we recruited the project's collaborators. In addition, we started the constitution of a new resource.

6.3. Common Basic Resources

Participants: Maxime Amblard, Clément Beysson, Philippe de Groote, Bruno Guillaume, Guy Perrier, Sylvain Pogodalla, Karën Fort.

6.3.1. Corpus Annotation

The Universal Dependencies project (UD) aims at building a syntactic dependency scheme which allows for similar analyses for several different languages. Bruno Guillaume and Guy Perrier are active in the UD community, and participate to the development and the improvement of the French data in this international initiative. Bruno Guillaume converted a new French treebank into UD: the French Question Bank (FQB), developped by Djamé Seddah and Marie Candito [35]. With the conversion system described in [2], the corpus UD_French-FQB was introduced in version 2.4 of UD in May 2019.

Bruno Guillaume, Marie-Catherine de Marneffe (Ohio State University, Columbus, Ohio, USA) and Guy Perrier improved the consistency of two French corpora annotated with the UD scheme [6]. They improved the annotations of the two French corpora to render them closer to the UD scheme, and evaluated the changes done to the corpora in terms of closeness to the UD scheme as well as of internal corpus consistency.

Bruno Guillaume and Guy Perrier developed and popularized the use of the GREW tool for various language applications and more particularly the pattern matching module Grew-match [22], [26], [17].

SUD is an annotation scheme for syntactic dependency treebanks, that is almost isomorphic to UD (Universal Dependencies). Contrary to UD, it is based on syntactic criteria (favoring functional heads) and the relations are defined on distributional and functional bases. In [14], Kim Gerdes (Sorbonne nouvelle, Paris 3), Bruno Guillaume, Sylvain Kahane (*Université Paris Nanterre*) and Guy Perrier recalled and specified the general principles underlying SUD, presented the updated set of SUD relations, discussed the central question of Multiword Expressions, and introduced an orthogonal layer of deep-syntactic features converted from the deep-syntactic part of the UD scheme.

6.3.2. FR-FraCas

Maxime Amblard, Clement Beysson, Philippe de Groote, Bruno Guillaume, Sylvain Pogodalla and Karën Fort carried on the development of FR-FraCas, a French version of the FraCas test suite [31] which is an inference test suite, in English, for evaluating the inferential competence of different NLP systems and semantic theories. There currently exists a multilingual version of the resource for Farsi, German, Greek, and Mandarin. Sémagramme completed the first translation into French of the test suite. The latter has been publicly released [0]. We also ran an experiment in order to test both the translation and the logical semantics underlying the problems of the test suite. The experiment was run with 18 French native speakers. Such an experiment provides a way of checking the hypotheses made by formal semanticists against the actual semantic capacity of speakers (in the present case, French speakers), and allows us to compare the results we obtained with the ones of similar experiments that have been conducted for other languages [30], [29].

[0] https://gitlab.inria.fr/semagramme-public-projects/resources/french-fracas

Auctus Team

7. New Results

7.1. Posture and motion capture by smart textile

The objective of Postex is to design an intelligent textile jacket, without the use of additional sensors, to determine an operator's posture.

Since 2017, we offer an innovative solution based on the electrical properties of a conductive stretch fabric that is used in the manufacture of an intelligent garment. We use Electrical Impedance Tomography (EIT) to identify fabric deformations. A neural network is used to correlate the different postures and movements measured using a reference device with the electric field measured in the intelligent textile.

By 2018, we had successfully identified the movements of an elbow and then filed a patent under the number FR1860192. In 2019, we identified the shoulder and worked on the design of the fabric parts. In order to valorize the technology, the Touch Sensity startup was created at the end of 2019.

7.2. Set-based evaluation of robot capabilities

Set-based approaches allow to model serial mechanisms with varying levels of geometric uncertainties. The Kinematic Chain Appropriate Design Library (KCADL) has been created for the purpose of modelling imprecise serial kinematic chains and provides numerous certified methods, implemented using the IBEX interval analysis library, for analyzing the capabilities of these modelled mechanisms. The KCADL software provides a set of public routines to build arbitrary serial mechanisms by incrementally adding rigid-body segments with associated parent-child uncertainties. Efficient Forward Kinematic (FK) and Inverse Kinematic (IK) solvers have been formulated and integrated into the software. These solvers are fully compatible with set-based inputs and are capable of handling sets of poses or sets of joint configurations. In addition to the FK and IK solvers, analysis routines which are applicable to imprecise kinematic chains with set-based inputs are also implemented in the software (e.g., evaluating the mechanism's: force/velocity/acceleration capabilities, precision). These routines provide offline analysis and design tools as well as online real-time capable tools for reliably evaluating current and future capabilities.

7.3. Redundancy tube

A set-based approach for modelling the human upper-limb and its complex articular constraints as a 7-degree-of-freedom (dof) constrained imprecise kinematic chain is formulated and implemented in the AUCTUS-RT software. This software allows to easily customize the geometric parameters and articular constraints. This permits to adapt the upper-limb model to each unique human subject and may also be used to model sets of human subjects. Various visualization tools are available for Python and Matlab to aid in the selection of appropriate articular constraints. When given a task with m redundant dofs and the desired workspace resolution, the software is capable of computing certified inside and outside regions of the (m+1)-dimensional redundant workspace associated with the task and upper-limb model. When a temporal dimension is added to the task description, the redundant workspace varies over time and produces a tube of redundant motions, which we refer to as the redundancy tube. The software accepts spatial-temporal task descriptions and allows to compute the full redundancy tube or slices of the tube. Furthermore, the software allows to model individual and/or sets of trajectories to describe tasks exactly or with varying level of uncertainties. Much effort has been put into improving the efficiency of the redundancy tube evaluations and parallel computing, both locally and non-locally via PlaFRIM, using OpenMP is supported. The AUCTUS-RT software is currently being used for the ongoing study of human motor-variabilities for the AUCTUS Mover project.

Related publications: [16]

7.4. Mover project

The Mover project is an experiment-based project to evaluate and study the links between human motor-variability, expertise, and fatigue associated with repetitive tasks. A wireloop game serves as the experimental task where a human subject is tasked with moving a metal ring along a fixed conductive wire while trying to maximize a score which is a function of the task time and the number of ring-wire contacts. To more easily study motor-variabilities, the wireloop ring is actively constrained by a collaborative robot (the Frank Emika Panda), allowing to isolate desired task variabilities and easily modify the task (e.g., through applied disturbances, changes in robot stiffness, task guidance). A preliminary version of the experiment is currently being developed to test all aspects of the project (i.e., experimental protocol, robot control, motion capture, human modelling, and redundancy tube evaluation). All experimental aspects of the project will be finalized before March 2020.

7.5. Interactions with a chatbot

In the context of the CIFRE Orange PhD work by Nicolas Simonazzi under the supervision of Jean-Marc Salotti and with the objective of analyzing and identifying emotions during interactions with a chatbot, a first experiment was conducted. It involved a user, the use of a smartphone, viewing videos and asking questions about the content of the video and the feeling of the user just after the answer. The collected data were numerous: the accuracy of the answers to the questions, the emotional feeling (choice of emoticons by the user) as well as the real-time measurements of the accelerometer of the smartphone. An analysis of the data was carried out with Russell's relatively simple emotional model as an explanatory framework based on two variables, the positive or negative valence of the emotion, and the degree of excitement. The experimental results showed that there was a slight correlation between the valence indicated by the user, the accuracy of the answers to the questions and the accelerations of the smartphone. However, it was hoped that the videos would have an impact on the valence, because their content had an intrinsic valence, but it proved impossible to find a correlation with the valence indicated by the user, probably due to a lack of the user engagement and also because of the focus on the questions that followed and the accuracy of the answers. A new experimental protocol is currently being studied with a priori more impactful videos (likely to produce an emotion with a greater degree of excitement).

Related publications:[10], [19]

7.6. Prediction of human error in robotics

The INRS provides a database of accidents at work, from which we can extract those concerning robotics. Many accidents are due to a deterioration of situational awareness. However, as there are many different causes and human factors are not well understood, it is very difficult for experts to provide probabilistic risk assessments. We proposed to simplify the problem by classifying the accidents according to the main demons that degrade the consciousness of the situation (Endsley model) and to use a Bayesian approach with the Noisy-Or nodes. We had already tried such an interpretation in the field of aeronautics. We propose to extend it to the field of robotics. Even if the approach remains empirical and approximate, it is possible to infer general probabilities of risk of human error leading to accidents and to deduce actions to reduce risks.

Related publications: [18]

7.7. Securing industrial operators with collaborative robots: simulation and experimental validation for a carpentry task

In this work, a robotic assistance strategy is developed to improve the safety in an artisanal task that involves a strong interaction between a machine-tool and an operator. Wood milling is chosen as a pilot task due to its importance in carpentry and its accidentogenic aspect.

In order to analyze the wood milling task, a wood shaping training was conducted in collaboration with a carpentry learning institute which allowed to collect information related to the task (perceived effort, position of the operator, accident circumstances).

To analyze the human-machine interaction, a formalization of the problem as a dynamic exchange of spatial forces inspired by the grasping theory has been performed. This theory presents structural similarities with the studied task. Based on this formalization, a behavior simulator of the system "wood + human + tool" has been developed.

To propose a credible and a realistic assistance solution, accidentogenic situations are simulated (see Woobotsim). Based on the observation made with these simulations, the use of a collaborative robot to secure wood instability cases has been explored and validated by an experiment. An operational space damping behaviour appears to be the most appropriate solution to improve safety in the studied cases.

The experiment was designed to reproduce two cases of instability during a carpentry milling task based on the entry and exit of the tool into and out of a wood node. For safety reasons, the experiment is performed on a safe but tangible simulation of the task. We then show how a robot ((Franka Emika's Panda, 7-DOF)) controlled in torque can instantly stabilize the wood to avoid an accident without modifying the carpenter's sensations.
Related publications: [22]

7.8. A software architecture for the control of a 7 dof robot for the conduct of several experiment

The Franka Emika Panda robot is a 7 dof robot. Using the Robot Operating System (ROS), an experimental setup has been built to exploit this robot. The experimental setup consists in:

- the Panda robot;
- several RGBD sensors (Kinects);
- a safety laser scanner.

Dedicated algorithms have been developed to exploit the capacities of the Kinects to visualize the environment surrounding the robot and compute the distance to the closest obstacle. Several Kinects can be used simultaneously. Specific drivers have been developed to exploit the data given by the laser scanner to also determine the closest distance between a human and a robot.

Within the framework of Arcol (see 6.6) A software architecture has been developed to ease the development of different controllers. The robot can be controlled in joint position, velocity and torque using standard state-of-the-art control technique or constrained convex optimization methods. Trajectories can be easily defined, played, and modified at run time. The robot can be simulated using the GAZEBO dynamic simulator or run on the real robot with similar behaviours. All this software architecture works on a real-time patched computer.

7.9. Modulation of the robot velocity capabilities according to the distance to a human

Using the setup described in 7.8 , a controller has been defined to constrain the robot maximum velocity according to the distance to a human. The aim of this work is to be able to use the robot optimally at all time. When a human comes near the robot workspace, the robot must stop to avoid any dangerous interaction. Several strategies exist to reduce the robot velocity as a function of the distance to the human. In this work, the goal is to determine the maximum deceleration capabilities of the robot in real time and determine if the robot has the capacity to stop before a contact. If not, the maximum allowable joint velocities of the robot are reduced. When the human reaches the robot workspace, these joint velocities must be null to ensure safety. Simply reducing the joint velocity of the robot without modulating the trajectory would induce a bad tracking of the robot task. Hence, the trajectory is updated in real time to take into account the capacities of the robot. This work has been submitted for publication at the ICRA 2020 conference.

Related publications: [23]

7.10. Human motion analysis in ecological environment

The estimation of human motion from sensors that can be used in an ecological environment is an important issue being it for home assistance for frail people or for human/robot interaction in industrial contexts. We are continuing our work on data fusion from RGB-D sensors using extended Kalman filters. The original approach uses a biomechanical model of the person to obtain anthropomorphically constrained joint angles to make their estimation physically coherent. In addition, we propose a method for the optimal adjustment of the covariance matrices of the extended Kalman filter. The proposed approach was tested with six healthy subjects performing 4 rehabilitation tasks. The accuracy of the joint estimates was evaluated with a reference stereophotogrammetric system. Our results show that an affordable RGB-D sensor can be used for simple home rehabilitation when using a constrained biomechanical model. This work has led to the writing of an article now in submission to the MBEC (Medical & Biological Engineering & Computing).

In a second step, we compared the joint centre estimates obtained with the new Kinect 3 (Azure Kinect) sensor, the Kinect 2 (Kinect for Windows) and a reference stereophotogrammetric system. Regardless of the system used, we have shown that our algorithm improves the body tracker data. This study also shows the importance of defining good heuristics to merge the data according to the body tracking operation. This study is submitted for publication at ICRA 2020.

7.11. Human motion decomposition

The aim of the work is to find ways of representing human movement in order to extract meaningful physical and cognitive information.

After the realization of a state of the art on human movement, several methods are compared: principal component analysis (PCA), Fourier series decomposition and inverse optimal control.

These methods are used on a signal comprising all the angles of a walking human being. PCA makes it possible to understand the correlations between the different angles during the trajectory. Fourier series decomposition methods are used for a harmonic analysis of the signal. Finally, inverse optimal control sets up a modeling of the engine control to highlight qualitative properties characteristic of the whole motion. These three methods are tested, combined and compared on data from the EWalk database (http://gamma.cs.unc.edu/GAIT/#EWalk) in order to test emotion recognition based on these decompositions and simple classifiers.

7.12. New method for cobotic task evaluation

Two industrial studies allowed us improving our methodology for cobotic task evaluation.

- Thanks to the partnership with Suez and the work of ENSC student Nina Docteur under Auctus supervision, there are several interesting results: first, the methodological approach has been reinforced. There was a detailed analysis of an accident-prone gesture (the pipe cover raising), meetings with field agents, supervisory teams, discussions with SUEZ ergonomic expert and field observations. Second, there was a theoretical framework - a model - for the general evaluation of a cobotic task, as well as the exploration of rules for evaluating the components of this model. Five main components have been proposed for the evaluation: bio-mechanics, cognitics, usability, hedonism and social impact. An important difficulty was to mix every component and to unify the evaluation. In order to mature the model, it has been decided to carry on the partnership with a PhD work.
- The PORTAGE project (Plateforme de RoboTisation et d'Automatisation GénériquE de bâtis industriels) involves AKKA Technologies, Ez-Wheel, IIDRE and IMS laboratory. It aims at developing semi-autonomous solutions for moving heavy structures within industrial environments (e.g. aircraft industry). Our contributions is concerned with human-robot interactions, and especially accounting for real-life constraints of operators' job within their industrial environment, and translate them into design choices and requirements for the to-be-developed robotic solution. In order to identify relevant elements from the work situation, three Human Factors models have been used: Reason's model

[54], Situation Awareness model [27], and Skill Rule Knowledge (SRK) model [53]. The Reason's model details the different layers to explain accidents, notably in the aircraft industry. These layers gather equipment, procedure, training, management, policies and even psychological precursors of the operator. Therefore, this model allows investigations on potential latent causes of accident in complement with "obvious" patent causes usually more easily identified. The Situation Awareness model of Endsley describes cognitive mechanisms involved in a given situation for a person when performing actions, based on 1) the perception of elements of the current situation, 2) the understanding of the current situation, and 3) the projection of the future status of the situation. This model leads to identifying 8 daemons where situation awareness can be deteriorated, potentially resulting in accidents. The SRK model describes the decision process of an operator (or any person) performing a given task, based on his/her familiarity with this task. This model, coupled with the Situation Awareness model, can be leveraged to identify elements to be accounted when developing collaborative robots in industrial environments.

7.13. Situation awareness analysis

Baptiste Prébot, PhD student under the direction of Jean-Marc Salotti developed a methodology to analyze and assess representation sharing and situation awareness in groups of humans, possibly involving robots or artificial intelligence systems. An experiment has been carried out with two persons and a vehicle. The first person was assigned the role of mission control and the second person the role of an astronaut driving a vehicle using a real driving interface but in a simulated environment (surface of Mars in virtual reality). The two persons were located in different rooms and could communicate only by voice. Mission control had to guide the driver to a specific location. Every minute, the experiment was stopped and the two persons were asked to make a cross on a map corresponding to what they believe was the location of the vehicle. Comparisons of crosses on the maps, including ground truth locations, enabled us to determine the exactness of the localization and the degree of correct sharing of the situation. This experiment helped us better understanding communication and sharing issues, which are particularly relevant for the design of tasks and procedures for robotic operations.

Related publications:[12], [13], [9]

CHORALE Team

6. New Results

6.1. Task based world modeling and understanding

6.1.1. Hidden robot

Participants: John Thomas (Master student), Philippe Martinet, Paolo Salaris, Sébastien Briot (LS2N-ARMEN)

When robots want to execute a task, they require to have an adequate representation of the environment where they will evolve. In model based approach, it is classical to describe environment using Metric Map where the function of perception (localization) and control (Path or trajectory tracking) refer to Cartesian state. In sensor based control, the methodology "teaching by showing" has been developed during the last 30 years. The concept of sensory memory has been then introduced in order to represent the task to be executed in sensor space. This concept is used in order to represent the task directly in the sensor space for a particular set of sensors. In summary, building the representation of the task (or the environment) is building the sensory memory, defining a particular motion (or trajectory) is defining a particular occurrence of sensor features, and executing the task is done when a control is designed to perceive the same as stored in the sensory memory. This approach has shown great ability in terms of robustness. However, it is still difficult to analyze the singularities and to demonstrate the stability property for those approaches (mainly when it is necessary to control 6 degree of freedom). In 2013, Sébastien Briot and Philippe Martinet have studied the visual servoing scheme of a Gough-Stewart Platform [18] and shown that it exists an hidden robot in the controller that can be used to study the behaviour and properties of it. The Hidden robot allows to transform the analysis of the controller by viewing it as a parallel robot. Recently, this concept has been applied to study the singularities of the visual servoing scheme of points and of lines [19]. This work continues in the framwork of the ANR project SESAME.

The idea of the new initial work done in 2019, is to find a methodology to design a task by using the Hidden robot concept. Navigation of a mobile robot has been considered in a first time. The followed methodology considers a topological navigation framework where a successive interaction situation are modelled by using an hidden robot: in some words, navigation is done by using a set of successive hidden parallel robots holding the robot when moving. At least two main question have been identified: What is the structure of the virtual robot which fits to task to be done? and Where to fix (or How to select) the anchors of this virtual robots?

For the first question, the idea is, considering different kind of features, to define a virtual parallel robot based on virtual legs. These virtual legs are directly linked to the considered feature. We have studied two cases, distance and angle, considering that existing sensors allow us to obtain the corresponding extracted features. After the modeling of sensors features, different control laws have been investigated allowing to produce motion of the mobile plateform. The corresponding hidden robots and the properties have been studied.

For the second question, two methods have been investigated using selection matrix of features or weighted features. The main used criteria is the transmibility index which relates the faculty of motion transmission of the virtual parallel mecanism.

This work [52] is preliminary and on going. We already have obtained preliminary results in simulation allowing a mobile plateform to evolve in a dedicated environment. It was the work done by John Thomas under the supervision of Philippe Martinet and Paolo Salaris.

6.1.2. End to end navigation

Participants: Renato Martins (Post-doc), Patrick Rives

This research deals with the problem of end-to-end learning for navigation in dynamic and crowded scenes solely from visual information. We investigate the problem of navigating an unknown space to reach a target of interest, for instance “doors", exploring the possibilities given by data-driven based models in the context of ANR MOBI-Deep project around the guidance of visually impaired people. A successful agent navigation policy requires learning general relationships between the agent actions, safety rules and its surrounding environment. We started studying a simple guidance model (turn left, right or stop), whose guidance task is to remain inside a specific region of the scene (to avoid collusion). This is equivalent to take the action to stay in the center of a corridor (indoor scene) or road (outdoor scenario). We first evaluate a relatively small supervised net composed of sixteen ResNet convolutional layers. This model was trained with real images from the Udacity autonomous driving challenge, but presented limited generalization when tested in either non-structured scenes or in scenes with humans. In order to overcome these limitations, we plan to train an A3C agent (Asynchronous Actor-Critic Agent) to learn the action policies in a reinforcement learning scheme, using data acquired of virtual environments with crowds. We also plan to evaluate the use of inputs from different levels as: scene semantic segmentation; depth inference from monocular images; and human and object detection information in the learning scheme.

6.1.3. Semantization of scene

Participants: Mohammed Boussaha (PhD, IGN), E. Fernandez-Moral, R. Martins, Patrick Rives
The work carried out in the ANR PlaTINUM project concerns the semantic labeling of images [17] acquired by agents (autonomous vehicles or pedestrians) moving in an urban-like environment and their accurate localization and guidance. A semantic labeling based on a machine learning approach (CNN) was developed. A same methodology is used to semantize virtual images built from a textured 3D mesh representation of the environment and images from the camera handled by the agents. Several strategies have been studied to exploit complementary information, such as color and depth for improving the accuracy of semantization. Our results show that exploiting this complementarity requires to perfectly align the different sources of informations. We proposed a new approach to the problem of calibration of heterogeneous multi sensors systems [41], [44]. We also looked for evaluating a new metric to quantify the accuracy of semantization provided by the CNN by taking into account the boundaries of semantized objects during the learning step. As a consequence, we show that weighting the boundary pixels in the images allows to segment more clearly the navigable areas used by different agents such as pedestrians (sidewalks) and cars (road). The results of this research were published in [29], [30]. The CNN used for labeling imagesacquired by different image sensors (perspective and spherical) was pre-trained from public datasets with perspective images of urban-like environments (simulated or real). In the context of the Platinum ANR project, a fine-tuning was done with some spherical images acquired in Rouen by the IGN Stereopolis vehicle and then hand-labelled. A Docker version of the software has been made available on the project server in order to be used by the other partners.

A localization method has also been implemented to exploit information of color, depth and semantics (when this information is available). An estimation of the agent position (6DOV, rotation and translation) is computed thanks to a dense method that minimizes the geometric, photometric and semantic differences between a spherical view provided by a SIG (Système d’Information Géographique) data base hosted in a cloud server and the current view of the agent.

During the last year of the project, the methods developed in PlaTINUM were consolidated and validated on the data acquired in Rouen. As originally planned in the project, Inria enlisted the help of iXblue-division Robopec to integrate the various functions developed during the project. This software, called Perception360, will be from now the software platform for all perception developments in the Inria CHORALE team.

6.1.4. Optical Flow Estimation Using Deep Learning In Spherical Images

Participants: Haozhou Zhang (Master), Cédric Demonceaux (Vibot), Guillaume Allibert
In a complex environment such as in a forest, the autonomous navigation is a challenging problem due to many constraints such as the loss of GPS signals because dense and unstructured environments (branches, foliage, ...) reduce the visibility. Without GPS signals, a vision system with the ability to capture everything going on around you seems more valuable than ever and crucial to navigate in this environment. Spherical images offer great benefits over classical cameras wherever a wide field of view is essential.

The equirectangular projection is a popular representation of images taken by spherical cameras. In this projection, the latitude and longitude of the spherical images are projected to horizontal and vertical coordinates on a 2D plane. However, this equirectangular projection suffers from distortions, especially in polar regions. In this case, the density of features is no longer regular at different latitudes of the images. As a result, traditional image processing methods that have been used for perspectives images do not have good performance when they are applied to equirectangular images.
Optical flow estimation is a basic problem of computer vision [50]. It is generally used as input of algorithm for autonomous navigation. Given two successive images, it estimates the motion vector in 2D (in x and y direction) for each pixel from between the two input images. Optical flow is usually considered as a good approximation of the true physical motion mapped on the image plane. It provides a concise description of the direction and velocity of the motion. In [24] and [36], CNNs which are capable of solving the optical flow estimation problem as a supervised learning task are proposed and became the standard for optical flow estimation. However, the dataset used to train [24], [36] is only based on perspective images. Even if they can be used directly with spherical images as input, the high distortions coming from equirectangular projection drastically reduce the global performance of these networks. One possible way to solve this issue is to train the networks proposed in [24], [36] with spherical images. Unfortunately, these databases do not exist and generating them would be a long and costly process.

In the Master's Hoazhou [55], we have proposed a solution to overcome this issue in proposing an adaptation of FlowNet networks to deal with the distortions in the equirectangular projection of spherical images. The proposed approach lies a distortion aware convolution used as convolution layers in the network to deal with distortions in equirectangular images. The proposed networks allows the models to be trained by perspective images and be applied to spherical images using an adapted convolution which is coherent with the spherical image. This solution avoids training a large number of spherical images which is not available and costly to generate.

6.2. Multi-sensory perception and control

6.2.1. Autonomous Parking Maneuvers

Participants: David Perez Morales (PhD, LS2N-ARMEN), Olivier Kermorgant (LS2N-ARMEN), Salvador Dominguez Quijada (LS2N-ARMEN), Philippe Martinet
Automated parking is used as new functionality to sell different model of cars right now. Mainly, the different versions of parking abilities are not autonomous and are based on motion planning only. There is no ability to evolve in dynamic environment: it remains automated in static environment, or even an assistant to park under the control of the driver. The purpose of the PhD work of David Perez Morales was to investigate how the problem of autonomous parking by using different sensor based techniques is able to handle any kind of parking situations (parallel, perpendicular, diagonal) for parking an unparking (backward and forward).
Two different frameworks has been developed. The first framework, using a Multi-Sensor-Based Control (MSBC) approach [47], [48], [46], [45] allows to formalize different parking and unparking operations in a single maneuver with either backward or forward motions. Building upon the first one and by using an MPC strategy [49], a Multi-Sensor-Based Predictive Control (MSBPC) framework has been developed, allowing the vehicle to park autonomously (with multiple maneuvers, if required) into perpendicular and diagonal parking spots with both forward and backward motions and into parallel ones with backward motions in addition to unpark from parallel spots with forward motions. These frameworks have been tested extensively using a robotized Renault ZOE with positive outcomes and now they are part of the autonomous driving architecture being developed at LS2N.

In 2019, the main focus was on MSBPC, and on taking into account the dynamic aspect in the environment (mainly pedestrians). Detection and tracking for pedestrian has been included in the perception aspect, in parallel to the detection of empty spots for parking. An additional terms has been added as a contraint in the cost function to be minimized in order to take into account the dynamic aspect, and a mecanism has been put in place in order to switch automatically the maneuver. In presence of pedestrian, an additional maneuver is engaged, which is what human are generally doing if place is enough for performing safely the maneuver. Comparison with state of the art motion planning approach have been done in simulation. The proposed method have demonstrated the efficiency while the others fails in a very long set of maneuvers. Real experiments have been done also in presence of pedestrians.

6.2.2. Platoon control and observer

Participants: Ahmed Khalifa (Post-Doc, LS2N-ARMEN), Olivier Kermorgant (LS2N-ARMEN), Salvador Dominguez Quijada (LS2N-ARMEN), Philippe Martinet

In the framework of the ANR Valet project, we are interested in platooning control of cars for a service of VALET Parking where it is necessary to join a platoon (after unparking), to evolve among the platoon, and leave the platoon (for parking). We are considering the case when the leader is autonomous (following an already defined path) or manually driven by a human (the path must be build on line). The lateral controller to follow a path has been designed earlier [23] and the localization technique largelly evaluated experimentally [33]. The main exteroceptive sensor is the Velodyne VLP16.

The first work [38] [15] concerned the design of a distributed longitudinal controller for car-like vehicles platooning that travel in an urban environment. The presented control strategy combines the platoon maintaining, gap closure, and collision avoidance functionality into a unified control law. A consensus-based controller designed in the path coordinates is the basis of the proposed control strategy and its role is to achieve position and velocity consensus among the platoon members taking into consideration the nature of the motion in an urban environment. For platoon creation, gap closure scenario is highly recommended for achieving a fast convergence of the platoon. For that, an algorithm is proposed to adjust the controller parameters online. A longitudinal collision between followers can occur due to several circumstances. Therefore, the proposed control strategy considers the assurance of collision avoidance by the guarantee of a minimum safe inter-vehicle distance. Convergence of the proposed algorithm is proved in the different modes of operations. Finally, studies are conducted to demonstrate and validate the efficiency of the proposed control strategy under different driving conditions. To better emulate a realistic setup, the controller is tested by an implementation of the car-like vehicles platoon in a vehicular mobility simulator called ICARS, which considers the real vehicle dynamics and other platooning staff in urban environments.

The second work [14] adresses the problem of controlling the longitudinal motion of car-like vehicles platoon navigating in an urban environment that can improve the traffic flow with a minimum number of required communication links. To achieve a higher traffic flow, a constant-spacing policy between successive vehicles is commonly used but this is at a cost of increased communication links as the leader information must broadcast to all the followers. Therefore, we propose a distributed observer-based control law that depends on a hybrid source of neighbours information in which a sensor-based link is used to get the predecessor position while the leader information is acquired through a communication-based link. Then, an observer is designed and integrated into the control law such that the velocity information of the predecessor can be estimated. We start by presenting the platoon model defined in the Curvilinear coordinates with the required transformation between that coordinate and the Cartesian Coordinates so that one can design the control law directly in the Curvilinear coordinates. After that, internal and string stability analysis are conducted. Finally, we provide simulation results, through dynamic vehicular mobility simulator called ICARS, to illustrate the feasibility of the proposed approach and corroborate our theoretical findings.

Both work have been tested in real with a platoon of 3 up to 4 cars.

6.2.3. High speed visual servoing

Participants: Franco Fusco (PhD, LS2N-ARMEN), Olivier Kermorgant (LS2N-ARMEN), Philippe Martinet

Controlling high speed robot with visual feedback may require to develop more complex models including the dynamics of the robots and the environment. Some previous work done in the field of dynamic visual feedback of parallel robots [42] have demonstrated the efficiency regarding the classical Joint computed torque control. Also, it has been shown that it is also possible to develop more complex interaction models [20].
In recent years, many efforts have been dedicated to extend Sampling-based planning algorithms to solve problems involving constraints, such as geometric loop-closure, which lead the valid Configuration Space to collapse to a lower-dimensional manifold. One proposed solution considers an approximation of the constrained Configuration Space that is obtained by relaxing constraints up to a desired tolerance. The resulting set has then non-zero measure, allowing therefore to exploit classical planning algorithms to search for a path that connects two given states. When the constraints involve kinematic loops in the system, relaxation generally bears to undesired contact forces, which needs to be compensated during execution by a proper control action. We propose a new tool that exploits relaxation to plan in presence of constraints [32]. Local motions inside the approximated manifold are found as the result of an iterative scheme that uses Quadratic Optimization to proceed towards a new sample without falling outside the relaxed region. By properly guiding the exploration, paths are found with smaller relaxation factors and the need of a dedicated controller to compensate errors is reduced. We complete the analysis by showing the feasibility of the approach with experiments on a real manipulator platform.
The commonly exploited approach in visual servoing is to use a model that expresses the rate of change of a set of features as a function of sensor twist. These schemes are commonly used to obtain a velocity command, which needs to be tracked by a low-level controller. Another approach that can be exploited consists in going one step further and to consider an acceleration model for the features. This strategy allows also to obtain a natural and direct link with the dynamic model of the controlled system. The work done in [13] aims at comparing the use of velocity and acceleration-based models in feed-back linearization for Visual Servoing. We consider the case of a redundant manipulator and discuss what this implies for both control techniques. By means of simulations, we show that controllers based on features acceleration give better results than those based on velocity in presence of noisy feedback signals.
We are working to propose new prediction models for Visual Predictive Control that can lead to both better motions in the feature space and shorter sensor trajectories in 3D. Contrarily to existing local models based only on the Interaction Matrix, it is proposed to integrate acceleration information provided by second-order models. This helps to better estimate the evolution of the image features, and consequently to evaluate control inputs that can properly steer the system to a desired configuration. By means of simulations, the performances of these new predictors are shown and compared to those of a classical model. Real experiments confirm the validity of the approach and show that the increased complexity.

6.2.4. Proactive and social navigation

Participants: Maria Kabtoul (PhD), Wanting Jin (Master), Anne Spalanzani (CHROMA), Philippe Martinet, Paolo Salaris

In the last decade, many works have been done concerning navigation of robots among humans [34], [27] or human robots interaction [22], [31]. In very few cases, a robot can realize an intention to move.

In this work, we would like that robots can express their needs for sharing spaces with humans in order to perform their task (i.e. navigation in crowdy environments). This requires to be proactive and adapt to the behavior by exploiting the potential collaborative charcateritics of the nearby environment of the robots.
In the framework of the ANR project HIANIC, Maria Kabtoul is doing her PhD on the topic Proactive Social navigation for autonomous vehicles among crowds. We consider shared spaces where humans and cars are able to evolve simultaneously. The first step done in this way is to introduce a pedestrian to vehicle interaction behavioral model. The model estimates the pedestrian's cooperation with the vehicle in an interaction scenario by a quantitative time-varying function. Then, the trajectory of the pedestrian is predicted based on its cooperative behavior. Both parts of the model are tested and validated using real-life recorded scenarios of pedestrian-vehicle interaction. The model is capable of describing and predicting agents' behaviors when interacting with a vehicle in both lateral and frontal crossing scenarios.

In the framework of the ANR project MOBI-DEEP, we have addressed the problem of navigating a robot in a constrained human-like environment. We provide a method to generate a control strategy that enables the robot to proactively move in order to induce desired and socially acceptable cooperative behaviors in neighboring pedestrians. Contrary to other control strategies that simply aim to passively avoid neighboring pedestrians, this approach greatly simplifies the navigation task for both robots and humans, especially in crowded and constrained environments. In order to reach this objective, the co-navigation process between humans and robots is formalized as a multi-objective optimization problem and a control strategy for the robot is obtained by using the Model Predictive Control (MPC) approach. The Social Force Model (SFM) is used to predict the human motion in cooperative situations. Different social behaviors of humans when moving in a group are also taken into account to generate the proper robot motion. Moreover, a switching strategy between purely reactive (if cooperation is not possible) and proactive-cooperative planning depending on the evaluation of the human intentions is also provided. Simulations under different navigation scenarios show how the proactive-cooperative planner enables the robot to generate more socially and understandable behaviors.
This work has been done by Wanting Jin during her Master thesis [37].

6.2.5. Safe navigation

Participants: Luiz Guardini (PhD), Anne Spalanzani (CHROMA), Christian Laugier (CHROMA), Philippe Martinet, Anh-Lam Do (Renault), Thierry Hermitte (Renault)
Today, car manufacturer are selling systems to brake in presence of obstacle. Those systems are based on the fact that the risk of collision is always detected and well evaluated. Their action are limited on brake only, which is in some case not sufficient to limit the risk. A global and safe system must be more efficient in environment perception awareness and also in action to be decided (break, steer, acceleration). In such a case, it is very complicated to find the best solution as long as we have to evaluate the different solutions in a near horizon in terms of risk of colisions and severity injuries. Car manufacturer are interested to find solution (i.e evaluation of trajectories (planification and action) in terms of risks and injuries.
Evaluating a scene to perform a collision avoidance maneuver is a hard task for both humans and (semi-) autonomous vehicles. There are some cases though that collision avoidance is inevitable. Interpreting the scene for a possible collision avoidance is difficult already a difficult task. Choosing how to mitigate the damage seems even harder, specially when humans have only a split of second to decide how to proceed.

Intending to decrease the reaction time and to increase safety on dangerous driving situations, one can rely on intelligent systems. Nevertheless, autonomous vehicles simulation and testing usually focus on risk assessment and path planing on regular driving conditions [40]. For instance, Waymo from Google, still do not have the full capability of avoiding collision initiated by other vehicles [28].

Developing Advanced Driver Assistance Systems (ADAS) technologies is one alternative for these emergency scenarios. It includes systems such as Active Braking System (ABS), Forward Collision Warning (FCW) and Collision Avoidance (CA). The latter is one of the most complex systems developed in order to assure safety. It perceives technologies such as Advanced Emergency Braking (AEB) and Autonomous Emergency Steering (AES) System. Those systems attempt to avoid the crash or at least reduce its severity. Developing a CA system starts by assessing the available information in the scene. This is made by establishing safe zones that the vehicle can access. The notion of safety of severity is usually addressed by the concept of risk. Risk can be intuitively understood as the likelihood and severity of the damage that an object of interest may suffer or cause in the future. Threat Assessment (also referred as Risk Assessment or Hazard Assessment) makes use of such concept.

The excellence of the data evidenced in the scene plays a major role in risk assessment and mitigation. Up to date, objects in the scene are not contextualized. For instance, pedestrians are treated as forbidden zones whereas cars are allowed to be collided when mitigation is necessary. This might be a correct assessment in some cases, but not always. The injury risk changes independently to each object according to aspects on the scene, such as the impact velocity and angle of collision.

This work focus on the development of a probabilistic cost map that expresses the Probability of Collision with Injury Risk (PCIR). On top of the information gathered by sensors, it includes the severity of injury in the event of a collision between ego and the objects in the scene. This cost map provides enhanced information to perform vehicle motion planning in emergency trajectories where collision is impending.

We represent the environment though probabilistic occupancy grids. It endures agile and robust sensor interpretation mechanisms and incremental discovery procedures. It also handles uncertainty thanks to probabilistic reasoning [25].

We use the Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT developed in CHROMA). It is a generic spatial occupancy tracker that infers dynamics of the scene through a hybrid representation of the environment. The latter consists of static and dynamic occupancy, empty spaces and unknown areas. This differentiation enables the use of state-specific models as well as relevant confidence estimation and management of dataless areas [51].

Although CMCDOT occupancy grid leads to a very reliable global occupancy of the environment, it works on a sub-object level, meaning that the grid by itself does not carry the information on object classification. To overcome this, Erkent et al [26] proposes a method, which estimates an occupancy grid containing detailed semantic information. The semantic characteristics include classes like road, car, pedestrian, sidewalk, building, vegetation, etc.

The proposed Probabilistic risk map has been built and validation has been done in simulation using Gazebo using different scenarios (identified by the car manufacturer).

6.2.6. 3D Autonomous navigation using Model Predictive Path Integral approach

Participants: Ihab Mohamed (PhD), Guillaume Allibert, Philippe Martinet

Having a safe and reliable system for autonomous navigation of autonomous systems such as Unmanned Aerial Vehicles (UAVs) is a highly challenging and partially solved problem for robotics communities, especially for cluttered and GPS-denied environments such as dense forests, crowded offices, corridors, and warehouses. Such a problem is very important for solving many complex applications, such as surveillance, search-and-rescue, and environmental mapping. To do so, UAVs should be able to navigate with complete autonomy while avoiding all kinds of obstacles in real-time. To this end, they must be able to (i) perceive their environment, (ii) understand the situation they are in, and (iii) react appropriately.

To solve this problem, the applications of the path-integral control theory have recently become more prevalent. One of the most noteworthy works is Williams's iterative path integral method, namely, Model Predictive Path Integral (MPPI) control framework Williams et al. [53]. In this method, the control sequence is iteratively updated to obtain the optimal solution based on importance sampling of trajectories. In Williams et al [54], authors derived a different iterative method in which the control- and noise-affine dynamics constraints, on the original MPPI framework, are eliminated. This framework is mainly based on the information-theoretic interpretation of optimal control using KL-divergence and free energy, while it was previously based on the linearization of Hamilton-Jacob Bellman (HJB) equation and application of Feynman-Kac lemma.

The attractive features of MPPI controller, over alternative methods, can be summarized as: (i) a derivative-free optimization method, i.e., no need for derivative information to find the optimal solution; (ii) no need for approximating the system dynamics and cost functions with linear and quadratic forms, i.e., non-linear and non-convex functions can be naturally employed, even that dynamics and cost models can be easily represented using neural networks; (iii) planning and execution steps are combined into a single step, providing an elegant control framework for autonomous vehicles.

In the context of autonomous navigation, it is observed that the MPPI controller has been mainly applied to the tasks of aggressive driving and UAVs navigation in cluttered environments. For instance, to navigate in cluttered environments, the obstacle map is assumed to be known (either available a priori or built off-line), and only static 2D floor-maps are used. Conversely, in practice, the real environments are often partially observable, with dynamic obstacles. Moreover, only 2D navigation tasks are performed, which limits the applicability of the control framework.

For this reason, our work focuses on MPPI for 2D and 3D navigation tasks in cluttered environments, which are inherently uncertain and partially observable. To the best of our knowledge, this point has not been reported in the literature, presenting a generic MPPI framework that opens up new directions for research.

We propose a generic Model Predictive Path Integral (MPPI) control framework that can be used for 2D or 3D autonomous navigation tasks in either fully or partially observable environments, which are the most prevalent in robotics applications. This framework exploits directly the 3D-voxel grid, e.g., OctoMap [35], acquired from an on-board sensing system for performing collision-free navigation.We test the framework, in realistic RotorS-based simulation, on goal-oriented quadrotor navigation tasks in a 2D/3D cluttered environment, for both fully and partially observable scenarios. Preliminary results demonstrate that the proposed framework works perfectly, under partial observability, in 2D and 3D cluttered environments.

We demonstrate our proposed framework on a set of simulated quadrotor navigation tasks in a 2D and 3D cluttered environment, assuming that: (i) there is a priori knowledge about the environment (namely, fully observable case); (ii) there is not any a priori information (namely, partially observable case); here, the robot is building and updating the map, which represents the environment, online as it goes along.

6.2.7. Perception-aware trajectory generation for robotic systems

Participant: Paolo Salaris, Marco Cognetti (PostDoc, RAINBOW), Valerio Paduano (Master, RAINBOW), Paolo Robuffo Giordano (RAINBOW)

We now focus on our planned research activities on task-oriented perception and control of a robotic system engaged in executing a task. The main objective is to improve the execution of a given task by fruitfully *coupling action and perception*. We aim at finding the correct balance between efficient task execution and quality of the information content since the amount of the latter has an impact on the possibility of correctly executing the task. Indeed, a robot needs to solve an estimation problem in order to safely move in unstructured environments and accomplishing a task. For instance, it has to self-calibrate and self-localize w.r.t. the environment while, at the same time, a map of the surroundings may be built. These possibilities are highly influenced by the quality and amount of sensor information (i.e., available measurements), especially in case of limited sensing capabilities and/or low cost (noisy) sensors.

For nonlinear systems (i.e., the most of the robotics systems of our interest) the amount and quality of the collected information depends on the robot trajectories. It is hence important to find, among all possible trajectories able to accomplish a task, the most informative ones. One crucial point in this context, also known as *active sensing control*, is the choice of an appropriate *measure of information* to be optimized. The Observability Gramian (OG) measures the level of observability of the *initial state* and hence, its maximization (e.g. by maximizing its smallest eigenvalue) actually increase the amount of information about the initial state and hence improves the performances in estimating (observing) the initial state of the robot. However, when the objective is to estimate the current/future state of the robot (which is implicitly the goal of most of the previous literature in this subject, and of our research too), the OG is *not* the right metric even if is often used in the literature for this goal. Recently, in [12], we showed that, the right metric is instead the *Constructibility Gramian* (CG) that indeed quantifies the amount of information about the current/future state, which is obviously the state of interest for the sake of motion control/task execution. We then propose an *online* optimal sensing control problem whose objective is to determine at *runtime*, i.e. anytime a new estimate is provided by the employed observer (an EKF in our case), the future trajectory that maximizes the smallest eigenvalue of the CG. We applied our machinery to two robotics platforms: a unicycle vehicle and a quadrotor UAV moving on a vertical plane, both measuring two distances w.r.t. two markers located in the environment. Results show the effectiveness of our solution not only for pure robot's state estimation, but also with instances of active self-calibration and map building.

The proposed solution is not able to cope with the process/actuation noise as CG is not able to measure its degrading effects on the current amount of the collected information and by consequence its negative effects in the estimation process. For all the cases where an EKF is used as an observer, we overcame this issue in [21] where we minimized the largest eigenvalue of the covariance matrix of the EKF that is the solution of the Riccati differential equation.

We also extended the methodology to the problem of shared control by proposing a shared control active perception method aimed at fusing the high-level skills of a human operator in accomplishing complex tasks with the capabilities of a mobile robot in maximizing the acquired information from the onboard sensors for improving its state estimation (localization). In particular, a persistent autonomous behaviour, expressed in terms of a cyclic motion represented by a closed B-Spline, is executed by the robot. The human operator is in charge of modifying online some geometric properties of this path for executing a given task (e.g., exploration). The path is then autonomously processed by the robot, resulting in an actual path that tries to follow the human's commands while, at the same time, maximizing online the acquired information from the sensors. This work has been done by Valerio Paduano during his Master thesis [43] and submitted to ICRA 2020.

Recently we are also working on extending the methodology to Multiple Robot Systems (in particular a group of quadrotor UAVs). In this context, the goal is to propose an optimal and *online* trajectory planning framework for addressing the localization problem of a group of multiple robots without requiring the rigidity condition. In particular, by leveraging our recent work on optimal online active estimation, we will propose the use of CG for quantifying the localization accuracy, and develop an *online* decentralized optimal trajectory planning able to optimize the CG during the robot motion. We particularly focus on the *online* component, since the planned trajectory are *continuously refined* during the robot motion by exploiting the (continuously converging) decentralized estimation of the robot relative poses. In order to illustrate the approach, we will consider the localization problem for a group of quadrotor UAVs measuring relative distances with maximum range sensing constraints and a decentralized Extended Kalman Filter [39] that estimates the relative configuration of each robot in the group w.r.t. a special one (randomly chosen in the group).

CHROMA Project-Team

7. New Results

7.1. Robust state estimation (Sensor fusion)

This research is the follow up of Agostino Martinelli's investigations carried out during the last five years, which are in the framework of the visual and inertial sensor fusion problem and the unknown input observability problem.

7.1.1. Visual-inertial structure from motion

Participant: Agostino Martinelli.

We have continued our study on the visual inertial sensor fusion problem in the cooperative case, with a special focus on the case of two agents. During this year, we have carried out an exhaustive analysis of all the singularities and minimal cases of this cooperative sensor fusion problem. As in the case of a single agent and in the case of other computer vision problems, the key of the analysis is the establishment of an equivalence between the cooperative visual-inertial sensor fusion problem and a Polynomial Equation System (PES). In the case of a single agent, the PES consists of linear equations and a single polynomial of second degree. In the case of two agents, the number of second degree equations becomes three and, also in this case, a complete analytic solution can be obtained [19], [20]. The power of the analytic solution is twofold. From one side, it allows us to determine the state without the need of an initialization. From another side, it provides fundamental insights into all the structural properties of the problem. The research of this year has focused on this latter issue. Specifically, we have obtained all the minimal cases and singularities depending on the number of camera images and the relative trajectory between the agents. The problem, when non singular, can have up to eight distinct solutions. The usefulness of this analysis has also been illustrated with simulations. In particular, we have quantitatively obtained how the performance of the state estimation worsens near a singularity. The results of this research will be published by the Robotics and Automation Letter (RA-L) journal [18].

7.1.2. Unknown Input Observability

Participant: Agostino Martinelli.

The Unknown Input Observability problem (UIO) in the nonlinear case was an open problem since the sixties years, when it was solved only in the linear case. In the last five years, I have obtained its general analytic solution. So far, I only published the solution for systems characterized by driftless dynamics. In particular, this solution was published as a full paper on the IEEE Transaction on Automatic Control [17]. In December 2018, I was invited by the Society for Industrial and Applied Mathematics (SIAM) to write a book with the general solution. This has been the main work of this year. Since this general solution is based on tensorial calculus (Ricci algebra) and many mathematics procedures and tricks borrowed from theoretical physics, the scope of book has gone much more beyond the presentation of the solution. Basically, by writing this book, I've obtained a new theory of observability.

The current theory of nonlinear observability, does not capture/exploit the key features that are intimately related to the concept of observability. This results in two important limitations:

- The theory, although simple and based on elementary mathematics, can be sometimes burdensome with the risk of easily loosing the meaning of the results and losing the meaning of their assumptions.
- More complex observability problems (e.g., the unknown input observability problem to which this book provides the complete analytic solution) remained unsolved for half a century.

The key to overcome the two above limitations, consists in building a new theory of observability that accounts for the **group of invariance that is inherent to the concept of observability**. This is the typical manner the research in physics has always proceeded. To this regard, I wish to emphasize that the derivation of the basic equations of any physics theory (e.g., the General Relativity, the Yang Mills theory, the Quantum Chromodynamics) starts precisely from the characterization of the group of invariance of the theory.

One of the major novelties introduced by this book is the characterization of the group of invariance of observability and, regarding the case of unknown inputs, the characterization of a subgroup that was called the *Simultaneous Unknown Input Output transformations' group*.

In summary, the book provides several novelties with respect to the existing literature in control theory. Specifically, the reader will learn the following:

- The solution of two open problems in control theory (the book provides separately the solution and the derivation), which are:
  - The extension of the observability rank condition to nonlinear systems driven by also unknown inputs.
  - The extension of the observability rank condition to nonlinear, time-variant systems (both in presence and in absence of unknown inputs)
- A new and more palatable derivation of the existing results in nonlinear observability.
- A new manner of approaching scientific and technological problems, borrowed from theoretical physics (a chapter summarizes in a very intuitive and quick manner the basic mathematics, which includes tensorial calculus).
- A new manner of dealing with the variable *time* in system theory, which is obtained by introducing a new framework, which was called the *chronospace*.

I believe this book could be an opportunity for control and information theory communities to borrow basic mathematics, tricks, types of reasoning from theoretical physics to revisit many aspects of control and information theory.

7.2. Bayesian Perception

Participants: Christian Laugier, Lukas Rummelhard, Jean-Alix David, Jerome Lussereau, Thomas Genevois, Nicolas Turro [SED], Rabbia Asghar, Mario Garzon.

Recognized as one of the core technologies developed within the team over the years (see related sections in previous activity report of Chroma, and previously e-Motion reports), the CMCDOT framework is a generic Bayesian Perception framework, designed to estimate a dense representation of dynamic environments and the associated risks of collision, by fusing and filtering multi-sensor data. This whole perception system has been developed, implemented and tested on embedded devices, incorporating over time new key modules. In 2019, this framework, and the corresponding software, has continued to be the core of many important industrial partnerships and academic contributions, and to be the subject of important developments, both in terms of research and engineering. Some of those recent evolutions are detailed below.

In 2019, the new results have been presented in several invited talks given in some of the major international conferences of the domain [30], [28], [26], [29], [27].

7.2.1. Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT) Framework

Participants: Lukas Rummelhard, Jerome Lussereau, Jean-Alix David, Thomas Genevois, Christian Laugier, Nicolas Turro [SED].

Important developments in the CMCDOT (Fig. 5), in terms of calculation methods and fundamental equations, were introduced and tested. These developments are currently being patented, and will then be used for academic publications. These changes lead to a much higher update frequency, greater flexibility in the management of transitions between states (and therefore a better system reactivity), as well as to the management of a high variability in sensor frequencies (for each sensor over time, and in the set of sensors). The changes include:

- Grid fusion: a new fusion of occupancy grids, enhanced with "unknown" variables, has been developed and implemented. The role of unknown variables has also been enlarged. Currently being patented, it should be the subject of an upcoming paper.
- Ground Estimator: a new method of occupancy grid generation, more accurately taking into account the height of each laser beam, has been developed. Currently being patented, it should be the subject of an upcoming paper.
- Software optimization: the whole CMCDOT framework has been developed on GPUs (implementations in C++/Cuda). An important focus of the engineering has always been, and continued to be in 2019, on the optimization of the software and methods to be embedded on low energy consumption embedded boards (Nvidia Jetson TX1, TX2, AGX Xavier).

Figure 5. CMCDOT results

7.2.2. Multimodal Bayesian perception

Participants: Thomas Genevois, Christian Laugier.

The objective is to extend the concept of Bayesian Perception to the fusion of multiple sensing modalities (including raw data provided by low cost sensors). In 2019, we have developed and implemented a Bayesian model dedicated to ultrasonic range sensors. For any given measurement provided by the sensor, the model computes the occupancy probability in a 2 dimensional grid around the sensor. This computation takes into account the accuracy and the possibility to "miss" an object. Thanks to various parameters, this model has been applied to the sensors of our Renault Zoe demonstrator and to the low cost sensors of our light vehicle demonstrator (flycar).

Fig. 6 .a shows an example, developed and implemented on our light vehicle demonstrator. In this example, the perception is relying on 1 lidar and 5 ultrasonic range sensors. An occupancy grid is generated for each sensor. Then they are fused in a single occupancy grid which is filtered using the CMCDOT approach.

7.2.3. Embedding deep learning for semantics

Participants: Thomas Genevois, Christian Laugier.

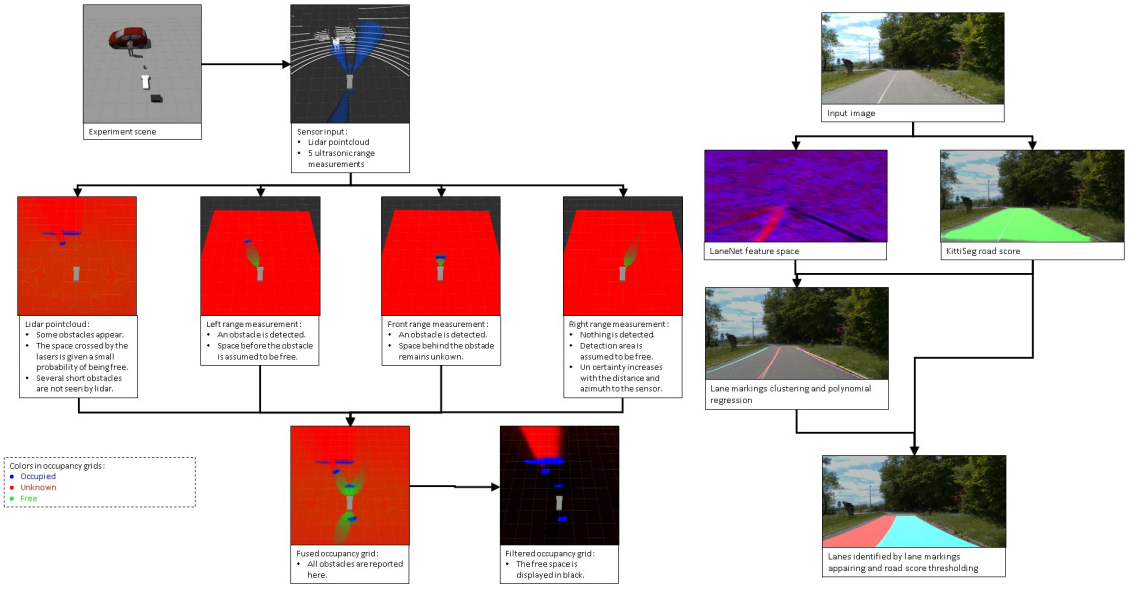


Figure 6. a. Example of multimodal perception, implemented both in simulation and on an actual vehicle demonstrator. b. Combining LaneNet and KittiSeg into a common lane recognition tool.

The objective is to improve embedded Bayesian Perception outputs in our experimental vehicle platforms (Renault Zoe and Flycar), by adding semantics obtained using RGB images and embedded deep learning approaches. In 2019, we have tested several networks for road scene semantic segmentation and implemented two of them in our vehicle platforms:

- LaneNet is a network that provides lane markings detection in road scenarios [83]
- KittiSeg is a network that performs the segmentation of roads [95]

Therefore, Kittiseg is used to identify the shape of the road within an RGB image and LaneNet is used to identify the lane markings that divide the road into lanes. Upon this, we have developed a post-processing technique based on filtering, clustering and regression (Fig. 6 .b). This post-processing technique makes the whole system far more robust and allows to express the lanes in a simple way (polynomial curves in the vehicle's base frame).

Since the objective is to embed semantic segmentation tools on our vehicle platforms, an emphasis has been put on the related embedded constraints (in particular strong real time constraints and appropriate light hardware such as the NVIDIA Jetson TX2). However, the networks LaneNet and KittiSeg have not been optimized neither for real-time inference nor for inference on light hardware. This is why we had to propose an approach for adapting these networks to our strong embedded constraints. This approach relies on the following three main steps: Reducing the resolution of the input image, Removing all computations not needed at inference (some parts of the networks are only needed in the learning phase), Adapting the network's shape to the hardware.

These optimization steps have been followed for KittiSeg and LaneNet networks. The improvement is obvious. Namely, for the network LaneNet the initial inference needed 334 operations while, after optimization, it needs only 10 operations. The inference initially runs at 0.3Hz on our board NVIDIA Jetson TX2 while, after optimization, it runs at 10Hz. Also the memory needed for inference is divided by two due to the optimization.

7.2.4. Online map-relative localization

Participants: Rabbia Asghar, Mario Garzon, Jerome Lussereau, Christian Laugier.

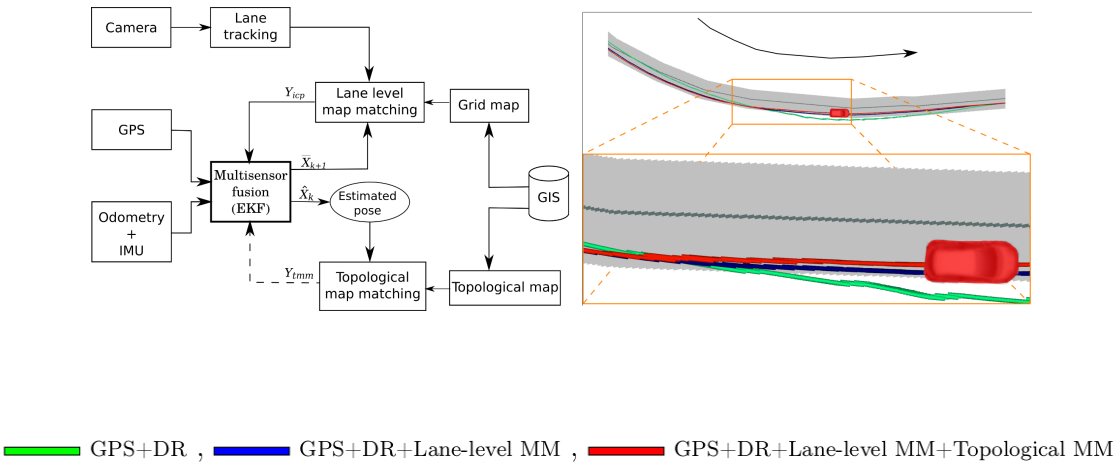


Figure 7. (a) Overview of the map relative localization approach. (b) Estimated pose of the vehicle using three different localization approaches on a curved section of road. The vehicle is provided as a reference where the estimate vehicle pose is just at the curb of the road. Black arrow represents direction of travel.

Localization is one of the key components of the system architecture of autonomous driving and Advanced Driver Assistance Systems (ADAS). Accurate localization is crucial to reliable vehicle navigation and acts as a prerequisite for the planning and control of autonomous vehicles. Offline digital maps are readily available especially in urban scenarios and they play an important role in the field of autonomous vehicles and ADAS. In this framework, we have developed a novel approach for online vehicle localization in a digital map. Two distinct map matching algorithms are proposed:

- Iterative Closest Point (ICP) based lane level map matching (Ll.MM) is performed with visual lane tracker and grid map.
- Decision-rule (DR) based approach is used to perform topological map matching (T. MM).

Results of both map matching algorithms are fused together with GPS and dead reckoning using Extended Kalman Filter to estimate the vehicle's pose relative to the map (see Fig. 7). The approach has been validated on real life conditions on a road-equipped vehicle using a readily available, open source map. Detailed analysis of the experimental results show improved localization using the two aforementioned map matching algorithms (see [50] for more details).
This research work has been carried out in the scope of Project Tornado. A paper on this work was submitted to ICRA2020 and is awaiting review.

7.2.5. System Validation using Simulation and Formal Methods

Participants: Alessandro Renzaglia, Anshul Paigwar, Mathieu Barbier, Philippe Ledent [Chroma/Convecs], Radu Mateescu [Convecs], Christian Laugier, Eduard Baranov [Tamis], Axel Legay [Tamis].

Since 2017, we are working on novel approaches, tools and experimental methodologies with the objective of validating probabilistic perception-based algorithms in the context of autonomous driving. To achieve this goal, a first approach based on Statistical Model Checking (SMC) has been mainly studied in the scope of the European project Enable-S3 and in collaboration with the Inria team Tamis. In this work, we studied the behavior of specifically defined Key Performance Indicators (KPIs), expressed as temporal properties depending on a set of identified metrics, during a large number of simulations via a statistical model checker. As a result, we obtained an evaluation of the probability for the system to meet the KPIs. In particular, we show how this method can be applied to two different subsystems of an autonomous vehicle: a perception system and a decision-making approach for intersection crossing [31]. A more detailed description of the validation scheme for the decision-making approach has been also presented in [49]. This work has been developed in the framework of M. Barbier's PhD thesis, which has been defended in December 2019 [11]. In parallel, in [38], we also proposed a methodology based on a combination of simulation, formal verification, and statistical analysis to validate the collision-risk assessment generated by the Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT), a probabilistic perception system developed in the team. This second work is in collaboration with the Inria team Convecs.

In both cases, the validation methodology relies on the simulation of realistic scenarios generated by using the CARLA simulator [0]. CARLA simulation environment consists of complex urban layouts, buildings and vehicles rendered in high quality, allowing for a realistic representation of real-world scenarios. The ego-vehicle and its sensors, as well as other moving vehicles can be so configured in the simulation to match with the actual system. In order to be able to efficiently generate a large number of execution traces, we have perfected a parameter-based approach which streamlines the process through which the dimensions and initial position and velocity of non-ego vehicles are specified.

We also collected several traces in real experiments by imitating the collision of the ego-vehicle (equipped Renault Zoe) with a pedestrian (by using a mannequin) and with another vehicle (by throwing a big ball). Since it is unfeasible to generate with real experiments a statistically significant number of traces, we focused our analysis on studying how close the simulation traces are to these real experiments by comparing analogous scenarios. These results have been recently submitted to ICRA and are currently under review [0].

7.2.6. Industrial partners and technological transfer

Participants: Christian Laugier, Lukas Rummelhard, Jerome Lussereau, Jean-Alix David, Thomas Genevois.

In 2019, a significant amount of work has been done with the objective to transfer our Bayesian Perception technologies to industrial companies. In a first step, we have developed a new version of CMCDOT based on a clear split of ROS middle-ware code and of GroundEstimator/CMCDOT CUDA code. This allowed us to develop a new version of CMCDOT using the RTMAPS middleware for Toyota Motor Europe. It also allowed us to transfer the CMCDOT technology to some other industrial partners (confidential), in the scope of the project "Security of Autonomous Vehicle" of IRT Nanoelec. Within the IRT Nanoelec framework, we also developed a new "light urban autonomous vehicle" operating using an appropriate version of the CMCDOT and having the capability to navigate with low cost sensors. A first demo of the prototype of this light vehicle has been shown in December 2019, and a start-up project (named Starlink) is currently in incubation.

7.2.7. Autonomous vehicle demonstrations

Participants: Lukas Rummelhard, Jean-Alix David, Thomas Genevois, Jerome Lussereau, Christian Laugier.

In 2019, Chroma has participated to two main public demonstrations:

- **IEEE IV 2019 Conference** (Versailles Satory, June 2019): A one day public demonstration of our Autonomous Vehicle Embedded Perception System has been done using our Renault Zoe platform. Fig. 8 .a and 8 .b show, respectively, the demonstration track (yellow track) and our booth & demonstration vehicle. During the day, we regularly drove people in our Zoe platform for demonstrating how the perception system was working in various situations.

[0] http://carla.org/

[0] A. Paigwar, E. Baranov, A. Renzaglia, C. Laugier and A. Legay, "Probabilistic Collision Risk Estimation for Autonomous Driving: Validation via Statistical Model Checking", *submitted to* IEEE ICRA20.

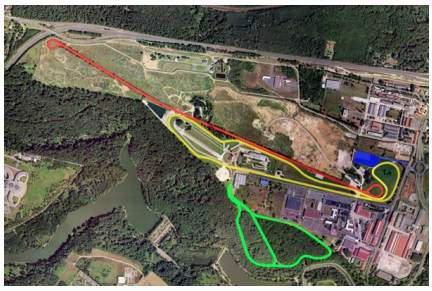


Figure 8. Demonstration at the IV2019 conference : a) track b) demonstration event.

- **FUI Tornado mid-project event** (Rambouillet, September 2019): This one week event included public demonstrations and several open-road tests. During this week, we tested the technologies developed in the scope of the project and we made public and official (for persons from the French Ministries) demonstrations with our Renault Zoe vehicle.

7.3. Situation Awareness & Decision-making for Autonomous Vehicles

Participants: Ozgur Erkent, Christian Wolf, Christian Laugier, Olivier Simonin, Mathieu Barbier, David Sierra-Gonzalez, Jilles Dibangoye, Mario Garzon, Anshul Paigwar, Manuel Alejandro Diaz-Zapata, Victor Romero-Cano [Universidad Autónoma de Occidente, Cali, Colombia], Andrés E. Gómez H., Luiz Serafim-Guardini.

In this section, we include all the novel results in the domains of perception, motion prediction and decision-making for autonomous vehicles. In 2019, these results have also been presented in several invited talks given in some of the major international conferences of the domain [30], [28], [26], [29], [27].

7.3.1. End-to-End Learning of Semantic Grid Estimation Deep Neural Network with Occupancy Grids

Participants: Özgür Erkent, Christian Wolf, Christian Laugier.

Semantic grid is a spatial 2D map of the environment around an autonomous vehicle consisting of cells which represent the semantic information of the corresponding region such as *car, road, vegetation, bikes, etc.*. It consists of an integration of an occupancy grid, which computes the grid states with a Bayesian filter approach, and semantic segmentation information from monocular RGB images, which is obtained with a deep neural network. The network fuses the information and can be trained in an end-to-end manner. The output of the neural network is refined with a conditional random field [15]. The contributions of the study are:

- An end-to-end trainable deep learning method to obtain the semantic grids by integrating the occupancy grids obtained by a Bayesian filter approach and the semantically segmented images by using the monocular RGB images of the environment.
- Grid refinement with conditional random fields (CRFs) on the output of the deep network.
- A comparison of the performances of three different semantic segmentation network architectures in the proposed end-to-end trainable setting.

The proposed method is tested in various datasets (KITTI dataset, Inria-Chroma dataset and SYNTHIA) and different deep neural network architectures are compared (Fig. 9).

7.3.2. Attentional PointNet for 3D object detection in Point Cloud

Participants: Anshul Paigwar, Özgür Erkent, Christian Wolf, Christian Laugier.

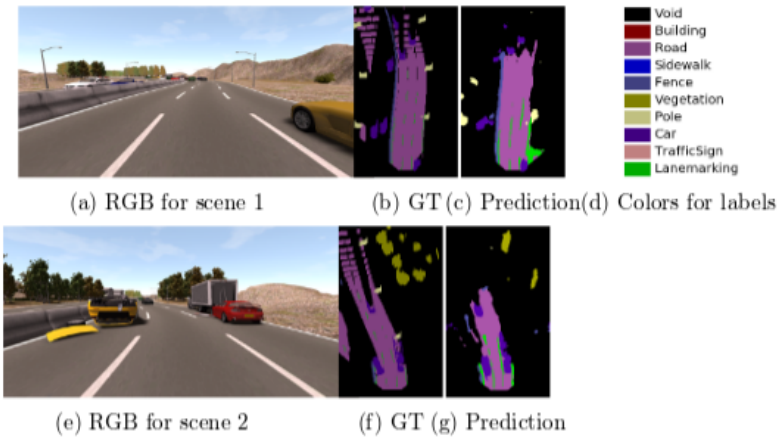


Figure 9. Two scenes with RGB image, ground truth (GT), semantic and segmentation predictions from SYNTHIA dataset.

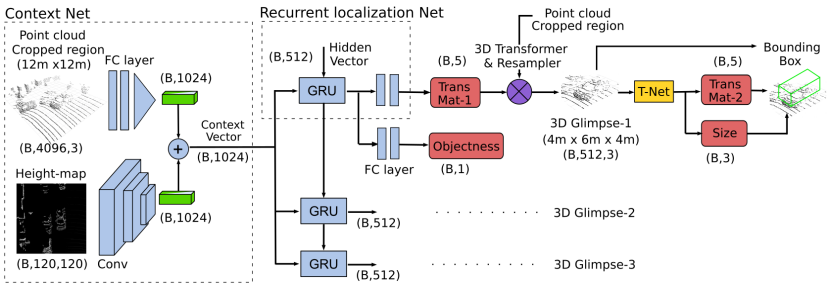


Figure 10. ***Attentional PointNet Architecture****: Given the point cloud and the corresponding height map, network sequentially regress parameters of a 3D Transformation matrix representing pose of a fixed size 3D glimpse. A modified PointNet (T-Net) then estimates another 3D transformation matrix and size representing the 3D bounding box of the object inside the glimpse. Where B is the batch size.*

Accurate detection of objects in 3D point clouds is a central problem for autonomous navigation. Approaches like PointNet [87] that directly operate on sparse point data have shown good accuracy in the classification of single 3D objects. However, LiDAR sensors on Autonomous Vehicles generate a large scale point cloud. Real-time object detection in such a cluttered environment still remains a challenge. In this study, we propose Attentional PointNet, which is a novel end-to-end trainable deep architecture for object detection in point clouds (Fig. 10). We extend the theory of visual attention mechanisms to 3D point clouds and introduce a new recurrent 3D Localization Network module. Rather than processing the whole point cloud, the network learns where to look (finding regions of interest), which significantly reduces the number of points to be processed and inference time. Evaluation on KITTI [72] car detection benchmark shows that our Attentional PointNet achieves comparable results with the *state-of-the-art* LiDAR-based 3D detection methods in detection (Fig. 11) and speed.

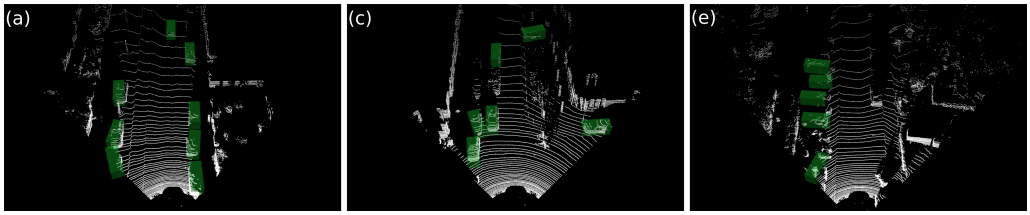


*Figure 11. Visualizations of **Attentional PointNet** results on KITTI dataset for the car category shows model's ability to detect multiple objects in cluttered environments*

This work has been published in CVPR 2019 - Workshop for Autonomous Driving, Long Beach, California, USA [39].

7.3.3. Panoptic Segmentation

Participants: Manuel Alejandro Diaz-Zapata, Victor Romero-Cano [Universidad Autónoma de Occidente, Cali, Colombia], Özgür Erkent, Christian Laugier.

This work has been accomplished during the internship of Manuel Alejandro Diaz Zapata at Inria-Rhone Alpes under supervision of Ozgur Erkent, Victor Romero-Cano and Christian Laugier at Chroma Project Team. Manuel Alejandro Diaz Zapata was a student of Mechatronic Engineering at Universidad Autónoma de Occidente, Colombia during his internship [52].

Semantic segmentation labels an image at the pixel level, where amorphous regions of similar texture or material such as grass, sky or road are given a label depending on the class. Instance segmentation focuses on countable objects such as people, cars or animals by delimiting them in the image using bounding boxes or a segmentation mask. To reduce the gap between the methods used to detect uncountable objects, and things or countable objects, panoptic segmentation has been proposed [75].

We propose a model consisting of three modules: the semantic segmentation module, the instance segmentation module and the panoptic head (Fig.12). Here the semantic segmentation is done by the MobileNetV2 [90] and the instance segmentation is done by Mask R-CNN [73]. The outputs of both networks are joint by the Panoptic Head. The results are provided on two different datasets.

7.3.4. Recognition of dynamic objects for risk assessment

Participants: Andrés E. Gómez H., Özgür Erkent, Christian Laugier.

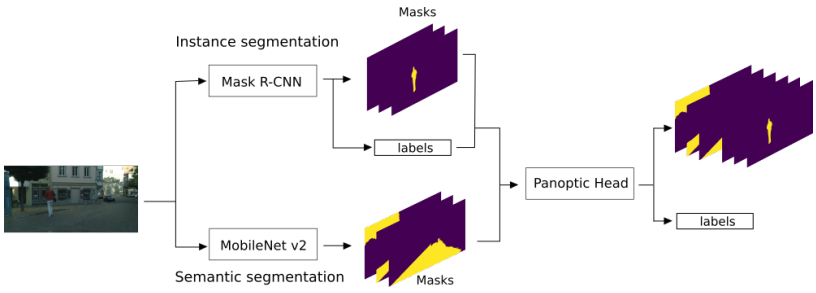


Figure 12. Proposed model for panoptic segmentation.

The Conditional Monte Carlo Dense Occupancy Tracker (*CMCDOT*) framework has proved its accuracy in describing *2D* spatial maps for the Zoe platform. However, this method nowadays cannot recognize the objects in the surrounding. Specifically, the identification of dynamical objects will let us consider different methodologies of risk assessment. This procedure can be possible, through the fusion of RGB and dynamical occupancy grids information.

In the fusion process development, we took into consideration the following steps: *i)* selection of a deep-learning approach, *ii)* development of the projective transformations and *iii)* joining the sub-results. In each step, we used real data from the Zoe platform. In the first step, the *YoloV3* was the deep-learning approach chosen for its accuracy and time performance. In the second step, the projective transformations let us compute the representation of the dynamical points obtained from the occupancy grid plane (i.e., *CMCDOT* framework) in the image plane. Finally, in the third step, we compare the result obtained between the last two-step to identify the dynamic objects around the Zoe platform.

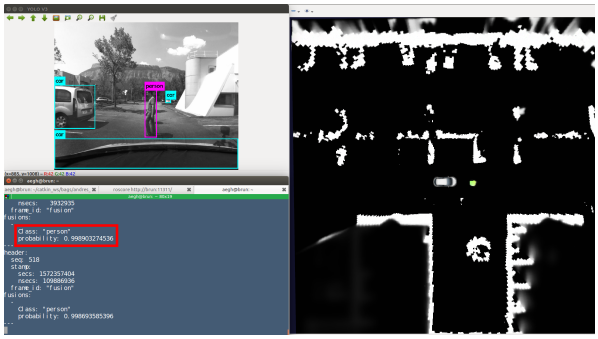


Figure 13. Identification of a pedestrian moving in front of the Zoe platform using the fusion process proposed.

Figure 13 lets us observe the inputs needed for the fusion process and its result.

The work described in this section was done during 2019, inside the activities developed for the Star project. The future work in our project aims to consider the velocity and direction of the dynamic points to define and implement risk behavior functions.

7.3.5. Driving behavior assessment and anomaly detection for intelligent vehicles

Participants: Chule Yang [Nanyang Technological University], Alessandro Renzaglia, Anshul Paigwar, Christian Laugier, Danwei Wang [Nanyang Technological University].

Ensuring safety of both traffic participants and passengers is an important challenge for rapidly growing autonomous vehicle technology. To this purpose, intelligent vehicles not only have to drive safe but must be able to safeguard themselves from other abnormally driving vehicles and avoid potential collisions [56]. Anomaly detection is one of the essential abilities in behavior analysis, which can be used to infer the moving intention of other vehicles and provide evidence for collision risk assessment. In this work, we propose a behavior analysis method based on Hidden Markov Model (HMM) to assess the driving behavior of vehicles on the road and detect anomalous moments. The algorithm uses the real-time velocity and position of the surrounding vehicles provided by the Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT) [89] framework. The movement of each vehicle can be classified into several observation states, namely, Approaching, Braking, Lane Changing, and Lane Keeping. Finally, by chaining these observation states using a Markov model, the abnormality of driving behavior can be inferred into Normal, Attention, and Risk. We perform experiments using CARLA simulator environment to simulate abnormal driving behaviors as shown in Fig. 14 , and we provide results showing the successful detection of abnormal situations.

This work has been published in IEEE CIS-RAM 2019, Bangkok [45].

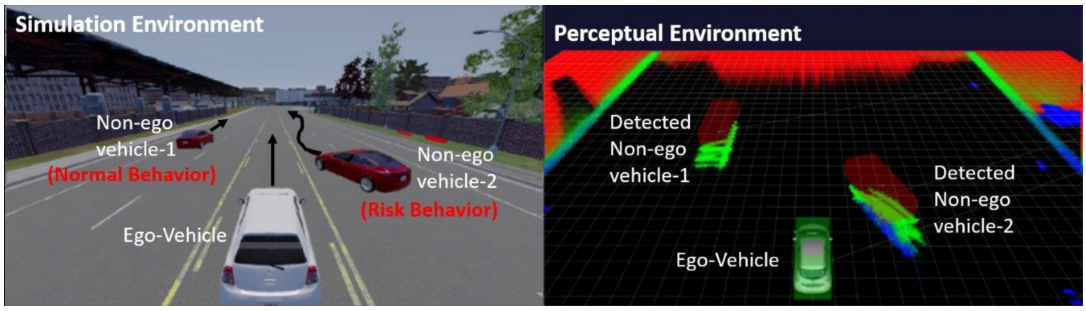


Figure 14. (Left) Simulation environment with CARLA simulator. The white vehicle is the ego-vehicle and two non-ego vehicles are simulated to perform anomaly movements. (Right) Perceptual environment with CMCDOT framework. By analyzing the real-time velocity and position of vehicles, the state and behavior of vehicles can be inferred.

7.3.6. Human-Like Decision-Making for Automated Driving in Highways

Participants: David Sierra-Gonzalez, Mario Garzon, Jilles Dibangoye, Christian Laugier.

Sharing the road with humans constitutes, along with the need for robust perception systems, one of the major challenges holding back the large-scale deployment of automated driving technology. The actions taken by human drivers are determined by a complex set of interdependent factors, which are very hard to model (e.g. intentions, perception, emotions). As a consequence, any prediction of human behavior will always be inherently uncertain, and becomes even more so as the prediction horizon increases. Fully automated vehicles are thus required to make navigation decisions based on the uncertain states and intentions of surrounding vehicles. Building upon previous work, where we showed how to estimate the states and maneuver intentions of surrounding drivers [91], we developed a decision-making system for automated vehicles in highway environments. The task is modeled as a Partially Observable Markov Decision Process and solved in an online fashion using Monte Carlo tree search. At each decision step, a search tree of beliefs is incrementally built and explored in order to find the current best action for the ego-vehicle. The beliefs represent the predicted state of the world as a response to the actions of the ego-vehicle and are updated using an interaction- and

intention-aware probabilistic model. To estimate the long-term consequences of any action, we rely on a lightweight model-based prediction of the scene that assumes risk-averse behavior for all agents. We refer to the proposed decision-making approach as human-like, since it mimics the human abilities of anticipating the intentions of surrounding drivers and of considering the long-term consequences of their actions based on an approximate, common-sense, prediction of the scene. We evaluated the proposed approach in two different simulated navigational tasks: lane change planning and longitudinal control. The results obtained demonstrated the ability of the proposed approach to make foresighted decisions and to leverage the uncertain intention estimations of surrounding drivers.

This work was published in ITSC 2019 [44]. It constitutes the last contribution of the PhD dissertation of David Sierra González, which was defended in April 2019 [12].

7.3.7. *Contextualized Emergency Trajectory Planning using severity curves*

Participants: Luiz Serafim Guardini, Anne Spalanzani, Christian Laugier, Philippe Martinet.

Perception and interpretation of the surroundings is essential for human drivers as well as for (semi-)autonomous vehicles navigation. To improve such interpretation, a lot of effort has been put in place, for example predicting the behavior of pedestrians and other drivers. Nevertheless, to date, cost maps still have considered simple contextualized objects (for instance, binary allowed/forbidden zones or a fixed weight to each type of object). In this work, the risk of injury issued by accidentology is employed to each class of object present in the scene. The scene is analyzed according to dynamic characteristics related to the Ego vehicle and enclosing objects. The aim is to have a better assessment of the surroundings by creating a navigation cost map and to get an improvement on the understanding of the collision severity in the scene. During the first year of his PhD, Luiz Serafim Gaurdini focused on the development of a probabilistic costmap that expresses the Probability of Collision with Injury Risk (PCIR) (see an example on Figure 15). On top of the information gathered by sensors, it includes the severity of injury in the event of a collision between ego and the objects in the scene. This cost map provides enhanced information to perform vehicle motion planning.

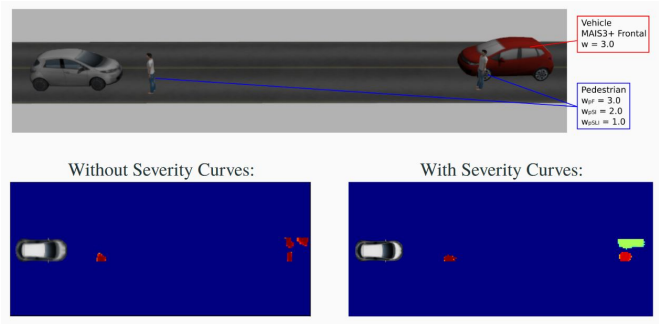


Figure 15. Illustration of the Probabilistic Costmap including the notion of Injury Risk

7.3.8. *Game theoretic decision making for autonomous vehicles' merge manoeuvre in high traffic scenarios*

Participants: Mario Garzon, Anne Spalanzani.

The goal of this work is to provide a solution for a very challenging task: the merge manoeuvre in high traffic scenarios (see Figure 16). Unlike previous approaches, the proposed solution does not rely on vehicle-to-vehicle communication or any specific coordination, moreover, it is capable of anticipating both the actions of other players and their reactions to the autonomous vehicle's movements. The game used is an iterative, multi-player level-k model,which uses cognitive hierarchy reasoning for decision making and has been proved

to correctly model human decisions in uncertain situations. This model uses reinforcement learning to obtain a near-optimal policy, and since it is an iterative model, it is possible to define a goal state so that the policy tries to reach it. To test the decision making process, a kinematic simulation was implemented. The resulting policy was compared with a rule-based approach. The experiments show that the decision making system is capable of correctly performing the merge manoeuvre, by taking actions that require reactions of the other players to be successfully completed. This work was published in [48].

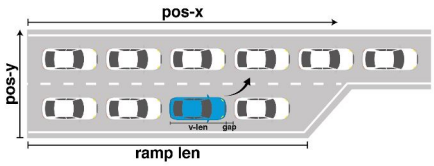


Figure 16. Typical scenario of changing lane in high traffic

7.4. Motion-planning in dense pedestrian environments

We study new motion planning algorithms to allow robots/vehicles to navigate in human populated environment, and to predict human motions. Since 2016, we investigate several directions exploiting vision sensors : prediction of pedestrian behaviors in urban environments (extended GHMM), mapping of human flows (statistical learning), and learning task-based motion planning (RL+Deep-Learning). The works of year 2019 are presented below.

7.4.1. Urban Behavioral Modeling

Participants: Pavan Vasishta, Anne Spalanzani, Dominique Vaufreydaz.

The objective of modeling urban behavior is to predict the trajectories of pedestrians in towns and around cars or platoons (PhD work of P. Vasishta). We first proposed to model pedestrian behaviour in urban scenes by combining the principles of urban planning and the sociological concept of Natural Vision. This model assumes that the environment perceived by pedestrians is composed of multiple potential fields that influence their behaviour. These fields are derived from static scene elements like side-walks, cross-walks, buildings, shops entrances and dynamic obstacles like cars and buses for instance. This work was published in [98]. We then developed an extension to the Growing Hidden Markov Model (GHMM) method that has been proposed to model behavior of pedestrian without observed data or with very few of them. This is achieved by building on existing work using potential cost maps and the principle of Natural Vision. As a consequence, the proposed model is able to predict pedestrian positions more precisely over a longer horizon compared to the state of the art. The method is tested over legal and illegal behavior of pedestrians, having trained the model with sparse observations and partial trajectories. The method, with no training data (see Fig. 17 .a), is compared against a trained state of the art model. It is observed that the proposed method is robust even in new, previously unseen areas. This work was published in [99] and won the **best student paper** of the conference. In 2019, Pavan Vasishta defended his PhD on this topic.

7.4.2. Proactive Navigation for navigating dense human populated environments

Participants: Maria Kabtoul, Anne Spalanzani, Philippe Martinet.

Figure 17. a. Prior Topological Map of the dataset from the Traffic Anomaly Dataset : first figure shows the generated potential cost map and second figure the "Prior Topology" of the image from scene. b. Illustration of the Principle of Proactive Navigation.

Developing autonomous vehicles capable of navigating safely and socially around pedestrians is a major challenge in intelligent transportation. This challenge cannot be met without understanding pedestrians' behavioral response to an autonomous vehicle, and the task of building a clear and quantitative description of the pedestrian to vehicle interaction remains a key milestone in autonomous navigation research. As a step towards safe proactive navigation in a spaceshared with pedestrians, we start to introduce in 2018 a pedestrian-vehicle interaction behavioral model. The model estimates the pedestrian's cooperation with the vehicle in an interaction scenario by a quantitative time-varying function. Using this cooperation estimation the pedestrian's trajectory is predicted by a cooperation-based trajectory planning model (see Figure 17 .b). Both parts of the model are tested and validated using real-life recorded scenarios of pedestrian-vehicle interaction. The model is capable of describing and predicting agents' behaviors when interacting with a vehicle in both lateral and frontal crossing scenarios.

7.4.3. Modelling crowds and autonomous vehicles using Extended Social Force Models

Participants: Manon Predhumeau, Anne Spalanzani, Julie Dugdale.

The focus of this work has been on the realistic simulation of crowds in shared spaces. We have developed a simulator, based on empirical studies and the state of the art, using PED-SIM software. The simulator takes into account the density of crowds, different social group structures in different contexts, inter and intra group forces, and collision avoidance strategies of pedestrians. The Social Force Model (SFM) successfully reproduces many collective phenomena in evacuations or dense crowds. However, pedestrians behaviour is context dependent and the SFM has some limitations when simulating crowds in an open environment under normal conditions. Specifically, in an urban public square pedestrians tend to expand their personal space and try to avoid dense areas to reduce the risk of collision. Based on the SFM, the proposed model splits the perception of pedestrians into a large perception zone and a restricted frontal zone to which they pay more attention. Through their perceptions, the agents estimate the crowd density and dynamically adapt their personal space. Finally, the original social force is tuned to reflect pedestrians preference of avoiding dense areas by turning rather than slowing down as long as there is enough space. Simulation results show that in the considered context the proposed approach produces more realistic behaviours than the original SFM. The simulated crowd is less dense with the same number of pedestrians and less collisions occur, which better fits the observations of sparse crowds in an open place under normal condition [40].

7.4.4. Deep Reinforcement Learning based Vehicle Navigation amongst pedestrians

Participants: Niranjan Deshpande, Anne Spalanzani, Dominique Vaufreydaz.

The objective of this work is to develop a navigation system for an autonomous vehicle in urban environments. The urban environment would consist of other road users as well including other vehicles and pedestrians. Specifically, the focus is on the decision making (behaviour planning) aspect of navigation. In this work, we propose to use Deep Reinforcement Learning as a method to learn decision making. We have developed a Deep Q-Network based agent for decision making amongst pedestrians using the SUMO simulator. This Deep Q-Network based agent is trained for a typical intersection crossing setup amongst pedestrians (see Figure 18). We propose a grid based representation as a state space input to the learning agent. With this grid based representation and our reward function the agent learns a policy capable of driving safely around pedestrians and also follow the traffic rule. This work was published in [35].

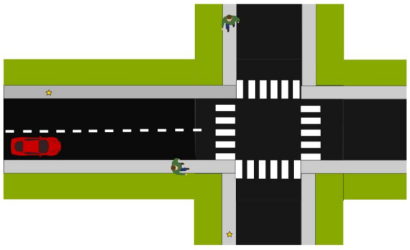


Figure 18. Typical intersection crossing used for training the behavior of the autonomous vehicle

7.5. Learning robot high-level behaviors

7.5.1. Learning task-based motion planning

Participants: Christian Wolf, Jilles Dibangoye, Laetitia Matignon, Olivier Simonin, Edward Beeching.

Our goal is the automatic learning of robot navigation in complex environments based on specific tasks and from visual input. The robot automatically navigates in the environment in order to solve a specific problem, which can be posed explicitly and be encoded in the algorithm (e.g. find all occurences of a given object in the environment, or recognize the current activities of all the actors in this environment) or which can be given in an encoded form as additional input, like text. Addressing these problems requires competences in computer vision, machine learning and AI, and robotics (navigation and paths planning).

A critical part for solving these kind of problems involving autonomous agents is handling memory and planning. An example can be derived from biology, where an animal that is able to store and recall pertinent information about their environment is likely to exceed the performance of an animal whose behavior is purely reactive. Many control problems in partially observed 3D environments involve long term dependencies and planning. Solving these problems requires agents to learn several key capacities: *spatial reasoning* — to explore the environment in an efficient manner and to learn spatio-temporal regularities and affordances. The agent needs to discover relevant objects, store their positions for later use, their possible interactions and the eventual relationships between the objects and the task at hand. Semantic mapping is a key feature in these tasks. A second feature is *discovering semantics from interactions* — while solutions exist for semantic mapping and semantic SLAM [64], [94], a more interesting problem arises when the semantics of objects and their affordances are not supervised, but defined through the task and thus learned from reward.

We started this work in the end of 2017, following the arrival of C. Wolf and his 2 year delegation in the team between Sept 2017. to Sept. 2019, through combinations of reinforcement learning and deep learning. The underlying scientific challenge here is to automatically learn representations which allow the agent to solve multiple sub problems required for the task. In particular, the robot needs to learn a metric representation (a map) of its environment based from a sequence of ego-centric observations. Secondly, to solve the problem,

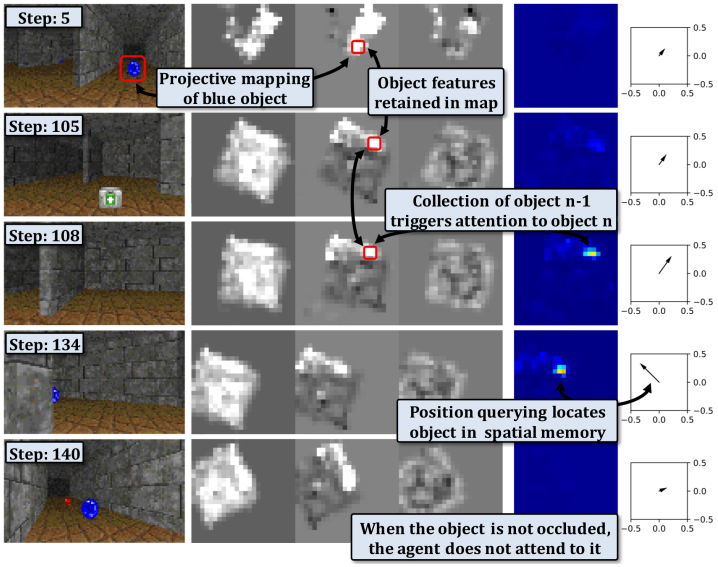


Figure 19. Analysis of the EgoMap for key steps (different rows) during an episode. Left column - RGB observations, central column - the three largest PCA components of features mapped in the spatially structured memory, right - attention heat map (result of the query) and x,y query position vector.

it needs to create a representation which encodes the history of ego-centric observations which are relevant to the recognition problem. Both representations need to be connected, in order for the robot to learn to navigate to solve the problem. Learning these representations from limited information is a challenging goal. This is the subject of the PhD thesis of Edward Beeching, which started on October 2018.

First work proposed a new 3D benchmark for Reinforcement learning, which requires high-level reasoning through the automatic discovery of object affordances [58]. Follow-up work proposed EgoMap, a spatially structured metric neural memory architecture integrating projective geometry in deep reinforcement learning, which we show to outperform classical recurrent baselines. In particular, we show that through visualizations that the agents learns to map relevant objects in its spatial memory without any supervision purely from reward (see Fig. 19). Ongoing work aims to propose a fully differentiable topological memory for Deep-RL.

Creating agents capable of high-level reasoning based on structured memory is main topic of the AI Chair "REMEMBER" obtained by C.Wolf in late 2019 and which involves O. Simonin and J. Dibangoye (Inria Chroma) as well as Laetitia Matignon (LIRIS/Univ Lyon 1). The chair is co-financed by ANR, Naver Labs Europe and INSA-Lyon.

7.5.2. Social robot : NAMO extension and RoboCup@home competition

Participants: Jacques Saraydaryan, Fabrice Jumel, Olivier Simonin, Benoit Renault, Laetitia Matignon, Christian Wolf.

Since 3 years, we investigate robot/humanoid navigation and complex tasks in populated environments such as homes :

- In 2018 we started to study NAMO problems (Navigation Among Movable Obstacles). In his PhD work, Benoit Renault is extending NAMO to Social-NAMO by modeling obstacle hindrance in regards to space access. Defining new spatial cost functions, we extend NAMO algorithms with the ability to maintain area accesses (connectivity) for humans and robots [41]. We also developed a simulator of NAMO problems and algorithms, called S-NAMO-SIM.

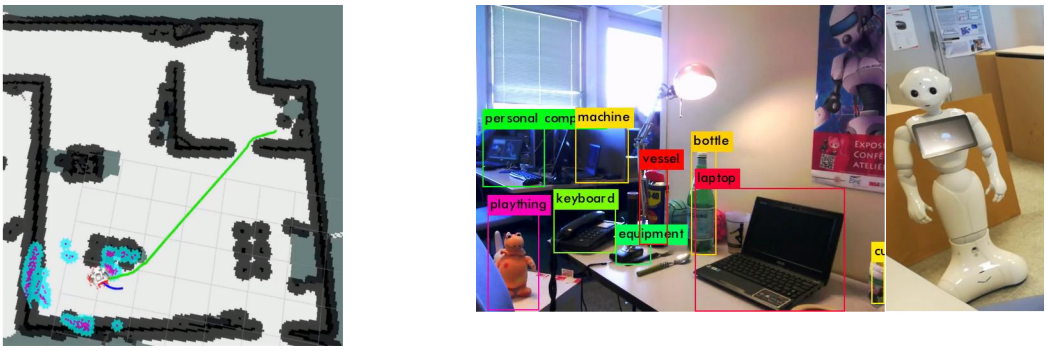


Figure 20. (a) Pepper's navigation and mapping (b) Object detection with Pepper based on vision/deep learning techniques.

- In the context of the **RoboCup** international competition, we created in 2017 the'LyonTech' team, gathering members from Chroma (INSA/CPE/UCBL). We investigated several issues to make humanoid robots able to evolve in a populated indoor environment : decision making and navigation (Fig. 20 .a), human and object recognition based on deep learning techniques (Fig. 20 .b) and human-robot interaction. In July 2018, we participated for the first time to the RoboCup and we reached the 5th place of the SSL league (Robocup@home with Pepper). In July 2019, we participated to the RoboCup organized in Sydney and we obtained the 3rd place of the SSL league. We also awarded the scientific Best Paper of the RoboCup conference [43].

7.6. Sequential decision-making

This research is the follow up of a subgroup led by Jilles S. Dibangoye carried out during the last four years, which include foundations of sequential decision making by a group of cooperative or competitive robots or more generally artificial agents. To this end, we explore combinatorial, convex optimization and reinforcement learning methods.

7.6.1. Optimally solving zero-sum games using centralized planning for decentralized control theory

Participants: Jilles S. Dibangoye, Olivier Buffet [Inria Nancy], Vincent Thomas [Inria Nancy], Abdallah Saffidine [Univ. New South Whales], Christopher Amato [Univ. New Hampshire], François Charpillet [Inria Nancy, Larsen team].

During the last two years, we investigated deep and standard reinforcement learning for solving systems with multiple agents and different information structures. Our preliminary results include:

1. (Theoretical) – As an extension of [68] in the competitive cases, we characterize the optimal solution of two-player fully and partially observable stochastic games.
2. (Theoretical) – We further exhibit new underlying structures of the optimal solution for both non-cooperative two-player settings with information asymmetry, one agent sees what the other does and sees.
3. (Algorithmic) – We extend a non-trivial procedure for computing such optimal solutions.

This work aims at reinforcing a recent theory and algorithms to optimally solving a two-person zero-sum POSGs (zs-POSGs). That is, a general framework for modeling and solving two-person zero-sum games (zs-Games) with imperfect information. Our theory builds upon a proof that the original problem is reducible to a zs-Game—but now with perfect information. In this form, we show that the dynamic programming theory applies. In particular, we extended Bellman equations [59] for zs-POSGs, and coined them maximin (resp. minimax) equations. Even more importantly, we demonstrated Von Neumann & Morgenstern's minimax theorem [102] [103] holds in zs-POSGs. We further proved that value functions—solutions of maximin (resp. minimax) equations—yield special structures. More specifically, the optimal value functions are Lipchitz-continuous. Together these findings allow us to extend planning techniques from simpler settings to zs-POSGs. To cope with high-dimensional settings, we also investigated low-dimensional (possibly non-convex) representations of the approximations of the optimal value function. In that direction, we extended algorithms that apply for convex value functions to lipschitz value functions.

7.6.2. Learning 3D Navigation Protocols on Touch Interfaces with Cooperative Multi-Agent Reinforcement Learning

Participants: Jilles S. Dibangoye, Christian Wolf [INSA Lyon], Quentin Debard [INSA Lyon], Stephane Canu [INSA Rouen].

During the last year, we investigated a number of real-life applications of deep multi-agent reinforcement learning techniques [34]. In particular, we propose to automatically learn a new interaction protocol allowing to map a 2D user input to 3D actions in virtual environments using reinforcement learning (RL). A fundamental problem of RL methods is the vast amount of inter- actions often required, which are difficult to come by when humans are involved. To overcome this limitation, we make use of two collaborative agents. The first agent models the human by learning to perform the 2D finger trajectories. The second agent acts as the interaction protocol, interpreting and translating to 3D operations the 2D finger trajectories from the first agent. We restrict the learned 2D trajectories to be similar to a training set of collected human gestures by first performing state representation learning, prior to reinforcement learning. This state representation learning is addressed by projecting the gestures into a latent space learned by a variational auto encoder (VAE).

7.7. Multi-Robot Routing

7.7.1. Global-local optimization in autonomous multi-vehicle systems

Participants: Guillaume Bono, Jilles Dibangoye, Laetitia Matignon, Olivier Simonin, Florian Peyreron [VOLVO Group, Lyon].

This work is part of the PhD thesis in progress of Guillaume Bono, with the VOLVO Group, in the context of the INSA-VOLVO Chair. The goal of this project is to plan and learn at both global and local levels how to act when facing a vehicle routing problem (VRP). We started with a state-of-the-art paper on vehicle routing problems as it currently stands in the literature [62]. We were surprised to notice that few attention has been devoted to deep reinforcement learning approaches to solving VRP instances. Hence, we investigated our own deep reinforcement learning approach that can help one vehicle to learn how to generalize strategies from solved instances of travelling salesman problems (an instance of VRPs) to unsolved ones.

The difficulty of this problem lies in the fact that its Markov decision process' formulation is intractable, i.e., the number of states grows doubly exponentially with the number of cities to be visited by the salesman. To gain in scalability, we build inspiration on a recent work by DeepMind, which suggests using pointer-net, i.e., a novel deep neural network architecture, to address learning problems in which entries are sequences (here cities to be visited) and output are also sequences (here order in which cities should be visited). Preliminary results are encouraging and we are extending this work to the multi-agent setting.

7.7.2. Towards efficient algorithms for two-echelon vehicle routing problems

Participants: Mohamad Hobballah, Jilles S. Dibangoye, Olivier Simonin, Elie Garcia [VOLVO Group, Lyon], Florian Peyreron [VOLVO Group, Lyon].

During the last year, Mohamad Hobballah (post-doc INSA VOLVO Chair) investigated efficient meta-heuristics for solving two-echelon vehicle routing problems (2E-VRPs) along with realistic logistic constraints. Algorithms for this problem are of interest in many real-world applications. Our short-term application targets goods delivery by a fleet of autonomous vehicles from a depot to the clients through an urban consolidation center using bikers. Preliminary results include:

1. (Methodological) Design of a novel meta-heuristic based on differential evolution algorithm [66] and iterative local search [101]. The former permits us to avoid being attracted by poor local optima whereas the latter performs the local solution improvement.
2. (Empirical) Empirical results on standard benchmarks available at http://www.vrp-rep.org/datasets.html show state-of-the-art performances on most VRP, MDVRP and 2E-VRP instances.

7.7.3. Multi-Robot Routing (MRR) for evolving missions

Participants: Mihai Popescu, Olivier Simonin, Anne Spalanzani, Fabrice Valois [INSA/Inria, Agora team].

After considering Multi-Robot Patrolling of known targets [86], we generalized to MRR (multi-robot routing) and to DMRR (Dynamic MRR) in the work of the PhD of M. Popescu. Target allocation problems have been frequently treated in contexts such as multi-robot rescue operations, exploration, or patrolling, being often formalized as multi-robot routing problems. There are few works addressing dynamic target allocation, such as allocation of previously unknown targets. We recently developed different solutions to variants of this problem :

- MRR-Sat : Multi-robot routing decentralized solutions consist in auction-based methods. Our work addresses the MRR problem and proposes MRR with saturation constraints (MRR-Sat), where the cost of each robot treating its allocated targets cannot exceed a bound (called saturation). We provided a NP-Complete proof for the problem of MRR-Sat. Then, we proposed a new auction-based algorithm for MRR-Sat and MRR, which combines ideas of parallel allocations with target-oriented heuristics. An empirical analysis of the experimental results shows that the proposed algorithm outperforms state-of-the art methods, obtaining not only better team costs, but also a much lower running time. Results are under review.
- DMRR : we defined the Dynamic-MRR problem as the continuous adaptation of the ongoing robot missions to new targets. We proposed a framework for dynamically adapting the existent robot missions to new discovered targets. Dynamic saturation-based auctioning (DSAT) is proposed for adapting the execution of robots to the new targets. Comparison was made with algorithms ranging from greedy to auction-based methods with provable sub-optimality. The results for DSAT shows it outperforms state-of-the-art methods.
- Synchronization : When patrolling targets along bounded cycles, robots have to meet periodically to exchange information, data (e.g. results of their tasks). Data will finally reach a delivery point. Hence, patrolling cycles sometimes have common points (rendezvous points), where the information needs to be exchanged between different cycles (robots). We investigated this problem by defining the following first solutions : random-wait, speed adaptation (first-multiple), primality of periods, greedy interval overlapping. In the context of the PHC 'DRONEM' project [0] we also developed a flow-based approach to the synchronization problem with the team of Prof. Gabriela Czibula from Babes-Bolyai University in Cluj-Napoca, Romania, see [37].

7.8. Multi-UAV exploration and communication

7.8.1. Multi-UAV Exploration and Visual Coverage of 3D Environments

Participants: Alessandro Renzaglia, Olivier Simonin, Jilles Dibangoye, Vincent Le Doze.

[0] Hubert Curien Partnership

Figure 21. (a) UAVs Chroma simulator (b) Intel Aero quadrotors platform (c) Crazyflie micro-UAV platform extended with UWB decawave chip.

Multi-robot teams, especially when involving aerial vehicles (UAVs [0]), are extremely efficient systems to help humans in acquiring information on large and complex environments. In these scenarios, two fundamental tasks are static coverage and exploration. In both cases, the robots' goal is to navigate through the environment and cooperate to maximize the observed area, either by finding the optimal static configuration which provides the best global view in the case of the coverage or by maximizing the new observed areas at every step until the environment becomes completely known in the case of the exploration.

Although these tasks are usually considered separately in the literature, we proposed a common framework where both problems are formulated as the maximization of online acquired information via the definition of single-robot optimization functions, which differs only slightly in the two cases to take into account the static and dynamic nature of coverage and exploration respectively [0]. A common derivative-free approach based on a stochastic approximation of these functions and their successive optimization is proposed, resulting in a fast and decentralized solution. The locality of this methodology limits however this solution to have local optimality guarantees and specific additional layers are proposed for the two problems to improve the final performance.

For the exploration problem, this resulted in a novel decentralized approach which alternates gradient-free stochastic optimization and a frontier-based approach [42] (IROS'19), [47]. Our method allows each robot to generate its own trajectory based on the collected data and the local map built integrating the information shared by its teammates. Whenever a local optimum is reached, which corresponds to a location surrounded by already explored areas, the algorithm identifies the closest frontier to get over it and restarts the local optimization. Its low computational cost, the capability to deal with constraints and the decentralized decision-making make it particularly suitable for multi-robot applications in complex 3D environments.

In the case of visual coverage, we studied how suitable initializations for the UAVs' positions can be computed offline based on a partial knowledge on the environment and how they can affect the final performance of the online measurements-based optimization. The main contribution of this work was thus to add another layer, based on the concept of Centroidal Voronoi Tessellation, to the optimization scheme in order to exploit an a priori sparse information on the environment to cover. The resulting method, taking advantages of the complementary properties of geometric and stochastic optimization, significantly improves the result of the uninitialized solution and notably reduces the probability of a far-to-optimal final configuration. Moreover, the number of iterations necessary for the convergence of the on-line algorithm is also reduced [88].

[0] Unmanned Aerial Vehicles

[0] A. Renzaglia, J. Dibangoye, V. Le Doze and O. Simonin, "A Common Optimization Framework for Multi-Robot Exploration and Coverage in 3D Environments," *sumbitted to* Journal of Intelligent & Robotic Systems, *under review*.

Both previous approaches have been tested in realistic simulations based on our extension of Gazebo, called SimuDronesGR (see Fig. 21 .a). The development of this UAVs simulator, which includes realistic models of both the environment and the aerial vehicle's dynamics and sensors, is an important current activity in Chroma. Such a simulator has the fundamental role of allowing for realistic tests to validate the developed algorithms and to better prepare the implementation of these solutions on the robotic platform of the team (Intel Aero quadrotors, Fig. 21 .b) for real experiments.

7.8.2. Communication-based control of swarm of UAVs

Participants: Remy Grunblatt, Olivier Simonin, Isabelle Guerin-Lassous [Inria/Lyon 1 Dante team], Alexandre Bonnefond.

Intel WiFi controllers are used in many common devices, such as laptops, but also in the Intel Aero Ready-to-Fly UAVs (Unmanned Aerial Vehicle). The mobility capabilities of these devices lead to greater dynamics in radio conditions, and therefore introduce a need for a suitable and efficient rate adaptation algorithm. In the context of the PhD of Remy Grunblatt, we have reverse-engineered the Intel rate adaptation mechanism from the source code of the IwlWifi Linux driver, and we have given, in a comprehensive form, the underlying rate adaptation algorithm named Iwl-Mvm-Rs. We have also implemented the Iwl-Mvm-Rs algorithm in the NS-3 simulator. Thanks to this implementation, we can evaluate the performance of Iwl-Mvm-Rs in different scenarios (static and with mobility, with and without fast fading). We also compared the performances of Iwl-Mvm-Rs with the ones of Minstrel-HT and IdealWifi, also implemented in the NS-3 simulator. This work has been published in ACM MSWiM conference (A) [36].

In the end of 2019, we obtained a DGA/Inria AI project, called "Dynaflock", aiming to extend the flocking approach to control swarm of communicating UAVs. Alexandre Bonnefond started a PhD to elaborate dynamic flocking models based on the link quality, which can be measured online.

7.8.3. Ultra-WideBand based localization & control of micro-UAVs fleets

Participants: Stephane d'Alu, Olivier Simonin, Oana Iova [Inria/INSA Agora team], Hervé Rivano [Inria/INSA Agora team].

The literature on autonomous flight of swarm of UAVs in indoor environments shows it requires the use of an external camera-based localization, i.e. a motion capture system. Indoor flying without such an expensive equipment installed in the infrastructure remains a challenge. To tackle this challenge, we investigate the Ultra-WideBand technology which can be embedded on micro UAVs as a way to estimate inter-drone distances (see Fig. 21 .c Crazyflie micro-UAV). In our approach, the distance information is a fundamental building block to perform a self-maintaining formation flight. We defined and experimented a time-of-flight distance computation, using UWB decawave chips. We showed a Crazyflie flying and computing its position in function of three fixed anchors. We also tested a two-UAV flight where inter-distance is measured to avoid collisions. See first results in [33].

DEFROST Project-Team

7. New Results

7.1. Soft robots locomotion and manipulation control using FEM simulation and quadratic programming

In this work, we proposed a method to control the motion of soft robots able to manipulate objects or roll from one place to another. We used the Finite Element Method (FEM) to simulate the deformations of the soft robot, its actuators, and its environment. To find the inverse model of the robot interacting with obstacles, and with constraints on its actuators, we wrote the problem as a quadratic program with complementarity constraints. The novelty of this work was that friction contacts (sticking contact only) is taken into account in the optimization process, allowing the control of these specific tasks that are locomotion and manipulation. We proposed a formulation that simplifies the optimization problem, together with a dedicated solver [22]. The algorithm had real-time performance and handles evolving environments as long as we know them. To show the effectiveness of the method, we presented several numerical examples, and a demonstration on a real robot (see Figure 2 and 3).

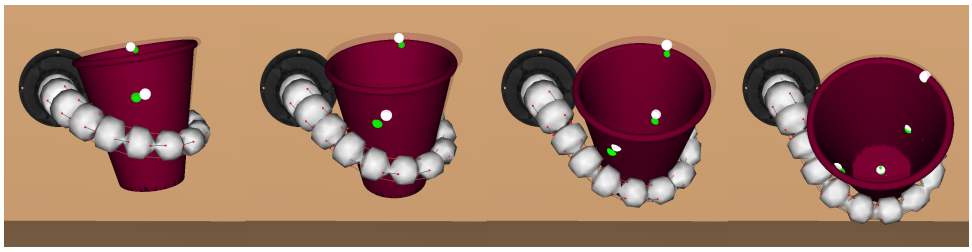


Figure 2. Simulation of a soft gripper holding a deformable cup subject to gravity. Here we optimize the cables displacements to control the position/orientation of the cup. A phantom of the cup target is shown in transparency. The cup have four controlled points represented by the green spheres. The corresponding targets are represented by the white spheres.

Figure 3. Real soft robot actuated online using the output of the simulation. In this scenario, using our control framework, we are able to control the orientation of the real plastic cup (see the attached video).

7.2. Toward Shape Optimization of Soft Robots

This year, we obtained new results on shape optimization for soft robotics where the shape is optimized for a given soft robot usage. To obtain a parametric optimization with a reduced number of parameters, we relied on an approach where the designer progressively refines the parameter space and the fitness function until a satisfactory design is obtained. In our approach, we automatically generate FEM simulations of the soft robot and its environment to evaluate a fitness function while checking the consistency of the solution. Finally, we have coupled our framework to an evolutionary optimization algorithm, and demonstrated its use for optimizing the design of a deformable leg of a locomotive robot. A paper presenting the approach was accepted at IEEE/International Conference On Soft-Robotics2019 [29].

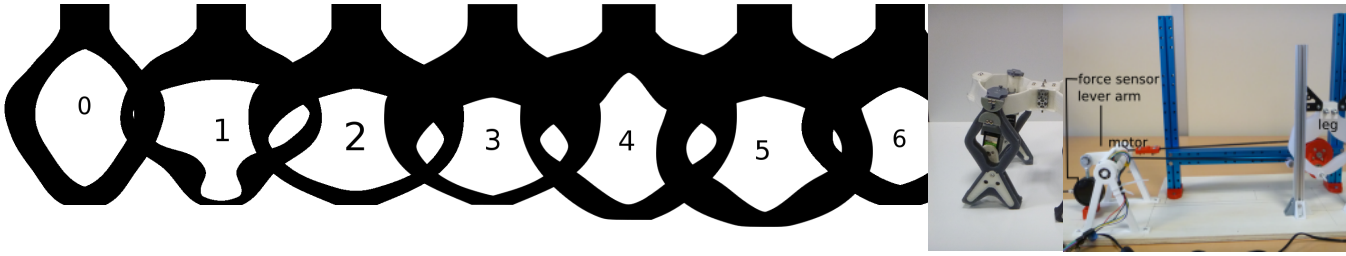


Figure 4. Initial shape (labelled 0) for the optimization and different shapes (labelled 1 to 6) obtained by numerical experiments. Pictures represents the robot with the legs and the test best that has been used to verify the accuracy of the simulation

7.3. Modeling Novel Soft Mechanosensors based on Air-Flow Measurements

In this work, we introduce a new pneumatic mechanosensor dedicated to Soft Robotics and propose a generic method to reconstruct the magnitude of a contact-force acting on it. This is illustrated by Fig. 5 . Changes in cavity volumes inside a soft silicon pad are measured by air-flow sensors. The resulting mechanosensor is characterized by its high sensitivity, repeatability, dynamic range and accurate localization capability in 2D. Using a regression found by machine learning techniques we can predict the contact location and force magnitude accurately when the force magnitudes are within the range of the training data. To be able to provide a more general model, a novel approach based on a Finite Element Method (FEM) is introduced. We formulate an optimization problem, which yields the contact load that best explains the observed changes in cavity volumes. This method makes no assumptions on the force range, the shape of the soft pad or the shape of its cavities. The prediction of the force also results in a model for the deformation of the soft pad. We characterize our sensor and evaluate two designs, a soft pad and a kidney-shaped sensor, in different scenarios. A paper was accepted for the journal *Robotics and Automation Letters (RA-L)* [5]. Furthermore, an extended abstract was accepted at the *RoboTac 2019 Workshop* at *IROS 2019*, leading to a presentation of a demo of the proposed technology.

7.4. Calibration and External Force Sensing for Soft Robots using an RGB-D Camera

Benefiting from the deformability of soft robots, calibration and force sensing for soft robots are possible using an external vision-based system, instead of embedded mechatronic force sensors. In this work, we first propose a calibration method to calibrate both the sensor-robot coordinate system and the actuator inputs. This task is addressed through a sequential optimization problem for both variables. We also introduce an external force

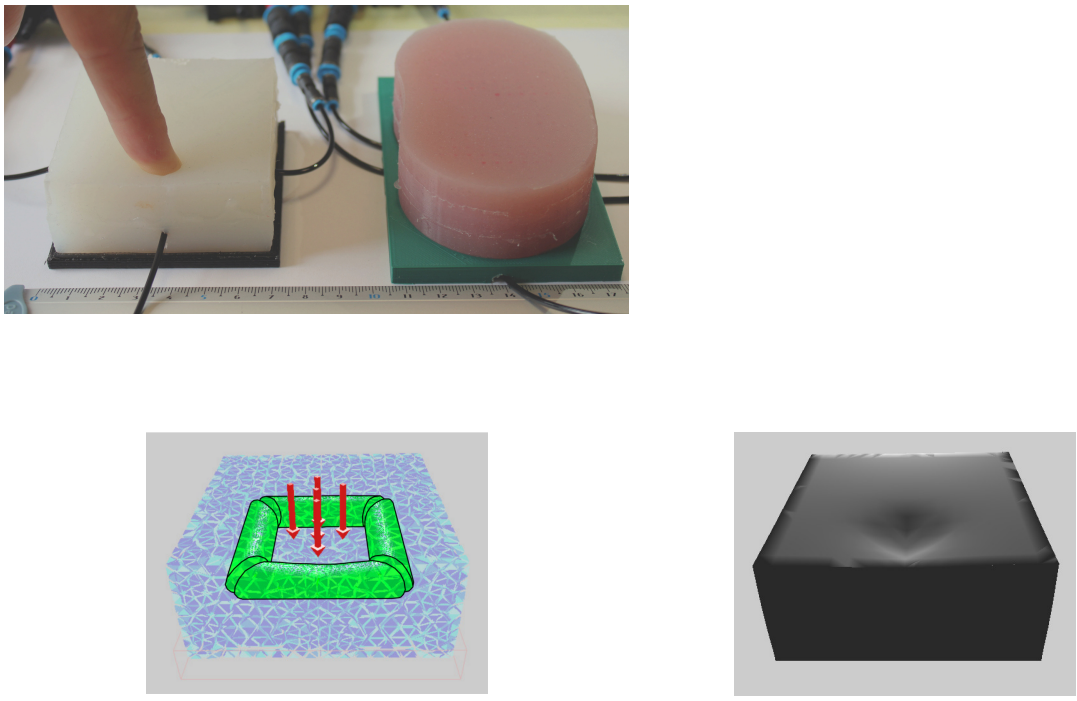


Figure 5. In this work, we show two designs of a novel soft mechanosensor made out of silicone (top, a soft pad and a kidney). When an external force is applied, the volume of cavities embedded in the silicone changes (left). This change in volume is registered through air-flow sensors. Using machine learning and FEM-based techniques, we show that it is possible to estimate the location and magnitude of an external force on the mechanosensor. Using the FEM also yields an estimation of the deformation of the sensor (left and right).

sensing system based on a real-time Finite Element (FE) model with the assumption of static configurations, and which consists of two steps: force location detection and force intensity computation. The algorithm that estimates force location relies on the segmentation of the point cloud acquired by an RGB-D camera. Then, the force intensities can be computed by solving an inverse quasi-static problem based on matching the FE model with the point cloud of the soft robot. As for validation, the proposed strategies for calibration and force sensing have been tested using a parallel soft robot driven by four cables (see figure 6 and reference [18]).

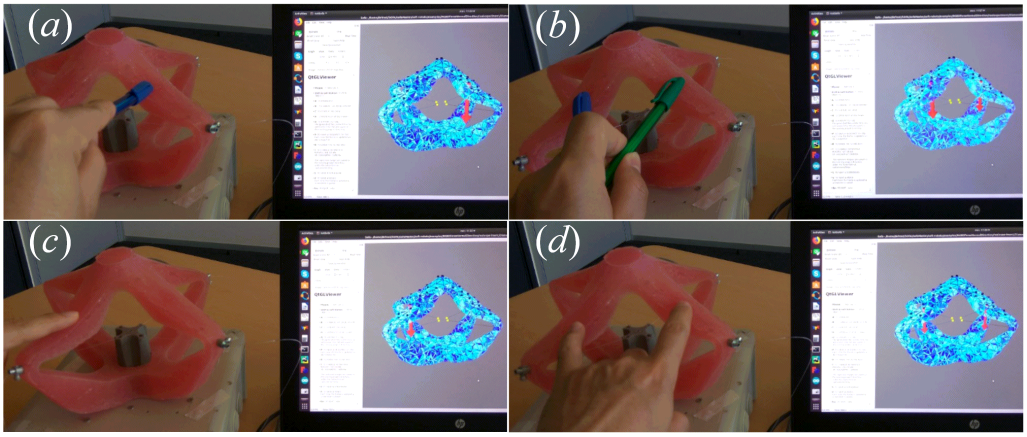


Figure 6. Screenshot of external force sensing. The robot has four cables with constant length for the experiment. (a) and (c) show one external force on the actuated soft robot. (b) and (d) show case with two external forces.

7.5. Motion Control of Cable-Driven Continuum Catheter Robot through Contacts

Catheter-based intervention plays an important role in minimally invasive surgery. For the closed-loop control of catheter robot through contacts, the loss of contact sensing along the entire catheter might result in task failure. To deal with this problem, we propose a decoupled motion control strategy which allows to control insertion and bending independently. We model the catheter robot and the contacts using the Finite Element Method. Then, we combine the simulated system and the real system for the closed-loop motion control. The control inputs are computed by solving a quadratic programming (QP) problem with a linear complementarity problem (LCP). A simplified method is proposed to solve this optimization problem by converting it into a standard QP problem. Using the proposed strategy, not only the control inputs but also the contact forces along the entire catheter can be computed without using force sensors. Finally, we validate the proposed methods using both simulation and experiments on a cable-driven continuum catheter robot for the real-time motion control through contacts [17].

7.6. Control Design for Soft Robots based on Reduced Order Model

Inspired by nature, soft robots promise disruptive advances in robotics. Soft robots are naturally compliant and exhibit nonlinear behavior, which makes their study challenging. No unified framework exists to control these robots, especially when considering their dynamics. This work proposes a methodology to study this type of robots around a stable equilibrium point. It can make the robot converge faster and with reduced oscillations to a desired equilibrium state. Using computational mechanics, a large-scale dynamic model of the robot is obtained and model reduction algorithms enable the design of low order controller and observer. A real robot is used to demonstrate the interest of the results [14].

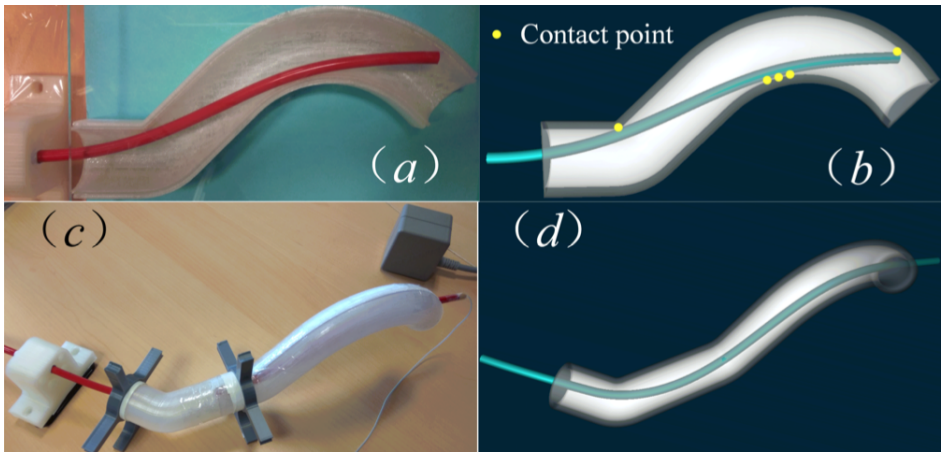


Figure 7. (a) and (c) present experimental setups for the validation. (b) and (d) show that we take into account the contact and provide accurate results in the simulation

FLOWERS Project-Team

7. New Results

7.1. Curiosity-Driven Learning in Humans

7.1.1. Computational Models Of Information-Seeking and Curiosity-Driven Learning in Human Adults

Participants: Pierre-Yves Oudeyer [correspondant], Sébastien Forestier, Alexandr Ten.

This project involves a collaboration between the Flowers team and the Cognitive Neuroscience Lab of J. Gottlieb at Columbia Univ. (NY, US), on the understanding and computational modeling of mechanisms of curiosity, attention and active intrinsically motivated exploration in humans.

It is organized around the study of the hypothesis that subjective meta-cognitive evaluation of information gain (or control gain or learning progress) could generate intrinsic reward in the brain (living or artificial), driving attention and exploration independently from material rewards, and allowing for autonomous lifelong acquisition of open repertoires of skills. The project combines expertise about attention and exploration in the brain and a strong methodological framework for conducting experimentations with monkeys, human adults and children together with computational modeling of curiosity/intrinsic motivation and learning.

Such a collaboration paves the way towards a central objective, which is now a central strategic objective of the Flowers team: designing and conducting experiments in animals and humans informed by computational/mathematical theories of information seeking, and allowing to test the predictions of these computational theories.

7.1.1.1. Context

Curiosity can be understood as a family of mechanisms that evolved to allow agents to maximize their knowledge (or their control) of the useful properties of the world - i.e., the regularities that exist in the world - using active, targeted investigations. In other words, we view curiosity as a decision process that maximizes learning/competence progress (rather than minimizing uncertainty) and assigns value ("interest") to competing tasks based on their epistemic qualities - i.e., their estimated potential allow discovery and learning about the structure of the world.

Because a curiosity-based system acts in conditions of extreme uncertainty (when the distributions of events may be entirely unknown) there is in general no optimal solution to the question of which exploratory action to take [108], [130], [141]. Therefore we hypothesize that, rather than using a single optimization process as it has been the case in most previous theoretical work [86], curiosity is comprised of a family of mechanisms that include simple heuristics related to novelty/surprise and measures of learning progress over longer time scales [128] [56], [118]. These different components are related to the subject's epistemic state (knowledge and beliefs) and may be integrated with fluctuating weights that vary according to the task context. Our aim is to quantitatively characterize this dynamic, multi-dimensional system in a computational framework based on models of intrinsically motivated exploration and learning.

Because of its reliance on epistemic currencies, curiosity is also very likely to be sensitive to individual differences in personality and cognitive functions. Humans show well-documented individual differences in curiosity and exploratory drives [106], [140], and rats show individual variation in learning styles and novelty seeking behaviors [80], but the basis of these differences is not understood. We postulate that an important component of this variation is related to differences in working memory capacity and executive control which, by affecting the encoding and retention of information, will impact the individual's assessment of learning, novelty and surprise and ultimately, the value they place on these factors [133], [149], [50], [155]. To start understanding these relationships, about which nothing is known, we will search for correlations between curiosity and measures of working memory and executive control in the population of children we test in our tasks, analyzed from the point of view of a computational models of the underlying mechanisms.

A final premise guiding our research is that essential elements of curiosity are shared by humans and non-human primates. Human beings have a superior capacity for abstract reasoning and building causal models, which is a prerequisite for sophisticated forms of curiosity such as scientific research. However, if the task is adequately simplified, essential elements of curiosity are also found in monkeys [106], [98] and, with adequate characterization, this species can become a useful model system for understanding the neurophysiological mechanisms.

7.1.1.2. Objectives

Our studies have several highly innovative aspects, both with respect to curiosity and to the traditional research field of each member team.

- Linking curiosity with quantitative theories of learning and decision making: While existing investigations examined curiosity in qualitative, descriptive terms, here we propose a novel approach that integrates quantitative behavioral and neuronal measures with computationally defined theories of learning and decision making.
- Linking curiosity in children and monkeys: While existing investigations examined curiosity in humans, here we propose a novel line of research that coordinates its study in humans and non-human primates. This will address key open questions about differences in curiosity between species, and allow access to its cellular mechanisms.
- Neurophysiology of intrinsic motivation: Whereas virtually all the animal studies of learning and decision making focus on operant tasks (where behavior is shaped by experimenter-determined primary rewards) our studies are among the very first to examine behaviors that are intrinsically motivated by the animals' own learning, beliefs or expectations.
- Neurophysiology of learning and attention: While multiple experiments have explored the single-neuron basis of visual attention in monkeys, all of these studies focused on vision and eye movement control. Our studies are the first to examine the links between attention and learning, which are recognized in psychophysical studies but have been neglected in physiological investigations.
- Computer science: biological basis for artificial exploration: While computer science has proposed and tested many algorithms that can guide intrinsically motivated exploration, our studies are the first to test the biological plausibility of these algorithms.
- Developmental psychology: linking curiosity with development: While it has long been appreciated that children learn selectively from some sources but not others, there has been no systematic investigation of the factors that engender curiosity, or how they depend on cognitive traits.

7.1.1.3. Current results: experiments in Active Categorization

In 2018, we have been occupied by analyzing data of the human adult experiment conducted in 2017. In this experiment we asked whether humans possess, and use, metacognitive abilities to guide task choices in two contexts motivational contexts, in which they could freely choose to learn about 4 competing tasks. Participants (n = 505, recruited via Amazon Mechanical Turk) were asked to play a categorization game with four distinct difficulty levels. Some participants had been explicitly prescribed a goal of maximizing their learning across the difficulty levels (across tasks), while others did not receive any specific instructions regarding the goal of the game. The experiment yielded a rich but complex set of data. The data includes records of participants' classification responses, task choices, reaction times, and post-task self-reports about various subjective evaluations of the competing tasks (e.g. subjective interest, progress, learning potential, etc.). We are now finalizing the results and a computational model of the underlying cognitive and motivational mechanisms in order to prepare them for public dissemination.

The central question going into the study was, how do active learners become interested in specific learning exercises: how do they decide which task to engage with, when none of the tasks provide external rewards. Last year, we identified some of the key behavioral observations that merited further attention. First, we saw a clear effect of an external goal prescription on the participants' overall task selection strategy. People who were explicitly instructed to try to maximize their learning across the 4 tasks challenged themselves more by giving preference towards harder tasks. In contrast, those who were simply familiarized with the rules of the game

and not given any explicit suggestions from the experiments did not show this overchallenge bias and had a slight preference for easier tasks (see figure 5). Second, we observed that although strategies varied between the two instruction groups, there was some considerable within-group variability in learning. We found that in both groups, people had varying success in learning the classification task for each task family resulting in four distinct performance based groups: learners of 0, 1, 2, or 3 tasks (task 4 was unlearnable), as shown in figure 6 . Importantly, successful 3-task learners in both instruction groups exhibited similar task preferences, suggesting that (1) even in the absence of external instruction, people can be motivated to explore the task space and (2) intrinsically motivated exploration is similar to strategies employed when a learner is trying to maximize her learning.

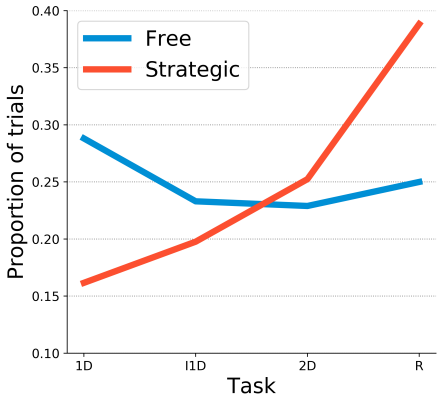


Figure 5. Proportion of trials on each task (1D, I1D, 2D, and R). The group with an externally prescribed learning maximization goal is referred to as "Strategic", and the unconstrained group is referred to as "Free". 1D was the task were categorization was determined by a single variable dimension. In I1D (ignore 1D), the stimuli varied across 2 dimensions, but only one determined the stimulus category. In 2D, there were 2 variable dimensions and both jointly determined the category. Finally in R, there were 2 variable dimensions, but none of them could reliably predict the stimulus class. The top plot shows data aggregated across experimental groups, shown separately in the bottom plot. In hte figure, task difficulty increases from left to right.

Assuming that task choice decisions are based on a subjective evaluative process that assigns value to choice candidates, we considered a simple choice model of task selection. In a classic conditional logit model [116], choices are made probabilistically and the choice probabilities are proportional to choice utilities (the inherent subjective value of a choice; also see ref [159]). We elaborated on the utility component of the basic choice model to consider two utility aspects of interest: a relative measure of learning progress (LP) and an absolute measure of proportion correct (PC). Although both measures are based on empirical feedback (correct / incorrect), the LP measure is considered relative, because it captures how performance changes over time by comparing performance estimates across different time scales, while PC is absolute in a sense that it only characterizes performance at a given instance. While PC alone does not differentiate between an unfamiliar (but potentially easy) task on which the performance might be low and a familiar but very hard task, the former can have markedly LP (due to the gradual improvement on that task) than the latter. Only the tasks characterized by high LP are then worthy of time and effort if the goal is to master tasks. The utility component in our model thus includes two principal quantities:

$$u_{i,t} = \alpha LP_{i,t} + \beta PC_{i,t}$$

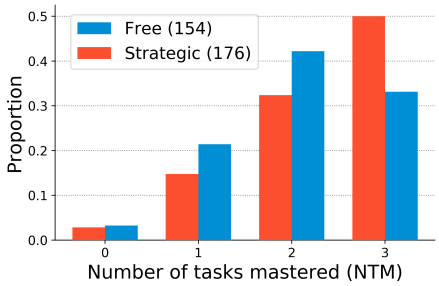


Figure 6. Learning outcomes by group. The group with an externally prescribed learning maximization goal is referred to as "Strategic", and the unconstrained group is referred to as "Free". Number of people in the groups are given in parentheses. The figure shows that some people seem to set and pursue learning maximization goals in the absence of external rewards and instructions. Additionally, learning goals have to be matched with appropriate behavioral strategies in order to be reached, which explains the failure to learn all 3 tasks by some people in the Strategic group.

where α and β are free parameters indexing the model's sensitivity to LP and PC, respectively. Index i designates the task, while t indexes time. Thus, in our model, utility is seen as a dual-component linear computation of both relative and absolute competence quantities. Task utilities enter the decision-making process that assigns relative preference to each task:

$$p(task_i) = \frac{e^{u_i/\tau}}{\sum_j e^{u_j/\tau}}$$

where τ is another free parameter (known as temperature) that controls the stochasticity of utility-based decision. The sum over j elements, $\sum_j$ constitutes the total exponentiated utility of each task in the task space, thus normalizing each individual exponentiated task utility u_i.

The computation of the utility components is important for the model, because it ultimately determines how well the model can fit to choice data. We started exploring the model with a simple definition for both LP and PC. Both components are based on averaging binary feedback over multiple trials in the past. Since the familiarization stage of our experiment was 15-trials-long, the first free choice was made based on feedback data from 15 trials on each task. Accordingly, we defined PC to be the proportion of correct guesses in 15 trials. LP was defined as the absolute difference between the first 9 and the last 6 trials of the recent most 15 trials in the past. While a participant was engaged with one of the tasks, LP and PC for that task changed according to her dynamic performance, while LPs and PCs for other tasks remained unchanged. We acknowledge that there are probably multiple other components at play when it comes to utility computation, some of which may have little to do with task competence. We also submit that there are multiple ways of defining the PC and LP components that are more biologically rooted and plausible given what we know of memory and metacognition. Finally, we do not rule out the possibility of dynamic changes of free parameters themselves, corresponding to changes in motivation during the learning process. All of these considerations are worthy directions of future research, but in this study we focused on finding some necessary evidence for the sensitivity to learning progress.

We fitted the model to each individual's choice data using maximum likelihood estimation. Assuming that choice probabilities on each trial come from a categorical distribution (also called a generalized Bernoulli distribution), where the probability of choosing item i is given by:

$$P(\mathbf{x} \mid \mathbf{p}) = \prod_{i=1}^{k} p_i^{x_i}$$

where $\mathbf{p}$ is a vector of probabilities associated with k events, and $\mathbf{p}$ is a one-hot encoded vector representing discrete items x_i. We add a time index to indicate the dynamic quality of choice probabilities, so that:

$$P_t(\mathbf{x} \mid \mathbf{p}_t) = \prod_{i=1}^{k} p_{i,t}^{x_i}$$

Then, the likelihood of the choice model ($p(task_i \mid \alpha, \beta, \tau)$) at time t is equal to the product of choice probabilities given by that model for that time step:

$$L_t(\alpha, \beta, \tau \mid \mathbf{x}) = \prod_{i}^{k} p_{i,t}^{x_i}$$

and since the empirical choice data can be represented in a one-hot format, the likelihood of the model for a given time point boils down to the predicted probability of the actualy choice:

$$L_t(\alpha, \beta, \tau \mid choice = i) = p_{i,t} = \frac{e^{u_{i,t}/\tau}}{\sum_j e^{u_{j,t}/\tau}}$$

The likelihood of the model across all m trials is obtained by applying the product rule of probability:

$$L_{overall}(\alpha, \beta, \tau \mid choice = i) = \prod_{t}^{m} p_{i,t} = \frac{e^{u_{i,t}/\tau}}{\sum_j e^{u_{j,t}/\tau}}$$

For convenience, we use the negative log transformation to avoid computational precision problems and convert a likelihood maximization objective into negative likelihood minimization problem solvable by publicly accessible optimization tools:

$$-\log L_{overall}(\alpha, \beta, \tau \mid choice = i) = -\sum_{t}^{m} \log p_{i,t}$$

Having formulated the likelihood function, we optimized the free parameters to obtain a model that fits the individual data best. We thus fit an individual-level model to each participant's choice data. The fitted parameters can be interpreted as relative sensitivity to the competence quantities of interest (LP and PC), since these quantities share the same range of values (0 to 1). Finally, we performed some group-level analyses on these individual-level parameter estimates to evaluate certain group-level effects that might influence them (e.g. effect of instruction or learning proficiency).

The group with a learning maximization goal devalued tasks with higher positive feedback expectation. Qualitatively, this matched our prior observations showing their strong preference for harder tasks. However, the best learners (3-task learners) across both instruction groups showed a slight preference for learning progress and a relatively strong aversion to positive feedback. It appears that what separated better and worse learners among the learning maximizers was whether they followed learning progress, and not just the feedback heuristic. On the other hand, while less successful learners in the unconstrained group seemed to choose tasks according to learning progress, they valued positive feedback over it, which prevented them from exploring more challenging learnable tasks. This is reflected in group mean values of the fitted parameters summarized in table 1

Table 1. Fitted parameter values averaged within number of tasks learned and instruction groups. The group with an externally prescribed learning maximization goal is referred to as "Strategic", and the unconstrained group is referred to as "Free". NTM stands for the number of tasks mastered

Group	NTM	Learning progress	Proportion correct	Temperature
	1	0.27	0.56	6.92
	2	0.07	0.32	5.90
	3	0.12	-0.15	6.14
	1	-0.01	-0.56	7.14
	2	-0.15	-0.41	6.42
	3	0.11	-0.44	6.59

We also looked at the relative importance of the utility model parameters by performing model comparisons. We compared 4 models based on combinations of 2 factors: PC and LP. According to the AIC scores (see Table 2), the best model was the one which included both LP and PC factors, followed by the PC-only model, and then by the LP-only model. The random-choice model came in last with the highest AIC score. The results of this model comparison show that both learning progress and positive feedback expectation factors provide substantive improvement to model likelihood compared to when these factors are included alone, or when neither of them is present. This is potentially important, as it provides some evidence for the role of relative competence kind of measure in autonomous exploration. We are planning to submit the work described about to a high impact peer-reviewed journal focusing on computational modeling of human behavior.

Table 2. Model comparisons.

Model	AIC	$AIC - AIC_{min}$
LP + PC	568.99	-
PC	593.51	24.52
LP	658.46	89.47
Random	693.60	124.61

7.1.2. Experimental study of the role of intrinsic motivation in developmental psychology experiments and in the development of tool use

Participants: Pierre-Yves Oudeyer, Sébastien Forestier [correspondant], Laurianne Rat-Fisher.

Children are so curious to explore their environment that it can be hard to focus their attention on one given activity. Many experiments in developmental psychology evaluate particular skills of children by setting up a task that the child is encouraged to solve. However, children may sometimes be following their own motivation to explore the experimental setup or other things in the environment. We suggest that considering the intrinsic motivations of children in those experiments could help understanding their role in the learning of related skills and on long-term child development. To illustrate this idea, we reanalyzed and reinterpretd a typical experiment aiming to evaluate particular skills in infants. In this experiment run by Lauriane Rat-Fischer et al, 32 21-month old infants have to retrieve a toy stuck inside a tube, by inserting several blocks in sequence into the tube. In order to understand the mechanisms of the motivations of babies, we studied in detail their behaviors, goals and strategies in this experiment. We showed that their motivations are diverse and do not always coincide with the target goal expected and made salient by the experimenter. Intrinsically motivated exploration seems to play an important role in the observed behaviors and to interfere with the measured success rates. This new interpretation provides a motivation for studying curiosity and intrinsic motivations in robotic models.

7.2. Intrinsically Motivated Learning in Artificial Intelligence

7.2.1. Intrinsically Motivated Goal Exploration and Goal-Parameterized Reinforcement Learning

Participants: Sébastien Forestier, Pierre-Yves Oudeyer [correspondant], Olivier Sigaud, Cédric Colas, Adrien Laversanne-Finot, Rémy Portelas, Grgur Kovac.

7.2.1.1. Intrinsically Motivated Exploration of Modular Goal Spaces and the Emergence of Tool use

A major challenge in robotics is to learn goal-parametrized policies to solve multi-task reinforcement learning problems in high-dimensional continuous action and effect spaces. Of particular interest is the acquisition of inverse models which map a space of sensorimotor goals to a space of motor programs that solve them. For example, this could be a robot learning which movements of the arm and hand can push or throw an object in each of several target locations, or which arm movements allow to produce which displacements of several objects potentially interacting with each other, e.g. in the case of tool use. Specifically, acquiring such repertoires of skills through incremental exploration of the environment has been argued to be a key target for life-long developmental learning [54].

Recently, we developed a formal framework called "Intrinsically motivated goal exploration processes" (IMGEPs), enabling study agents that generate and sequence their own goals to learn world models and skill repertoirs, that is both more compact and more general than our previous models [82]. We experimented several implementations of these processes in a complex robotic setup with multiple objects (see Fig. 7), associated to multiple spaces of parameterized reinforcement learning problems, and where the robot can learn how to use certain objects as tools to manipulate other objects. We analyzed how curriculum learning is automated in this unsupervised multi-goal exploration process, and compared the trajectory of exploration and learning of these spaces of problems with the one generated by other mechanisms such as hand-designed learning curriculum, or exploration targeting a single space of problems, and random motor exploration. We showed that learning several spaces of diverse problems can be more efficient for learning complex skills than only trying to directly learn these complex skills. We illustrated the computational efficiency of IMGEPs as these robotic experiments use a simple memory-based low-level policy representations and search algorithm, enabling the whole system to learn online and incrementally on a Raspberry Pi 3.

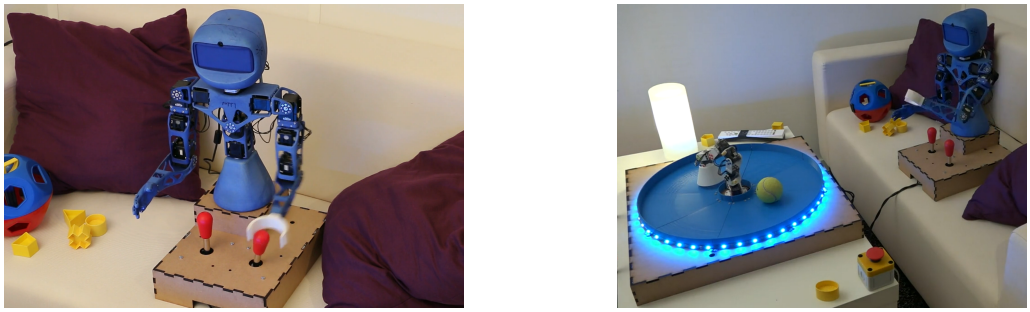


Figure 7. Robotic setup. Left: a Poppy Torso robot (the learning agent) is mounted in front of two joysticks. Right: full setup: a Poppy Ergo robot (seen as a robotic toy) is controlled by the right joystick and can hit a tennis ball in the arena which changes some lights and sounds.

In order to run many systematic scientific experiments in a shorter time, we scaled up this experimental setup to a platform of 6 identical Poppy Torso robots, each of them having the same environment to interact with. Every robot can run a different task with a specific algorithm and parameters each (see Fig. 8). Moreover, each Poppy Torso can also perceives the motion of a second Poppy Ergo robot, than can be used, this time, as a distractor performing random motions to complicate the learning problem. 12 top cameras and 6 head

cameras can dump video streams during experiments, in order to record video datasets. This setup is now used to perform more experiments to compare different variants of curiosity-driven learning algorithms.

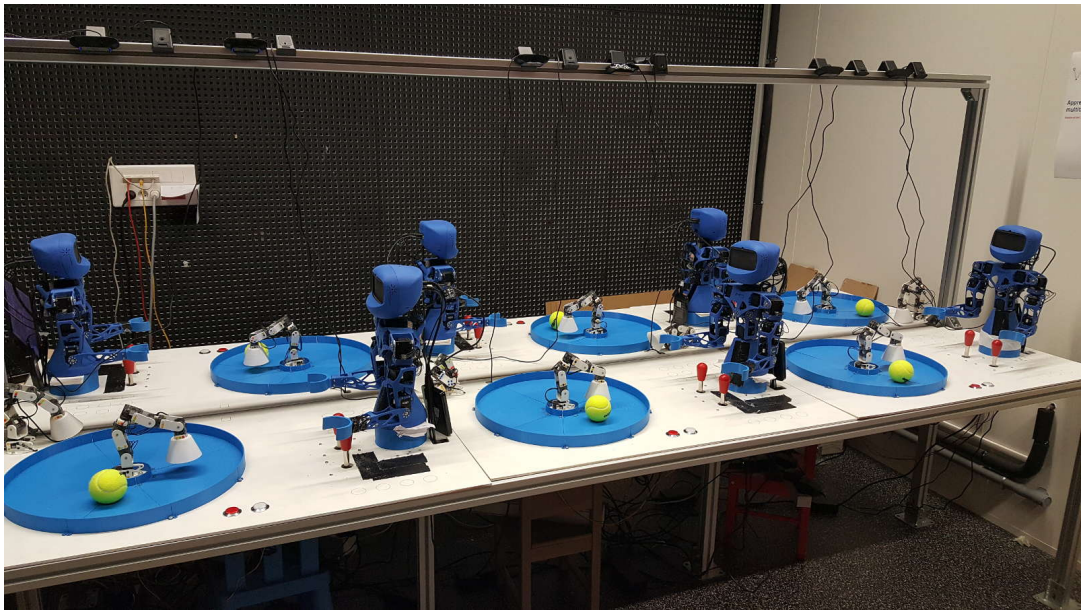


Figure 8. Platform of 6 robots with identical environment: joysticks, Poppy Ergo, ball in an arena, and a distractor. The central bar supports the 12 top cameras.

7.2.1.2. Leveraging the Malmo Minecraft platform to study IMGEP in rich simulations

We continued to leverage the Malmo platform to study curiosity-driven learning applied to multi-goal reinforcement learning tasks (https://github.com/Microsoft/malmo). The first step was to implement an environment called Malmo Mountain Cart (MMC), designed to be well suited to study multi-goal reinforcement learning (see figure [9]). We then showed that IMGEP methods could efficiently explore the MMC environment without any extrinsic rewards. We further showed that, even in the presence of distractors in the goal space, IMGEP methods still managed to discover complex behaviors such as reaching and swinging the cart, especially Active Model Babbling which ignored distractors by monitoring learning progress.

7.2.1.3. Unsupervised Learning of Modular Goal Spaces for Intrinsically Motivated Goal Exploration

Intrinsically motivated goal exploration algorithms enable machines to discover repertoires of policies that produce a diversity of effects in complex environments. These exploration algorithms have been shown to allow real world robots to acquire skills such as tool use in high-dimensional continuous state and action spaces, as shown in previous sections. However, they have so far assumed that self-generated goals are sampled in a specifically engineered feature space, limiting their autonomy. We have proposed an approach using deep representation learning algorithms to learn an adequate goal space. This is a developmental 2-stage approach: first, in a perceptual learning stage, deep learning algorithms use passive raw sensor observations of world changes to learn a corresponding latent space; then goal exploration happens in a second stage by sampling goals in this latent space. We made experiments with a simulated robot arm interacting with an object, and we show that exploration algorithms using such learned representations can closely match, and even sometimes improve, the performance obtained using engineered representations. This work was presented at ICLR 2018 [136].

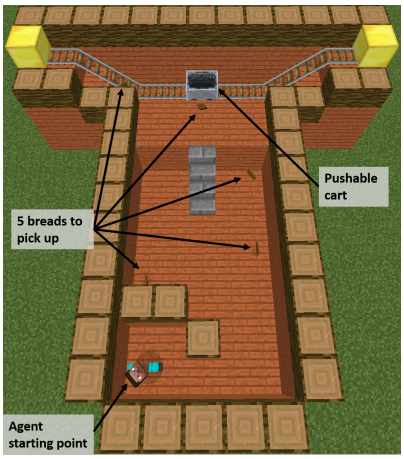


Figure 9. Malmo Mountain Cart. In this episodic environment the agent starts at the bottom left corner of the arena and is able to act on the environment using 2 continuous action commands: move and strafe. If the agent manages to get out of its starting area it may be able to collect items dispatched within the arena. If the agent manages to climb the stairs it may get close enough to the cart to move it along its railroad.

However, in the case of more complex environments containing multiple objects or distractors, an efficient exploration requires that the structure of the goal space reflects the one of the environment. We studied how the structure of the learned goal space using a representation learning algorithm impacts the exploration phase. In particular, we studied how disentangled representations compare to their entangled counterparts in a paper published at CoRL 2019 [101], associated with a blog post available at: https://openlab-flowers.inria.fr/t/discovery-of-independently-controllable-features-through-autonomous-goal-setting/494.

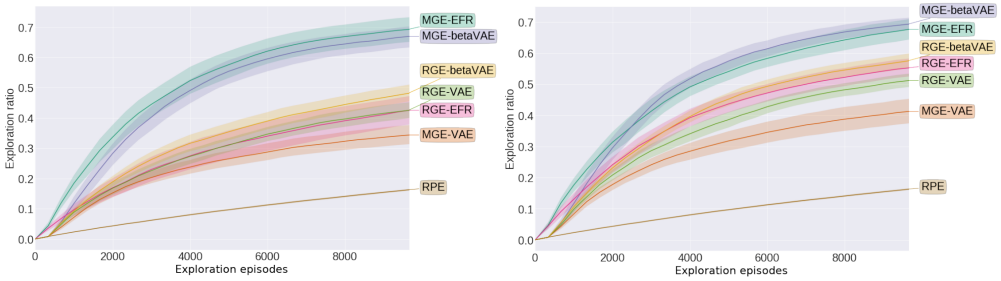
Those ideas were evaluated on a simple benchmark where a seven joints robotic arm evolves in an environment containing two balls. One of the ball can be grasped by the arm and moved around whereas the second one acts as a distractor: it cannot be grasped by the robotic arm and moves randomly across the environment.

Our results showed that using a disentangled goal space leads to better exploration performances than an entangled goal space: the goal exploration algorithm discovers a wider variety of outcomes in less exploration steps (see Figure 10). We further showed that when the representation is disentangled, one can leverage it by sampling goals that maximize learning progress in a modular manner. Lastly, we have shown that the measure of learning progress, used to drive curiosity-driven exploration, can be used simultaneously to discover abstract independently controllable features of the environment.

Finally, we experimented the applicability of those principles on a real-world robotic setup, where a 6-joint robotic arm learns to manipulate a ball inside an arena, by choosing goals in a space learned from its past experience, presented in this technical report: https://arxiv.org/abs/1906.03967.

7.2.1.4. Monolithic Intrinsically Motivated Modular Multi-Goal Reinforcement Learning

In this project we merged two families of algorithms. The first family is the population-based Intrinsically Motivated Goal Exploration Processes (IMGEP) developped in the team (see [83] for a presentation). In this family, autonomous learning agents sets their own goals and learn to reach them. Intrinsic motivation under the form of absolute learning progress is used to guide the selection of goals to target. In some variations of this framework, goals can be represented as coming from different *modules* or *tasks*. Intrinsic motivations are then used to guide the choice of the next task to target.



(a) Small exploration noise ($\sigma = 0.05$) (b) Large exploration noise ($\sigma = 0.1$)

Figure 10. Exploration ratio during exploration for different exploration noises. Architectures with disentangled representations as a goal space (βVAE) explore more than those with entangled representations (VAE). Similarly modular architectures (MGE) explore more than flat architecture (RGE).

The second family encompasses goal-parameterized reinforcement learning algorithms. The first algorithm of this category used an architecture called Universal Value Function Approximators (UVFA), and enabled to train a single policy on an infinite number of goals (continuous goal spaces) [144] by appending the current goal to the input of the neural network used to approximate the value function and the policy. Using a single network allows to share weights among the different goals, which results in faster learning (shared representations). Later, HER [51] introduced a goal replay policy: the actual goal aimed at, could be replaced by a fictive goal when learning. This could be thought of as if the agent were pretending it wanted to reach a goal that it actually reached later on in the trajectory, in place of the true goal. This enables cross-goal learning and speeds up training. Finally, UNICORN [111] proposed to use UVFA to achieve multi-task learning with a discrete task-set.

In this project, we developed CURIOUS [33] (ICML 2019), an intrinsically motivated reinforcement learning algorithm able to achieve both multiple tasks and multiple goals with a single neural policy. It was tested on a custom multi-task, multi-goal environment adapted from the OpenAI Gym Fetch environments [61], see Figure 11 . CURIOUS is inspired from the second family as it proposes an extension of the UVFA architecture. Here, the current task is encoded by a one-hot code corresponding to the task id. The goal is of size $\sum_{i=1}^{N} dim(\mathcal{G}_i)$ where $\mathcal{G}_i$ is the goal space corresponding to task T_i. All components are zeroed except the ones corresponding to the current goal g_i of the current task T_i, see Figure 12 .

CURIOUS is also inspired from the first family, as it self-generates its own tasks and goals and uses a measure of learning progress to decide which task to target at any given moment. The learning progress is computed as the absolute value of the difference of non-overlapping window average of the successes or failures

$$LP_i(t) = \frac{|\sum_{\tau=t-2l}^{t-l} S_\tau - \sum_{\tau=t-l}^{t} S_\tau|}{2l},$$

where S_τ describes a success (1) or a failure (0) and l is a time window length. The learning progress is then used in two ways: it guides the selection of the next task to attempt, and it guides the selection of the task to replay. Cross-goal and cross-task learning are achieved by replacing the goal and/or task in the transition by another. When training on one combination of task and goal, the agent can therefore use this sample to learn about other tasks and goals. Here, we decide to replay and learn more on tasks for which the absolute learning progress is high. This helps for several reasons: 1) the agent does not focus on already learned tasks, as the corresponding learning progress is null, 2) the agent does not focus on impossible tasks for the same reason.

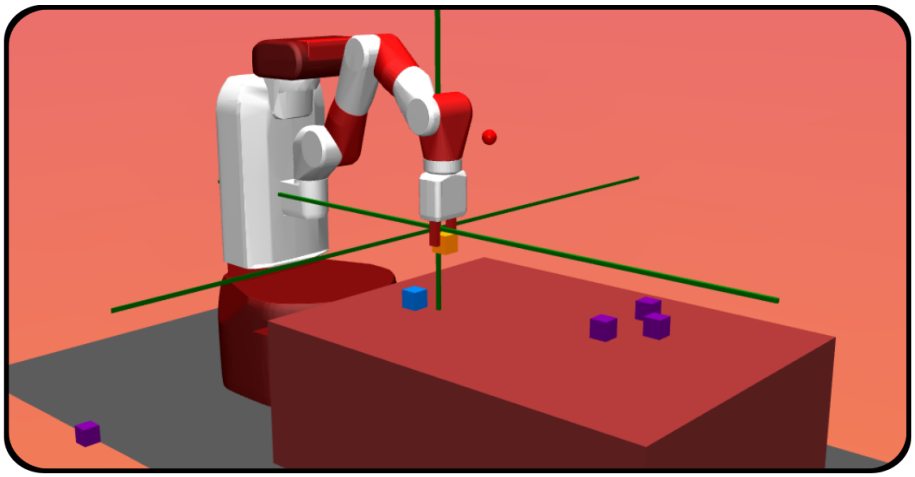


Figure 11. Custom multi-task and multi-goal environment to test the CURIOUS algorithm.

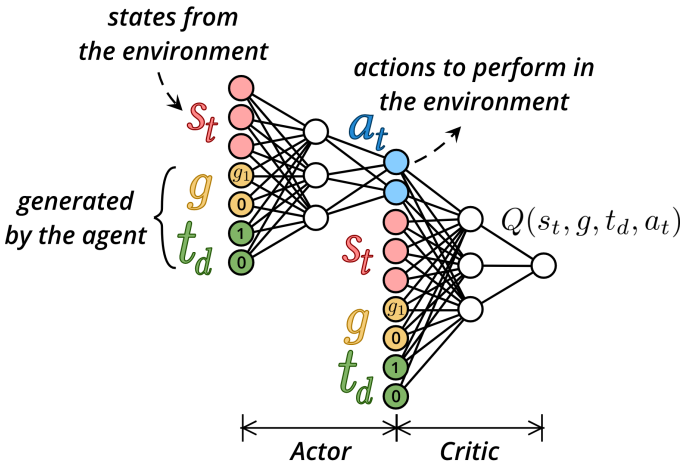


Figure 12. Architecture extended from Universal Value Function Approximators. In this example, the agent is targeting task T_1 among two tasks, each corresponding to a 1 dimension goal.

The agent focuses more on tasks that are being learned (therefore maximizing learning progress), and on tasks that are being forgotten (therefore fighting the problem of forgetting). Indeed, when many tasks are learned in a same network, chances are tasks that are not being attempted often will be forgotten after a while.

In this project, we compare CURIOUS to two baselines: 1) a flat representation algorithm where goals are set from a multi dimensional space including all tasks (equivalent to HER); 2) a task-expert algorithm where a multi-goal UVFA expert policy is trained for each task. The results are shown in Figure 13 .

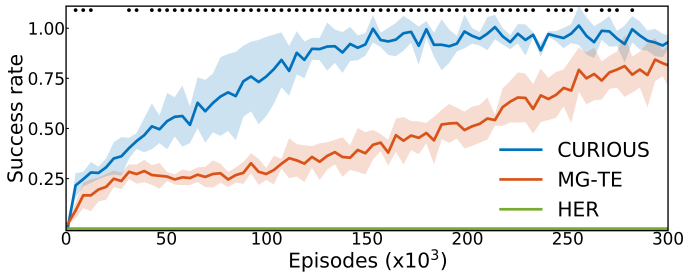


Figure 13. Comparison of CURIOUS to alternative algorithms.

7.2.1.5. Autonomous Multi-Goal Reinforcement Learning with Natural Language

This project follows the CURIOUS project on intrinsically motivated modular multi-goal reinforcement learning [33]. In the CURIOUS algorithm, we presented an agent able to tackle multiple goals of multiple types using a single controller. However, the agent needed to have access to the description of each of the goal types, and their associated reward functions. This represents a considerable amount of prior knowledge for the engineer to encode into the agent. In our new project, the agent builds its own representations of goals, can tackle a growing set of goals, and learns its own reward function, all this through interactions in natural language with a social partner.

The agent does not know any potential goal at first, and act randomly. As it reaches outcomes that are meaningful for the social partner, the social partner provides descriptions of the scene in natural language. The agent stores descriptions and corresponding states for two purposes. First it builds a list of potential goals, reaching back these outcomes that the social partner described. Second, it uses the combination of state and state description to learn a reward function, mapping current state and language descriptions to a binary feedback: 1 if the description is satisfied by the current state, 0 if not.

The agent sets goals to itself from the set of previously discovered descriptions, and is able to learn how to reach them thanks to its learned internal reward function. Concretely, the agent learns a set of 50+ goals from these interactions. We showed co-learning of the reward function and the policy did not produce consequent overhead compared to using an oracle reward function (see Figure 14). This project led to an article accepted at the NeurIPS workshop Visually Grounded Interaction and Language [35]. Current work aims at learning the language model mapping the description into a continuous goal space used as input to the policy and reward function using recurrent networks.

7.2.1.6. Intrinsically Motivated Exploration and Multi-Goal RL with First-Person Images

The aim of this project is to create an exploration process in first-person 3D environments. Following the work presented in [121], [135], [157] the current algorithm is setup in a similar manner. The agent observes the environment in a first person manner and is given a goal to reach. Furthermore, the goal policy samples goals that encourage the agent to explore the environment.

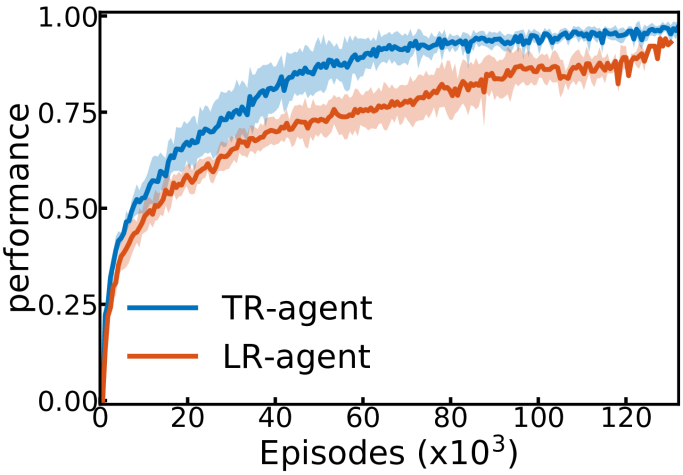


Figure 14. Comparison of performance of agents having access to the true oracle reward function (TR-agent), and agents co-learning policy and reward function (LR-agent).

Currently, there are two ways of representing and sampling goals. Following [157] the goal policy has a buffer of states previously visited by the agent and samples from this buffer the next goals as first person observations. In this setting a goal conditioned reward function is also learned in the form of a reachability network introduced in [142]. On the other hand, following [136], [121] and [135] one can learn a latent representation of a goal using an autoencoder (VAE) and then sample from this generative model or in the latent space. Then, we can use the L2 distance in the latent space as a reward function. The experiments are conducted on a set of Unity environments created in the team.

7.2.2. Teacher algorithms for curriculum learning of Deep RL in continuously parameterized environments

Participants: Remy Portelas [correspondant], Katja Hoffman, Pierre-Yves Oudeyer.

In this work we considered the problem of how a teacher algorithm can enable an unknown Deep Reinforcement Learning (DRL) student to become good at a skill over a wide range of diverse environments. To do so, we studied how a teacher algorithm can learn to generate a learning curriculum, whereby it sequentially samples parameters controlling a stochastic procedural generation of environments. Because it does not initially know the capacities of its student, a key challenge for the teacher is to discover which environments are easy, difficult or unlearnable, and in what order to propose them to maximize the efficiency of learning over the learnable ones. To achieve this, this problem is transformed into a surrogate continuous bandit problem where the teacher samples environments in order to maximize absolute learning progress of its student. We presented ALP-GMM (see figure 15), a new algorithm modeling absolute learning progress with Gaussian mixture models. We also adapted existing algorithms and provided a complete study in the context of DRL. Using parameterized variants of the BipedalWalker environment, we studied their efficiency to personalize a learning curriculum for different learners (embodiments), their robustness to the ratio of learnable/unlearnable environments, and their scalability to non-linear and high-dimensional parameter spaces. Videos and code are available at https://github.com/flowersteam/teachDeepRL.

Overall, this work demonstrated that LP-based teacher algorithms could successfully guide DRL agents to learn in difficult continuously parameterized environments with irrelevant dimensions and large proportions of unfeasible tasks. With no prior knowledge of its student's abilities and only loose boundaries on the task

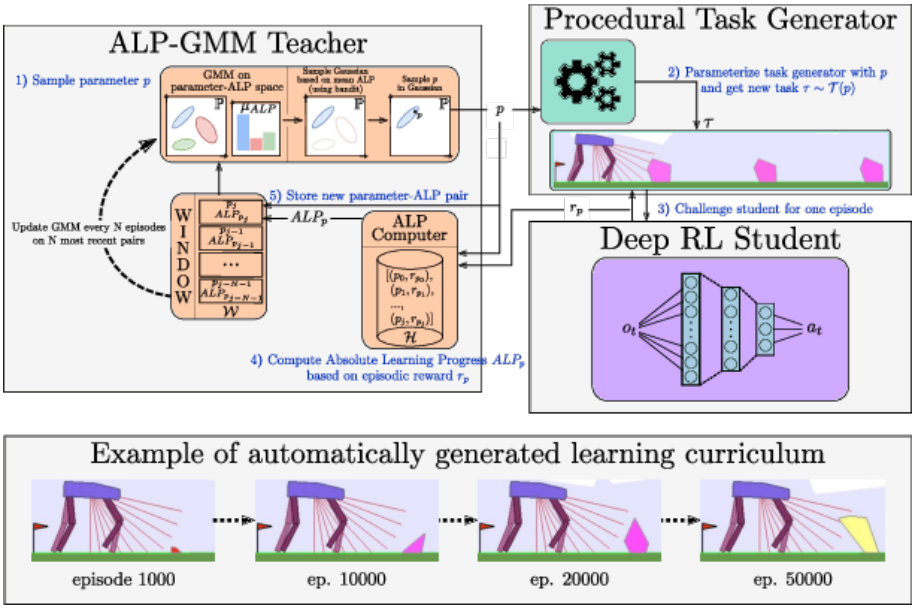


Figure 15. Schematic view of an ALP-GMM teacher's workflow

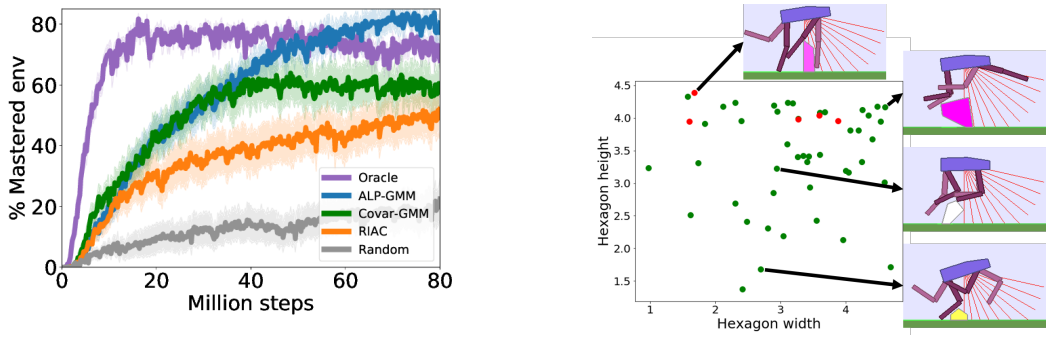


Figure 16. ***Teacher-Student approaches in Hexagon Tracks.Left:*** *Evolution of mastered tracks for Teacher-Student approaches in Hexagon Tracks. 32 seeded runs (25 for Random) of 80 Millions steps where performed for each condition. The mean performance is plotted with shaded areas representing the standard error of the mean.* ***Right:*** *A visualization of which track distributions of the test-set are mastered (i.e* $r_t > 230$*, shown by green dots) by an ALP-GMM run after 80 million steps.*

space, ALP-GMM, our proposed teacher, consistently outperformed random heuristics and occasionally even expert-designed curricula (see figure 16). This work was presented at CoRL 2019 [38].

ALP-GMM, which is conceptually simple and has very few crucial hyperparameters, opens-up exciting perspectives inside and outside DRL for curriculum learning problems. Within DRL, it could be applied to previous work on autonomous goal exploration through incremental building of goal spaces [101]. In this case several ALP-GMM instances could scaffold the learning agent in each of its autonomously discovered goal spaces. Another domain of applicability is assisted education, for which current state of the art relies heavily on expert knowledge [68] and is mostly applied to discrete task sets.

7.3. Automated Discovery in Self-Organizing Systems

7.3.1. Curiosity-driven Learning for Automated Discovery of Physico-Chemical Structures

Participants: Chris Reinke [correspondant], Mayalen Etcheverry, Pierre-Yves Oudeyer.

7.3.1.1. Introduction

Intrinsically motivated goal exploration algorithms (IMGEPs) enable machines to discover repertoires of action policies that produce a diversity of effects in complex environments. In robotics, these exploration algorithms have been shown to allow real world robots to acquire skills such as tool use [81] [55]. In other domains such as chemistry and physics, they open the possibility to automate the discovery of novel chemical or physical structures produced by complex dynamical systems [134]. However, they have so far assumed that self-generated goals are sampled in a specifically engineered feature space, limiting their autonomy. Recent work has shown how unsupervised deep learning approaches could be used to learn goal space representations [136] but they have used precollected data to learn the representations. This project studies how IMGEPs can be extended and used for automated discovery of behaviours of dynamical systems in physics or chemistry without using assumptions or knowledge about such systems.

As a first step towards this goal we choose Lenia [66], a simulated high-dimensional complex dynamical system, as a target system. Lenia is a continuous cellular automaton where diverse visual structures can self-organize (Fig.17 , c). It consists of a two-dimensional grid of cells $A \in [0,1]^{256\times256}$ where the state of each cell is a real-valued scalar activity $A^t(x) \in [0,1]$. The state of cells evolves over discrete time steps t. The activity change is computed by integrating the activity of neighbouring cells. Lenia's behavior is controlled by its initial pattern $A^{t=1}$ and several settings that control the dynamics of the activity change. Lenia can produce diverse patterns with different dynamics. Most interesting, spatially localized coherent patterns that resemble in their shapes microscopic *animals* can emerge. Our goal was to develop methods that allow to explore a high diversity of such animal patterns.

We could successfully accomplish this goal [30] based on two key contributions of our research: 1) the usage of compositional pattern producing networks (CPPNs) for the generation of initial states for Lenia, and 2) the development of a novel IMGEP algorithm that learns goal representations online during the exploration of the system.

7.3.1.2. 1) CPPNs for the generation of initial states

A key role in the generation of patterns in dynamical systems is their initial state $A^{t=1}$. IMGEPs sample these initial states and apply random perturbations to them during the exploration. For Lenia this state is a two-dimensional grid with 256×256 cells. Performing directly a random sampling of the 256×256 grid cells results in initial patterns that resemble white noise. Such random states result mainly in the emergence of global patterns that spread over the whole state space, complicating the search for spatially localized patterns. We solved the sampling problem for the initial states by using compositional pattern producing networks (CPPNs) [148]. CPPNs are recurrent neural networks that allow the generation of structured initial states (Fig.17 , a). The CPPNs are used as part of the system parameters which are explored by the algorithms. They are defined by their network structure (number of neurons, connections between neurons) and their connection weights. They include a mechanism for random mutation of the weights and structure.

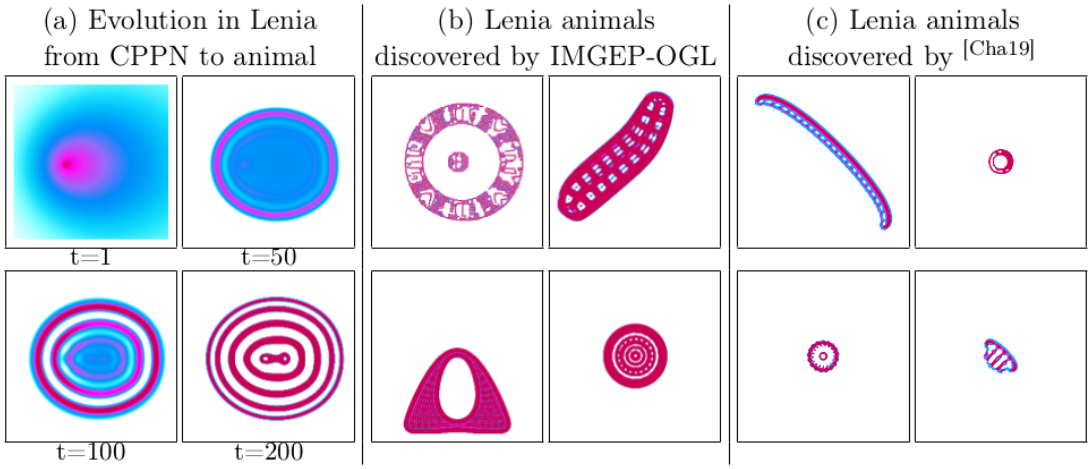


Figure 17. Example patterns produced by the Lenia system. Illustration of the dynamical morphing from an initial CPPN image to an animal (a). The automated discovery (b) is able to find similar complex animals as a human-expert manual search (c) by [66].

7.3.1.3. 2) IMGEP for Online Learning of Goal Space Representations

We proposed an online goal space learning IMGEP (IMGEP-OGL), which learns the goal space incrementally during the exploration process. A variational autoencoder (VAE) is used to encode Lenia patterns into a 8-dimensional latent representation used as goal space. The training procedure of the VAE is integrated in the goal sampling exploration process by first initializing the VAE with random weights. The VAE network is then trained every K explorations for E epochs on the previously idetnfied patterns during the exploration.

7.3.1.4. Experiments

We evaluated the performance of the novel IMGEP-OGL to other exploration algorithms by comparing the diversity of their identified patterns. Diversity is measured by the spread of the exploration in an *analytic behavior space*. This space is defined by a latent representation space that was build through the training of a VAE to learn the important features over a very large dataset of Lenia patterns identified during the many experiments over all evaluated algorithms. We then augmented that space by concatenating hand-defined features. Each identified Lenia pattern is represented by a specific point in this space. The space was then discretized in a fixed number of areas/bins of equal size. The final diversity measure of each algorithm is the number of areas/bins in which at least one explored pattern exists.

We compared different exploration algorithms to the novel IMGEP-OGL: 1) Random exploration of system parameters, 2) IMGEP-HGS: IMGEP with a hand-defined goal space, 3) IMGEP-PGL: IMGEP with a learned goal space via an VAE by a precollected dataset of Lenia patterns, and 4) IMGEP-RGS: IMGEP with a VAE with random weights that defines the goal space.

The system parameters θ consisted of a CPPN that generates the initial state $A^{t=1}$ for Lenia and 6 further settings defining Lenia's dynamics: $\theta = [\text{CPPN} \rightarrow A^{t=1}, R, T, \mu, \sigma, \beta_1, \beta_2, \beta_3]$. The CPPNs were initialized and mutated by a random process that defines their structure and connection weights as done. The random initialization of the other Lenia settings was done by an uniform distribution and their mutation by a Gaussian distribution around the original values.

7.3.1.5. Results

The diversity of identified patterns in the analytic behavior space show that IMGEP approaches with learned goal spaces via VAEs (PGL, OGL) could identify the highest diversity of patterns overall (Fig. 18 , a). They were followed by the IMGEP with a hand-defined goal space (HGS). The lowest performance had the random exploration and the IMGEP with a random goal space (RGS). The advantage of learned goals space approaches (PGL, OGL) over all other approaches was even stronger for the diversity of animal patterns, i.e. the main goal of our exploration (Fig. 18 , b).

(a) Diversity of all identified Patterns (b) Diversity of Animal Patterns

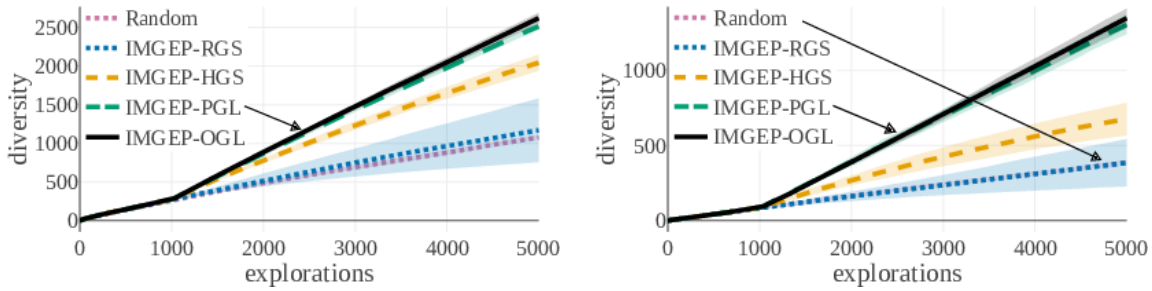


Figure 18. (a) All IMGEPs reach a higher diversity in the analytic behavior space over all patterns than random search. (b) IMGEPs with a learned goal space are especially successful in identifying a diversity of animal patterns. Depicted is the average diversity ($n = 10$) with the standard deviation as shaded area (for some not visible because it is too small).

7.3.1.6. Conclusion

Our goal was to investigate new techniques based on intrinsically motivated goal exploration for the automated discovery of patterns and behaviors in complex dynamical systems. We introduced a new algorithm (IMGEP-OGL) which is capable of learning unsupervised goal space representations during the exploration of an unknown system. Our results for Lenia, a high-dimensional complex dynamical system, show its superior performance over hand-defined goal spaces or random exploration. It shows the same performance as a learned goal space based on precollected data, showing that such a precollection of data is not necessary. We furthermore introduced the usage of CPPNs for the successful initialization of the intial states of the dynamical systems. Both advances allowed us to explore an unknown and high-dimensional dynamical system which shares many similarities with different physical or chemical systems.

7.4. Representation Learning

7.4.1. State Representation Learning in the Context of Robotics

Participants: David Filliat [correspondant], Natalia Diaz Rodriguez, Timothee Lesort, Antonin Raffin, René Traoré, Ashley Hill, Te Sun, Lu Lin, Guanghang Cai, Bunthet Say.

During the DREAM project, we participated in the development of a conceptual framework of open-ended lifelong learning [77] based on the idea of representational re-description that can discover and adapt the states, actions and skills across unbounded sequences of tasks.

In this context, State Representation Learning (SRL) is the process of learning without explicit supervision a representation that is sufficient to support policy learning for a robot. We have finalized and published a large state-of-the-art survey analyzing the existing strategies in robotics control [103], and we developed unsupervised methods to build representations with the objective to be minimal, sufficient, and that encode the relevant information to solve the task. More concretely, we used the developed and open sourced[0] the S-RL

[0] https://github.com/araffin/robotics-rl-srl

toolbox [137] containing baseline algorithms, data generating environments, metrics and visualization tools for assessing SRL methods. Part of this study is the [105] where we present a robustness analysis on Deep unsupervised state representation learning with robotic priors loss functions.

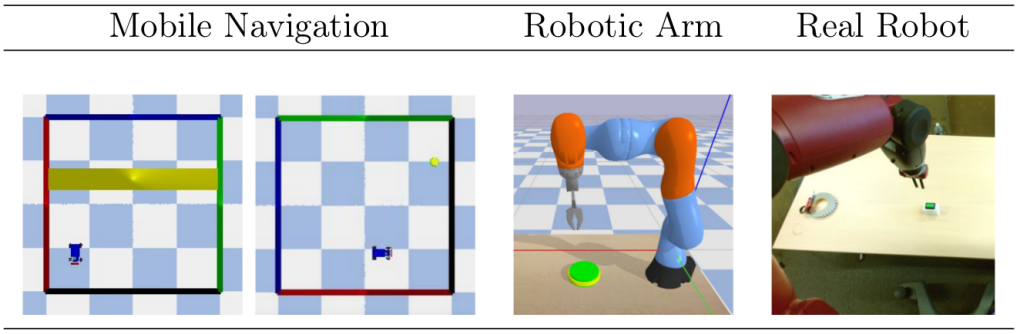


Figure 19. Environments and datasets for state representation learning.

The environments proposed in Fig. 19 are variations of two environments: a 2D environment with a mobile robot and a 3D environment with a robotic arm. In all settings, there is a controlled robot and one or more targets (that can be static, randomly initialized or moving). Each environment can either have a continuous or discrete action space, and the reward can be sparse or shaped, allowing us to cover many different situations.

The evaluation and visualization tools are presented in Fig. 20 and make it possible to qualitatively verify the learned state space behavior (e.g., the state representation of the robotic arm dataset is expected to have a continuous and correlated change with respect to the arm tip position).

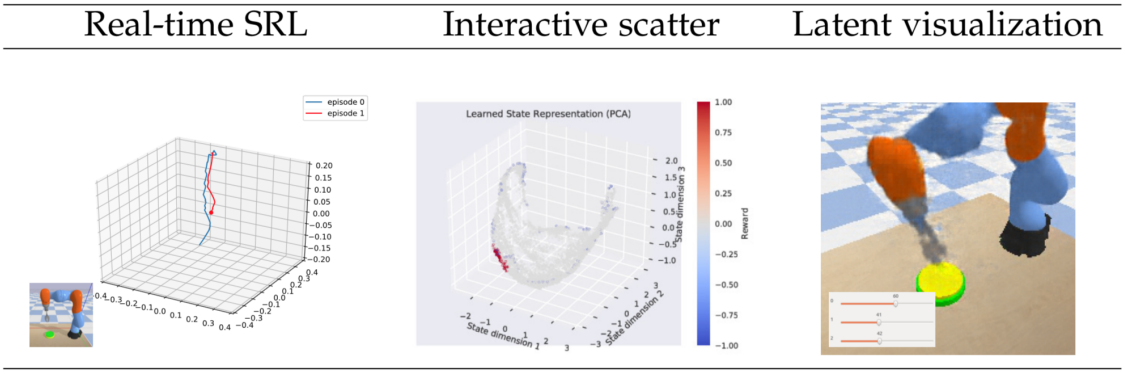


Figure 20. Visual tools for analysing SRL; Left: Live trajectory of the robot in the state space. Center: 3D scatter plot of a state space; clicking on any point displays the corresponding observation. Right: reconstruction of the point in the state space defined by the sliders.

We also proposed a new approach that consists of learning a state representation that is split into several parts where each optimizes a fraction of the objectives. In order to encode both target and robot positions, auto-encoders, reward and inverse model losses are used.

The latest work on decoupling feature extraction from policy learning, was presented at the SPIRL workshop at ICLR2019 in New Orleans, LA [138]. We assessed the benefits of state representation learning in goal based robotic tasks, using different self-supervised objectives.

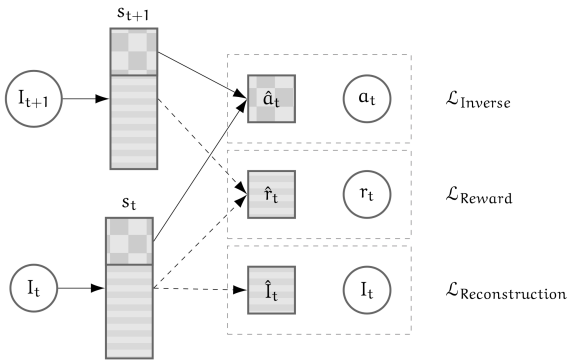


Figure 21. SRL Splits model: combines a reconstruction of an image I, a reward (r) prediction and an inverse dynamic models losses, using two splits of the state representation s. Arrows represent model learning and inference, dashed frames represent losses computation, rectangles are state representations, circles are real observed data, and squares are model predictions.

Because combining objectives into a single embedding is not the only option to have features that are *sufficient* to solve the tasks, by stacking representations, we favor *disentanglement* of the representation and prevent objectives that can be opposed from cancelling out. This allows a more stable optimization. Fig. 21 shows the split model where each loss is only applied to part of the state representation.

As using the learned state representations in a Reinforcement Learning setting is the most relevant approach to evaluate the SRL methods, we use the developed S-RL framework integrated algorithms (A2C, ACKTR, ACER, DQN, DDPG, PPO1, PPO2, TRPO) from Stable-Baselines [92], Augmented Random Search (ARS), Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) and Soft Actor Critic (SAC). Due to its stability, we perform extensive experiments on the proposed datasets using PPO and states learned with the approaches described in [137] along with ground truth (GT).

Table 22 illustrates the qualitative evaluation of a state space learned by combining forward and inverse models on the mobile robot environment. It also shows the performance of PPO algorithm based on the states learned by several baseline approaches.

We verified that our new approach (described in Task 2.1) makes it possible for reinforcement learning to converge faster towards the optimal performance in both environments with the same amount of budget timesteps. Learning curve in Fig. 23 shows that our unsupervised state representation learned with the split model even improves on the supervised case.

7.4.2. Continual learning

Participants: David Filliat [correspondant], Natalia Díaz Rodríguez, Timothee Lesort, Hugo Caselles-Dupré.

Continual Learning (CL) algorithms learn from a stream of data/tasks continuously and adaptively through time to better enable the incremental development of ever more complex knowledge and skills. The main problem that CL aims at tackling is catastrophic forgetting [115], i.e., the well-known phenomenon of a neural network experiencing a rapid overriding of previously learned knowledge when trained sequentially on new data. This is an important objective quantified for assessing the quality of CL approaches, however, the almost exclusive focus on catastrophic forgetting by continual learning strategies, lead us to propose a

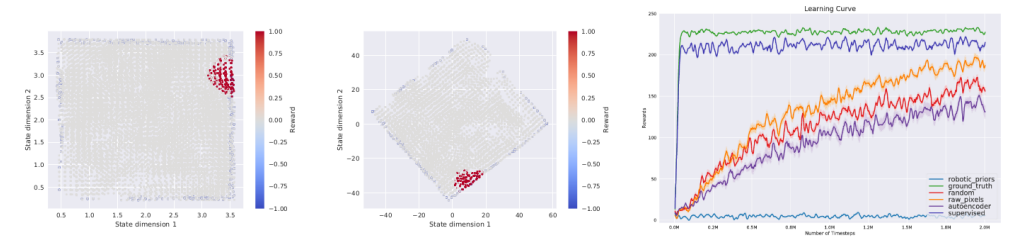


Figure 22. Ground truth states (left), states learned (Inverse and Forward) (center), and RL performance evaluation (PPO) (right) for different baselines in the mobile robot environment. Colour denotes the reward, red for positive, blue for negative and grey for null reward (left and center).

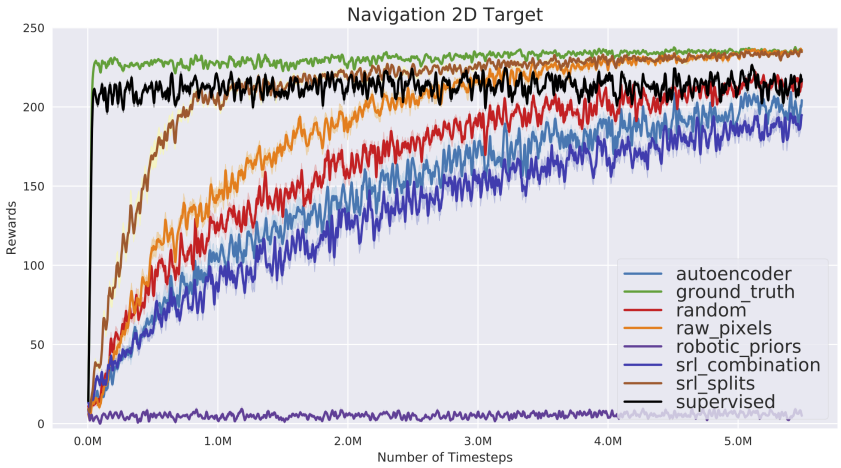


Figure 23. Performance (mean and standard error for 10 runs) for PPO algorithm for different state representations learned in Navigation 2D random target environment.

set of comprehensive, implementation independent metrics accounting for factors we believe have practical implications worth considering with respect to the deployment of real AI systems that learn continually, and in "Non-static" machine learning settings. In this context we developed a framework and a set of comprehensive metrics [78] to tame the lack of consensus in evaluating CL algorithms. They measure Accuracy (A), Forward and Backward (/*remembering*) knowledge transfer (FWT, BWT, REM), Memory Size (MS) efficiency, Samples Storage Size (SSS), and Computational Efficiency (CE). Results on iCIFAR-100 classification sequential class learning is in Table 24 .

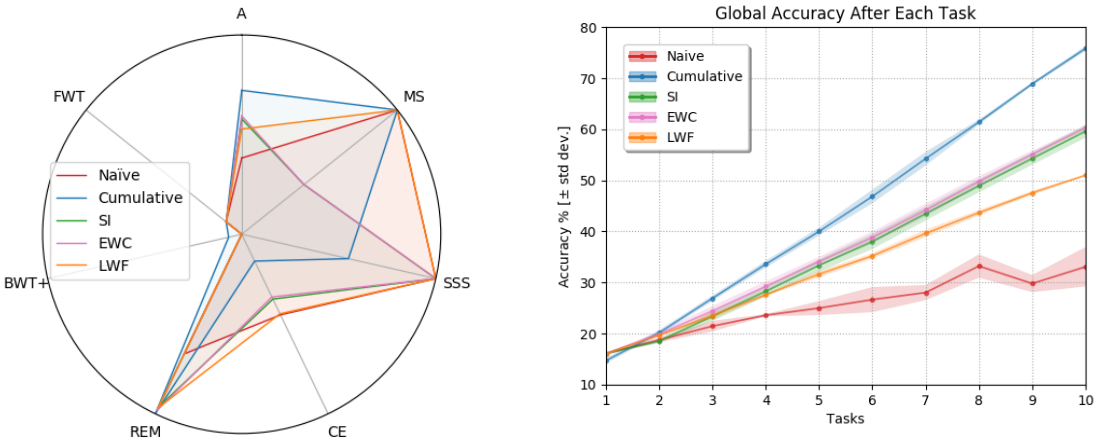


Figure 24. (left) Spider chart: CL metrics per strategy (larger area is better) and (right) Accuracy per CL strategy computed over the fixed test set.

Generative models can also be evaluated from the perspective of Continual learning which we investigated in our work [102]. This work aims at evaluating and comparing generative models on disjoint sequential image generation tasks. We study the ability of Generative Adversarial Networks (GANS) and Variational Auto-Encoders (VAEs) and many of their variants to learn sequentially in continual learning tasks. We investigate how these models learn and forget, considering various strategies: rehearsal, regularization, generative replay and fine-tuning. We used two quantitative metrics to estimate the generation quality and memory ability. We experiment with sequential tasks on three commonly used benchmarks for Continual Learning (MNIST, Fashion MNIST and CIFAR10). We found (see Figure 26) that among all models, the original GAN performs best and among Continual Learning strategies, generative replay outperforms all other methods. Even if we found satisfactory combinations on MNIST and Fashion MNIST, training generative models sequentially on CIFAR10 is particularly instable, and remains a challenge. This work has been published at the NIPS workshop on Continual Learning 2018.

Another extension of previous section on state representation learning (SRL) to the continual learning setting is in our paper [65]. This work proposes a method to avoid catastrophic forgetting when the environment changes using generative replay, i.e., using generated samples to maintain past knowledge. State representations are learned with variational autoencoders and automatic environment change is detected through VAE reconstruction error. Results show that using a state representation model learned continually for RL experiments is beneficial in terms of sample efficiency and final performance, as seen in Figure 26 . This work has been published at the NIPS workshop on Continual Learning 2018 and is currently being extended.

The experiments were conducted in an environment built in the lab, called Flatland [64]. This is a lightweight first-person 2-D environment for Reinforcement Learning (RL), designed especially to be convenient for Continual Learning experiments. Agents perceive the world through 1D images, act with 3 discrete actions,

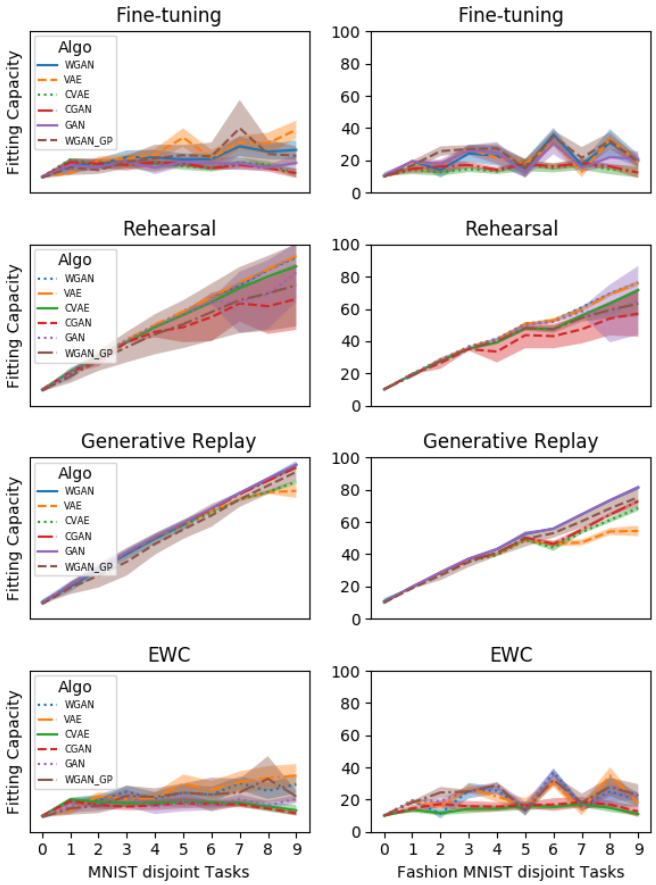


Figure 25. Means and standard deviations over 8 seeds of Fitting Capacity metric evaluation of VAE, CVAE, GAN, CGAN and WGAN. The four considered CL strategies are: Fine Tuning, Generative Replay, Rehearsal and EWC. The setting is 10 disjoint tasks on MNIST and Fashion MNIST.

and the goal is to learn to collect edible items with RL. This work has been published at the ICDL-Epirob workshop on Continual Unsupervised Sensorimotor Learning 2018, and was accepted as oral presentation.

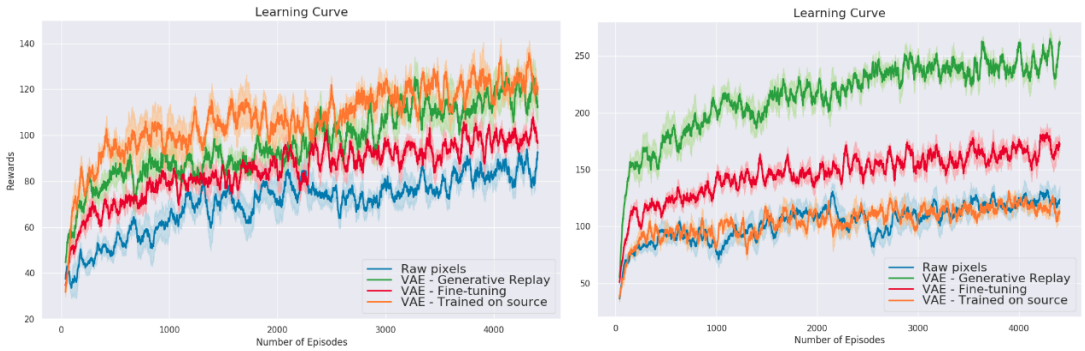


Figure 26. Mean reward and standard error over 5 runs of RL evaluation using PPO with different types of inputs. Fine-tuning and Generative Replay models are trained sequentially on the first and second environment, and then used to train a policy for both tasks. Generative Replay outperforms all other methods. It shows the need for continually learning features in State Representation Learning in settings where the environment changes.

In the last year, we published a survey on continual learning models, metrics and contributed a CL framework to categorize the approaches on this area [104]. Figure 27 shows the different approaches cited and the strategies proposed and a small subset of examples analyzed.

We also worked on validating a distillation approach for multitask learning in a continual learning reinforcement learning setting [152], [153].

Applying State Representation Learning (SRL) into a continual learning setting of reinforcement learning was possible by learning a compact and efficient representation of data that facilitates learning a policy. The proposed a CL algorithm based on distillation does not manually need to be given a task indicator at test time, but learns to infer the task from observations only. This allows to successfully apply the learned policy on a real robot.

We present 3 different 2D navigation tasks to a 3 wheel omni-directional robot to be learned to be solved sequentially. The robot has first access to task 1 only, and then to task 2 only, and so on. It should learn a single policy that solves all tasks and be applicable in a real life scenario. The robot can perform 4 high level discrete actions (move left/right, move up/down). The tasks where the method was validated are in Fig. 28 :

Task 1: Target Reaching (TR): Reaching a red target randomly positioned.

Task 2: Target Circling (TC): Circling around a fixed blue target.

Task 3: Target Escaping (TE): Escaping a moving robot.

DisCoRL (Distillation for Continual Reinforcement learning) is a modular, effective and scalable pipeline for continual RL. This pipeline uses policy distillation for learning without forgetting, without access to previous environments, and without task labels in order to transfer policies into real life scenarios [152]. It was presented as an approach for continual reinforcement learning that sequentially summarizes different learned policies into a dataset to distill them into a student model. Some loss in performance may occur while transferring knowledge from teacher to student, or while transferring a policy from simulation to real life. Nevertheless, the experiments show promising results when learning tasks sequentially, in simulated environments and real life settings.

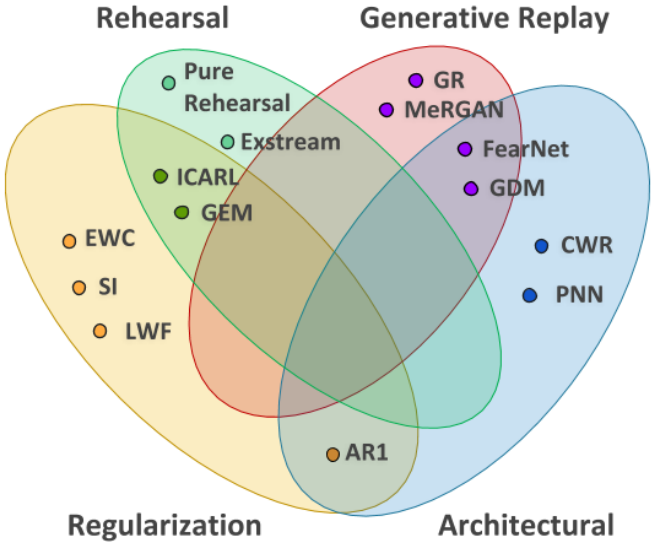


Figure 27. Venn diagram of some of the most popular CL strategies w.r.t the main approaches in the literature (CWR,PNN EWC, SI, LWF, ICARL, GEM, FearNet, GDM, ExStream, GR, MeRGAN, and AR1. Rehearsal and Generative Replay upper categories can be seen as a subset of replay strategies. Better viewed in color [104].

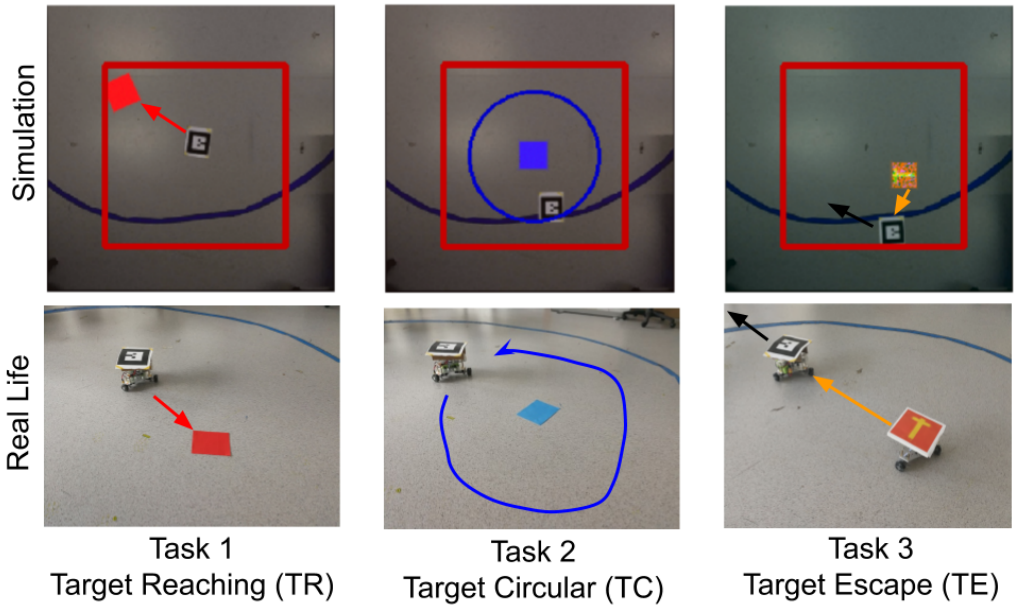


Figure 28. The three tasks, in simulation (top) and in real life (bottom), sequentially experienced. Learning is performed in simulation, the real life setting is only used at test time.

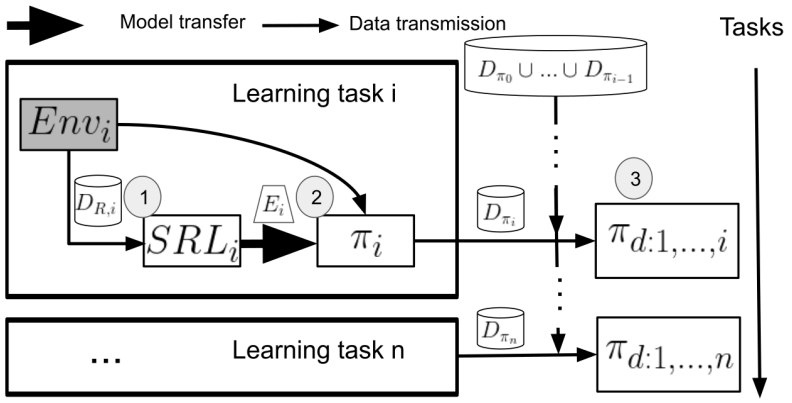


Figure 29. White cylinders are for datasets, gray squares for environments, and white squares for learning algorithms, whose name correspond to the model trained. Each task i is learned sequentially and independently by first generating a dataset $D_{R,i}$ with a random policy to learn a state representation with an encoder E_i with an SRL method (1), then we use E_i and the environment to learn a policy π_i in the state space (2). Once trained, π_i is used to create a distillation dataset D_{π_i} that acts as a memory of the learned behaviour. All policies are finally compressed into a single policy $\pi_{d:1,..,i}$ by merging the current dataset D_{π_i} with datasets from previous tasks $D_{\pi_1} \cup ... \cup D_{\pi_{i-1}}$ and using distillation (3).

The overview of DisCoRL full pipeline for Continual Reinforcement Learning is in Fig. 29 .

7.4.3. Disentangled Representation Learning for agents

Participants: Hugo Caselles-Dupré [correspondant], David Filliat.

Finding a generally accepted formal definition of a disentangled representation in the context of an agent behaving in an environment is an important challenge towards the construction of data-efficient autonomous agents. Higgins et al. (2018) recently proposed Symmetry-Based Disentangled Representation Learning, a definition based on a characterization of symmetries in the environment using group theory. We build on their work and make observations, theoretical and empirical, that lead us to argue that Symmetry-Based Disentangled Representation Learning cannot only be based on static observations: agents should interact with the environment to discover its symmetries.

Our research was published in NeuRIPS 2019 [32] at Vancouver, Canada.

7.5. Tools for Understanding Deep Learning Systems

7.5.1. Explainable Deep Learning

Participants: Natalia Díaz Rodríguez [correspondant], Adrien Bennetot.

Together with Segula Technologies and Sorbonne Université, ENSTA Paris has been working on eXplainable Artificial Intelligence (XAI) in order to make machine learning more interpretable. While opaque decision systems such as Deep Neural Networks have great generalization and prediction skills, their functioning does not allow obtaining detailed explanations of their behaviour. The objective is to fight the trade-off between performance and explainability by combining connectionist and symbolic paradigms [47].

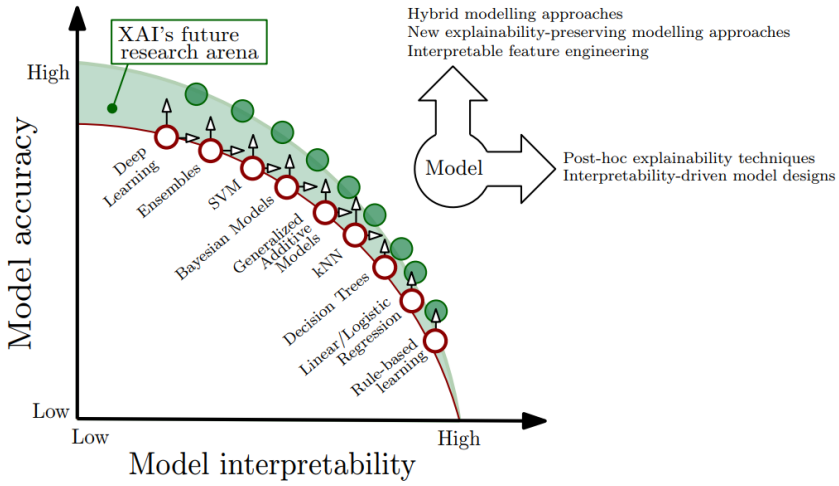


Figure 30. Trade-off between model interpretability and performance, and a representation of the area of improvement where the potential of XAI techniques and tools resides [46].

Broad consensus exists on the importance of interpretability for AI models. However, since the domain has only recently become popular, there is no collective agreement on the different definitions and challenges that constitute XAI. The first step is therefore to summarize previous efforts made in this field. We presented a taxonomy of XAI techniques in [46] and we are currently working on a prediction model that generates itself an explanation of its rationale in natural language while keeping performance as close as possible to the the state of the art [47].

7.5.2. Methods for Statistical Comparison of RL Algorithms

Participants: Cédric Colas [correspondant], Pierre-Yves Oudeyer, Olivier Sigaud.

Following a first article in 2018 [71], we pursued the objective of providing key tools to robustly compare reinforcement learning (RL) algorithms to practioners and researchers. In this year's extension, we compiled a hitchhicker's guide for statistical comparisons of RL algorithms. In particular, we provide a list of statistical tests adapted to compare RL algorithms and compare them in terms of false positive rate and statistial power. In particular, we study the robustness of these tests when their assumptions are violated (non-normal distributions of performances, different distributions, unknown variance, unequal variances etc). We provided an extended study using data from synthetic performance distributions, as well as empirical distributions obtained from running state-of-the-art RL algorithms (TD3 [84] and SAC [88]). From these results we draw a selection of advice for researchers. This study led to an article accepted at the ICLR conference workshop on Reproducibility in Machine Learning [34], to be sumitted to the Neural Networks journal.

7.5.3. Knowledge engineering tools for neural-symbolic learning

Participants: Natalia Díaz Rodríguez [correspondant], Adrien Bennetot.

Symbolic artificial intelligence methods are experiencing a come-back in order to provide deep representation methods the explainability they lack. In this area, a survey on RDF stores to handle ontology-based triple databases has been contributed [97], as well as the use of neural-symbolic tools that aim at integrating both neural and symbolic representations [58].

7.6. Applications in Educational Technologies

7.6.1. Machine Learning for Adaptive Personalization in Intelligent Tutoring Systems

Participants: Pierre-Yves Oudeyer [correspondant], Benjamin Clément, Didier Roy, Helene Sauzeon.

7.6.1.1. The Kidlearn project

Kidlearn is a research project studying how machine learning can be applied to intelligent tutoring systems. It aims at developing methodologies and software which adaptively personalize sequences of learning activities to the particularities of each individual student. Our systems aim at proposing to the student the right activity at the right time, maximizing concurrently his learning progress and its motivation. In addition to contributing to the efficiency of learning and motivation, the approach is also made to reduce the time needed to design ITS systems.

We continued to develop an approach to Intelligent Tutoring Systems which adaptively personalizes sequences of learning activities to maximize skills acquired by students, taking into account the limited time and motivational resources. At a given point in time, the system proposes to the students the activity which makes them progress faster. We introduced two algorithms that rely on the empirical estimation of the learning progress, **RiARiT** that uses information about the difficulty of each exercise and **ZPDES** that uses much less knowledge about the problem.

The system is based on the combination of three approaches. First, it leverages recent models of intrinsically motivated learning by transposing them to active teaching, relying on empirical estimation of learning progress provided by specific activities to particular students. Second, it uses state-of-the-art Multi-Arm Bandit (MAB) techniques to efficiently manage the exploration/exploitation challenge of this optimization process. Third, it leverages expert knowledge to constrain and bootstrap initial exploration of the MAB, while requiring only coarse guidance information of the expert and allowing the system to deal with didactic gaps in its knowledge. The system was evaluated in several large-scale experiments relying on a scenario where 7-8 year old schoolchildren learn how to decompose numbers while manipulating money [68]. Systematic experiments were also presented with simulated students.

7.6.1.2. Kidlearn Experiments 2018-2019: Evaluating the impact of ZPDES and choice on learning efficiency and motivation

An experiment was held between mars 2018 and July 2019 in order to test the Kidlearn framework in classrooms in Bordeaux Metropole. 600 students from Bordeaux Metropole participated in the experiment. This study had several goals. The first goal was to evaluate the impact of the Kidlearn framework on motivation and learning compared to an Expert Sequence without machine learning. The second goal was to observe the impact of using learning progress to select exercise types within the ZPDES algorithm compared to a random policy. The third goal was to observe the impact of combining ZPDES with the ability to let children make different kinds of choices during the use of the ITS. The last goal was to use the psychological and contextual data measures to see if correlation can be observed between the students psychological state evolution, their profile, their motivation and their learning. The different observations showed that generally, algorithms based on ZPDES provided a better learning experience than an expert sequence. In particular, they provide a better motivating and enriching experience to self-determined students. Details of these new results, as well as the overall results of this project, are presented in Benjamin Clément PhD thesis [69] and are currently being processed to be published.

7.6.1.3. Kidlearn and Adaptiv'Math

The algorithms developed during the Kidlearn project and Benjamin Clement thesis [69] are being used in an innovation partnership for the development of a pedagogical assistant based on artificial intelligence intended for teachers and students of cycle 2. The algorithms are being written in typescript for the need of the project. The expertise of the team in creating the pedagogical graph and defining the graph parameters used for the algorithms is also a crucial part of the role of the team for the project. One of the main goal of the team here is to transfer technologies developed in the team in a project with the perspective of industrial scaling and see the impact and the feasibility of such scaling.

7.6.1.4. Kidlearn for numeracy skills with individuals with autism spectrum disorders

Few digital interventions targeting numeracy skills have been evaluated with individuals with autism spectrum disorder (ASD) [114]. Yet, some children and adolescents with ASD have learning difficulties and/or a significant academic delay in mathematics. While ITS are successfully developed for typically developed students to personnalize learning curriculum and then to foster the motivation-learning coupling, they are not or fewly proposed today to student with specific needs. The objective of this pilot study is to test the feasibility of a digital intervention using an STI with high school students with ASD and/or intellectual disability. This application (KidLearn) provides calculation training through currency exchange activities, with a dynamic exercise sequence selection algorithm (ZPDES). 24 students with ASD and/or DI enrolled in specialized classrooms were recruited and divided into two groups: 14 students used the KidLearn application, and 10 students received a control application. Pre-post evaluations show that students using KidLearn improved their calaculation performance, and had a higher level of motivation at the end of the intervention than the control group. These results encourage the use of an STI with students with specific needs to teach numeracy skills, , but need to be replicated on a larger scale. Suggestions for adjusting the interface and teaching method are suggested to improve the impact of the application on students with autism. (Paper is in progress).

7.6.2. Curiosity-driven interaction systems for education

Participants: Pierre-Yves Oudeyer, Hélène Sauzéon [correspondant], Mehdi Alami, Didier Roy, Edith Law.

Three studies have been developed and conducted to newly design curiosity-driven interaction systems aiming to foster learning performance across lifespan : the first two studies include children and the last one includes the older adults.

The first study regards a new interactive robotic system to foster curiosity-driven learning. This led to an article in CHI 2019 [29]. In this work, we explored whether a social peer robot's verbal expression of curiosity can be perceived by participants, produce emotional or behavioural contagion effects, and impact learning. In a between-subject experiment involving 30 participants, a peer robot was manipulated to verbally express: curiosity, curiosity plus rationale, or no curiosity (neutral), within the context of LinkedIt!, a cooperative game we designed for teaching students how to classify rocks. Results show that participants were able to reliably recognize curiosity in the robot and curious robots can be used to elicit significantly more curiosity-driven behaviours among participants.

The second study regards a new interactive educational application to foster curiosity-driven question-asking in children. This study has been performed during the Master 2 internship of Mehdi Alaimi co-supervised by H. Sauzéon, E. Law and PY Oudeyer. The paper submission to CHI'20 is just accepted in december 2019 (« Pedagogical Agents for Fostering Question-Asking Skills in Children »). It addresses a key challenge for 21st-century schools, i.e., teaching diverse students with varied abilities and motivations for learning, such as curiosity within educational settings. Among variables eliciting curiosity state, one is known as « knowledge gap », which is a motor for curiosity-driven exploration and learning. It leads to question-asking which is an important factor in the curiosity process and the construction of academic knowledge. However, children questions in classroom are not really frequent and don't really necessitate deep reasoning. Determined to improve children's curiosity, we developed a digital application aiming to foster curiosity-related question-asking from texts and their perception of curiosity. To assess its efficiency, we conducted a study with 95 fifth grade students of Bordeaux elementary schools. Two types of interventions were designed, one trying to focus children on the construction of low-level question (i.e. convergent) and one focusing them on high-level questions (i.e. divergent) with the help of prompts or questions starters models. We observed that both interventions increased the number of divergent questions, the question fluency performance, while they did not significantly improve the curiosity perception despite high intrinsic motivation scores they have elicited in children. The curiosity-trait score positively impacted the divergent question score under divergent condition, but not under convergent condition. The overall results supported the efficiency and usefulness of digital applications for fostering children's curiosity that we need to explore further.

Finaly, the third study investigates the role of intrinsic motivation in spatial learning in late adulthood [25]. We investigated age differences in memory for spatial routes that were either actively (i.e., intinsic

motivation condition) or passively (i.e., control condition) encoded. A series of virtual environments were created and presented to 20 younger (Mean age = 19.71) and 20 older (Mean age = 74.55) adults, through a cardboard viewer. During encoding, participants explored routes presented within city, park, and mall virtual environments, and were later asked to re-trace their travelled routes. Critically, participants encoded half the virtual environments by passively viewing a guided tour along a pre-selected route, and half through active exploration with volitional control of their movements by using a button press on the viewer. During retrieval, participants were placed in the same starting location and asked to retrace the previously traveled route. We calculated the percentage overlap in the paths travelled at encoding and retrieval, as an indicator of spatial memory accuracy, and examined various measures indexing individual differences in their cognitive approach and visuo-spatial processing abilities. Results showed that active navigation, compared to passive viewing during encoding, resulted in a higher accuracy in spatial memory, with the magnitude of this memory enhancement being significantly larger in older than in younger adults. Results suggest that age-related deficits in spatial memory can be reduced by active encoding. In other words, this means that conditions where intrinsic motivation is involved, reduce negative effects of aging on spatial learning.

7.6.3. Poppy Education: Designing and Evaluating Educational Robotics Kits

Participants: Pierre-Yves Oudeyer, Didier Roy [correspondant], Thibault Desprez.

The Poppy Education project aims to create, evaluate and disseminate all-inclusive pedagogical kits, open-source and low cost, for teaching computer science and robotics in secondary education and higher education, scientific literacy centers and Fablabs.

It is designed to help young people to take ownership with concepts and technologies of the digital world, and provide the tools they need to allow them to become actors of this world, with a considerable socio-economic potential. It is carried out in collaboration with teachers and several official french structures (French National Education, High schools, engineering schools, ...).

Poppy Education is based on the robotic platform poppy (open-source platform for the creation, use and sharing of interactive 3D printed robots), including:

- web interface connection (see figure 31)

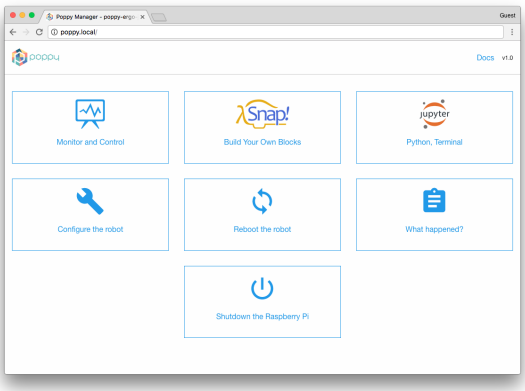


Figure 31. Home page on http://poppy.local

- Poppy Humanoid, a robust and complete robotics platform designed for genuine experiments in the real world and that can be adapted to specific user needs.
- Poppy Torso, a variant of Poppy Humanoid that can be easily installed on any flat support.

- Ergo Jr, a robotic arm. Durable and inexpensive, it is perfect to be used in class. It can be programmed in Python, directly from a web browser, using Ipython notebooks (an interactive terminal, in a web interface for the Python Programming Language).
- Snap. The visual programming system Snap (see figure 32), which is a variant of Scratch. Its features allow a thorough introduction of information technology. Several specific "blocks" have been developed for this.

Figure 32. The visual programming system Snap

- C++, Java, Matlab, Ruby, Javascript, etc. thanks to a REST API that allows you to send commands and receive information from the robot with simple HTTP requests.
- Virtual robots (Poppy Humanoid, Torso and Ergo) can be simulated with the free simulator V-REP (see figure 33). It is possible in the classroom to work on the simulated model and then allow students to run their program on the physical robot.
- Virtual robots (Poppy Ergo) can also be simulated with a 3D web viewer (see figure 34).

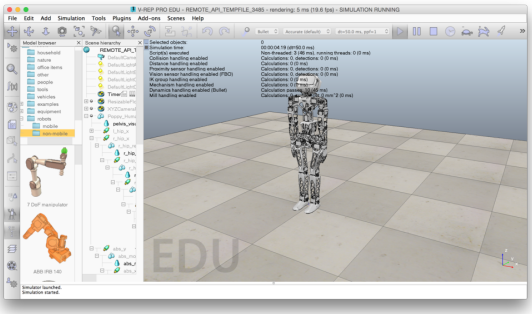


Figure 33. V-rep

7.6.3.1. Pedagogical experimentations : Design and experiment robots and the pedagogical activities in classroom.

The robots are designed with the final users in mind. The pedagogical tools of the project (robots and resources) are being created directly with the users and evaluated in real life by experiments. So teachers and researchers co-create activities, test them with students in class-room, share their experience and develop the platform as needed [126].

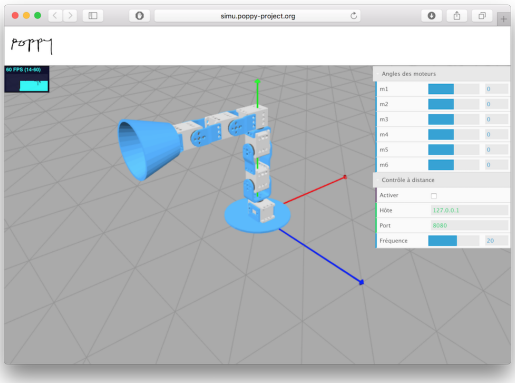


Figure 34. 3D viewer

The activities were designed mainly with Snap! and Python. Most activities use Poppy Ergo Jr, but some use Poppy Torso (mostly in higher school due to its cost).
The pedagogical experiments in classroom carried out during the first year of the project notably allowed to create and experiment many robotic activities. These activities are designed as pedagogical resources introducing robotics. The main objective of the second year was to make all the activities and resources reusable (with description, documentation and illustration) easily and accessible while continuing the experiments and the diffusion of the robotic kits.

Figure 35. Experiment robots and pedagogical activities in classroom

- Pedagogical working group : the teacher partners continued to use the robots in the classroom and to create and test new classroom activities. We organized some training to help them to discover and learn how to use the robotics platform. Also, an engineer of the Poppy Education team went to visit the teachers in their school to see and to evaluate the pedagogical tools (robots and activities) in a real context of use.

Five meetings have been organized during the year including all teachers part of the project as well as the Poppy Education team in order to exchange about their experience using the robots as a pedagogical tool, to understand their need and to get some feedback from them. This is helping us to understand better the educational needs, to create and improve the pedagogical tools.
You can see the videos of pedagogical robotics activities here:
https://www.youtube.com/playlist?list=PLdX8RO6QsgB7hM_7SQNLvyp2QjDAkkzLn

7.6.3.2. Pedagogical documents and resources

- We continued to improve the documentation of the robotic platform Poppy (https://docs.poppy-project.org/en/) and the documentation has been translated into French (https://docs.poppy-project.org/fr/.
  We configured a professional platform to manage the translation of the documentation (https://crowdin.com/project/poppy-docs. This platforms allows anybody to participate in the translation of the documentation to the language of their choice.
- To complete the pedagogical booklet [125] that provides guided activities and small challenges to become familiar with Poppy Ergo Jr robot and the Programming language Snap! (https://hal.inria.fr/hal-01384649/document) we provided a list of Education projects. Educational projects have been written for each activity carried out and tested in class. Each project has its own web page including resources allowing any teacher to carry out the activity (description, pedagogical sheet, photos / videos, pupil's sheet, teacher's sheet with correction etc.).
  The activities are available here:
  https://www.poppy-education.org/activites/activites-lycee
  The pedagogical activities are also available on the Poppy project forum where everyone is invited to comment and create new ones:
  https://forum.poppy-project.org/t/liste-dactivites-pedagogiques-avec-les-robots-poppy/2305

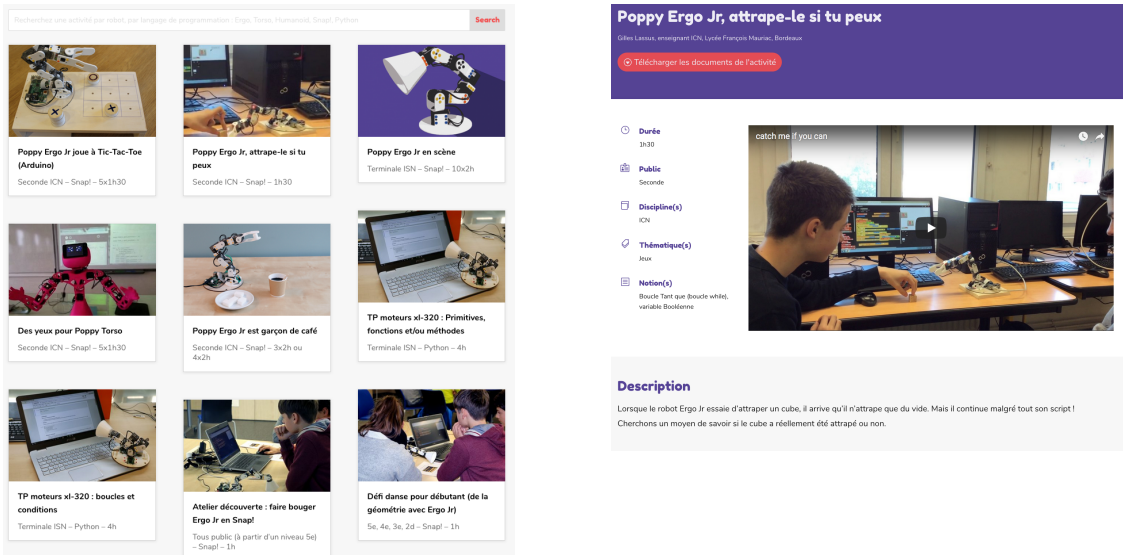


Figure 36. Open-source educational activities with Poppy robots are available on Poppy-Education.org

- A FAQ have been written with the most frequents questions to help the users: https://www.poppy-education.org/aide/
- A website has been created to present the project and to share all resources and activities.
  https://www.poppy-education.org/

7.6.3.3. Evaluation of the pedagogical kits

The impact of educational tools created in the lab and experimented in class had to be evaluated qualitatively and quantitatively. First, the usability, efficiency and user satisfaction must be evaluated. We must therefore assess, at first, if these tools offer good usability (i.e. effectiveness, efficiency, satisfaction). Then, in a second step, select items that can be influenced by the use of these tools. For example, students' representations of robotics, their motivation to perform this type of activity, or the evolution of their skills in these areas. In 2017 we conducted experiments to evaluate the usability of kits. We also collected data on students' perceptions of robotics.

- Population

  Our sample is made up of 28 teachers and 146 students from the region Nouvelle Aquitaine. Each subject completed an online survey in June 2017. Here, we study several groups of individuals: teachers and students. Among the students we are interested in those who practiced classroom activities with the Ergo Jr kit during the school year 2016 - 2017 (N = 68) (age = 16, std = 2.44). Among these students, 37 where High School students following the "Computer Science and Digital Sciences" stream (BAC S option ISN), 12 followed the stream "Computer and Digital Creation" (BAC S option ICN) and 18 where in Middle School.

  Among the 68 students, 13 declared having used the educational booklet provided in the kit and 16 declared having used other robotic kits. Concerning the time resource dedicated to activities with the robot, 30 students declared having spent less than 6 hours, 22 declared between 6 and 25 hours, and 16 declared having spent more than 25 hours.

  have practiced less than 6 hours of activity with the robot (N = 30), between 6 and 25 hours (N = 22) or more than 25 hours (N = 16); having built the robot (N = 12); have used the visual programming language Snap! (N = 46), the language of Python textual programming (N = 21), both (N = 8) or none (N = 9), it should be noted that these two languages are directly accessible via the main interface of the robot.

- Evaluation of the tool

  We have selected two standardized surveys dealing with this issue: SUS (The Systeme Usability Scales) [62] and The AttrakDiff [100]. These two surveys are complementary and allow to identify the design problems and to account for the perception of the user during the activities. The results of these surveys are available in the article (in French) [76] published at the conference Didapro (Lausanne Feb, 2018). Figures 37 and 38 show the averages of the 96 respondents (68 students + 28 teachers) for each of the 10 statements from the SUS and 28 pairs of antonyms to be scored on a scale of 1 to 5 and a 7-point scale, respectively.

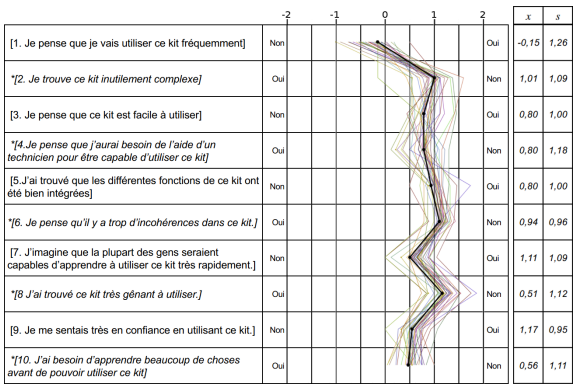


Figure 37. Result of SUS survey

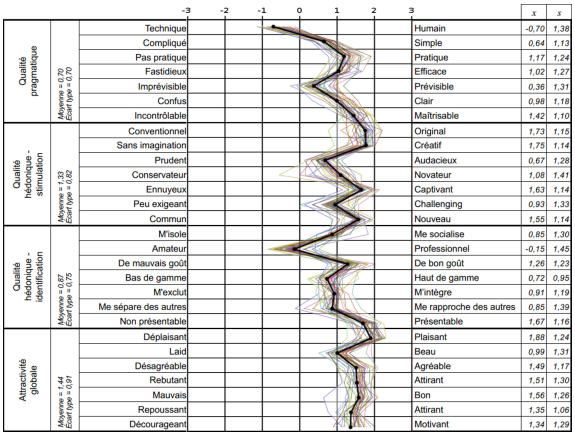


Figure 38. Result of AttrakDiff survey

- Evaluation of impact on learner

  One of the objectives of the integration of digital sciences in school is to allow students to have a better understanding of the technological tools that surround them daily (i.e. web, data, algorithm, connected object, etc.). So, we wanted to measure how the practice of activities with ErgoJr robot had changed this apprehension; especially towards robots. For that, we used a standardized survey: "attitude towards robot" *EuroBarometer 382* originally distributed in 2012 to more than 1000 people in each country of the European Union. On the one hand, we sought to establish whether there had been a change in response between 2012 and 2017, and secondly whether there was an impact on the responses of 2017 according to the participation, or not, in educational activities with ErgoJr robot. The analysis of the results is in progress and will be published in 2019.

- Web page for the experimentations

  To facilitate the storage of documents, their availability, and to highlight some information and news, a page dedicated to the experimentations is now available on the website. https://www.poppy-education.org/evaluation/

7.6.3.4. Partnership on education projects

- Ensam

  The Arts and Métiers campus at Bordeaux-Talence in partnership with Inria wishes to contribute to its educational and scientific expertise to the development of new teaching methods and tools. The objective is to develop teaching sequences based on a project approach, relying on an attractive multidisciplinary technological system: the humanoid Inria Poppy robot.

  The humanoid Inria Poppy robot offers an open platform capable of providing an unifying thread for the different subjects covered during the 3-years of the Bachelor training: mechanics, manufacturing (3D printing), electrical, mecha-tronics, computer sciences, design.

- Poppy entre dans la danse (Poppy enters the dance)

  The project "Poppy enters the dance" (Canope 33) took place for the second year. It uses the humanoid robot Poppy. This robot is able to move and experience the dance. The purpose of this project is to allow children to understand the interactions between science and choreography, to play with the random and programmable, to experience movement in dialogue with the machine. At the beginning of the project they attended two days of training on the humanoid robot (Inria -

Poppy Education). During the project, they met the choreographer Eric Minh Cuong Castaing and the engineer Segonds Theo (Inria - Poppy Education).

You can see a description and an overview of the project here:
https://www.youtube.com/watch?v=XfxXaq899kY

- DANE

  The Academic Delegation for Digital Educational is in charge of supporting the development of digital uses for pedagogy. It implements the educational digital policy of the academy in partnership with local authorities. She accompanies institutions daily, encourages innovations and participates in their dissemination.

- RobotCup Junior

  RoboCupJunior OnStage invites teams to develop a creative stage performance using autonomous robots that they have designed, built and programmed. The objective is to create a robotic performance between 1 to 2 minutes that uses technology to engage an audience. The challenge is intended to be open-ended. This includes a whole range of possible performances, for example dance, storytelling, theatre or an art installation. The performance may involve music but this is optional. Teams are encouraged to be as creative, innovative and entertaining, in both the design of the robots and in the design of the overall performance.

7.7. Other applications

7.7.1. Applications in Robotic myoelectric prostheses

Participants: Pierre-Yves Oudeyer [correspondant], Aymar de Rugy, Daniel Cattaert, Mick Sebastien.

Together with the Hybrid team at INCIA, CNRS (Sébastien Mick, Daniel Cattaert, Florent Paclet, Aymar de Rugy) and Pollen Robotics (Matthieu Lapeyre, Pierre Rouanet), the Flowers team continued to work on a project related to the design and study of myoelectric robotic prosthesis. The ultimate goal of this project is to enable an amputee to produce natural movements with a robotic prosthetic arm (open-source, cheap, easily reconfigurable, and that can learn the particularities/preferences of each user). This will be achieved by 1) using the natural mapping between neural (muscle) activity and limb movements in healthy users, 2) developing a low-cost, modular robotic prosthetic arm and 3) enabling the user and the prosthesis to co-adapt to each other, using machine learning and error signals from the brain, with incremental learning algorithms inspired from the field of developmental and human-robot interaction.

7.7.1.1. Reachy, a 3D-printed Human-like Robotic Arm as a Test Bed for Prosthesis Control Strategies

To this day, despite the increasing motor capability of robotic prostheses, elaborating efficient control strategies is still a key challenge for their design. To provide an amputee with efficient ways to drive a prosthesis, this task requires thorough testing prior to integration into finished products. To preserve consistency with prosthetic applications, employing an actual robot for such testing requires it to show human-like features. To fulfill this need for a biomimetic test platform, we developed the Reachy robotic platform, a seven-joint human-like robotic arm that can emulate a prosthesis. Although it does not include an articulated hand and is therefore more suitable for studying reaching than manipulation, a robotic hand from available research prototypes could be integrated to Reachy. Its 3D-printed structure and off-the-shelf actuators make it inexpensive relatively to the price of a genuine prosthesis. Using an open-source architecture, its design makes it broadly connectable and customizable, so it can be integrated into many applications. To illustrate how Reachy can connect to external devices, we developed several proofs of concept where it is operated with various control strategies, such as tele- operation or vision-driven control. In this way, Reachy can help researchers to develop and test innovative control strategies on a human-like robot.

7.7.2. Ship Motion estimation from sea wave vision

Participants: David Filliat [correspondant], Natalia Díaz Rodríguez, Zhi Zhou, Manuel Cortés-Batet, Nazar-Mykola Kaminskyi.

Together with Naval Group, ENSTA Paris has been working on a set of software tools for simulating sea waves and motion estimation from images. The objective is predicting variables of interest in order to compensate the position and inclination of large boats at deep sea, seconds ahead of time to preserve stability. Work being currently done in partnership with Abo Akademi University (Turku, Finland) will validate the soon to be published Blender wave generator and machine learning algorithms, with real data gathered from the Baltic Sea archipelago.

HEPHAISTOS Project-Team

6. New Results

6.1. Robotics

6.1.1. Analysis of Cable-driven parallel robots

Participants: Jean-Pierre Merlet [correspondant], Yves Papegay.

We have continued the analysis of suspended CDPRs for control and design purposes. This analysis is heavily dependent on the behavior of the cable. Three main models can be used: *ideal* (no deformation of the cable due to the tension, the cable shape is a straight line between the attachments points), *elastic* (cable length changes according to the tension to which it is submitted, straight line cable shape) and *sagging* (cable shape is not a line as the cable is submitted to its own mass). The different models leads to very different analysis with a complexity increasing from ideal to sagging. All cables exhibit sagging but the sagging effect is often neglected if the CDPR is relatively small while it definitively cannot be neglected for large CDPRs. The most used sagging model is the Irvine model [24]. This is a non algebraic planar model with the upper attachment point of the cable is supposed to be grounded: it provides the coordinates of the lowest attachment point B of the cable if the cable length L_0 at rest and the force applied at this point are known. It takes into account both the elasticity and deformation of the cable due to its own mass. A drawback of this model is that we will be more interested in a closed-form of the L_0 for a given pose of B (for the inverse kinematics of CDPR) and in alternate form of the model that will provide constraint on the force components (for the direct kinematics). We have proposed new original formulations of the Irvine model in [13] and have shown that their use drastically improve the solving time for both the inverse and direct kinematics (i.e finding all possible solutions for both problems) that are required for CDPRs control. Still the solving time of the direct kinematics is too large for the real-time direct kinematics and in that case only the current pose of the platform is of interest.

The direct kinematics relies on an accurate estimation of the cable lengths that is usually based on the measurement of the winch drum rotation. We have evaluated the influence of uncertainties in the cable length measurement on the result of the FK [19] and have shown that for a poor robot geometry (which was for example the case for the prototype described in section 6.1.2 for which the geometry was imposed) this influence may be quite large . An usual strategy to decrease this uncertainty for small to medium-sized CDPR is to use a drum with a cable spiral guide for the coiling which impose a coiling path for the cable. However this strategy is unfeasible for large and very large CDPR (that we called *Ultrabot*) for which the large length of the cables impose to have several layers on the drum and therefore leads to a more erratic coiling process that leads to possibly large errors of the cable lengths estimation. To get a better estimation of the cable lengths we have proposed an original method, based on the Vernier principle [21]. The idea is to have several small colored marks on the cable at known distances from the end-point of the cable and to have several color sensors in the mast of the CDPR. We have first shown that if 3 colors (e.g. RGB) were used, then an appropriate disposition of the marks on the cable allows to have up to 29 marks on the cable so that the sequence of 3 successive colors is always unique. Hence by coiling the cable and detecting the 3 successive color detected by a sensor allows to determine exactly the distance between the sensor and the cable end-point, i.e. to *calibrate* the cable length. Calibration is always an issue for CDPR which uses usually incremental encoders for measuring the drum rotation (which explain why we have also proposed another approach [18]. Then we have considered the sequence of color detection when coiling the cable, starting from its largest length. We have looked at the distribution of cable length changes $\Delta\rho$ between two successive detection and have proposed a strategy that provide the distance between the marks so that this distribution is quasi-uniform with a mean value that is minimal. For example we have shown that for a 60 meters length cable having 29 marks we were able to have an almost constant $\Delta\rho$ of 40 cm, meaning that when the cable length changes by this value, then we get an exact evaluation of the cable length at each detection. In between such detection we rely on the drum rotation measurement to estimate the cable length. Furthermore we have shown that the difference between the

expected detection time and the real one allows one to update the estimate of the drum radius, thus enabling to manage an erratic coiling process. We have initially installed this system on the prototype presented in section 6.1.2 . The few initial tests were really promising but on-site we have had problems for ensuring a constant positioning of the marks on the synthetic cables. Being given the very short deployment time we have not been able to fix this problem. Consequently we have decided to use another approach based on direct measurement of the load pose with lidars, this approach being described in section 6.1.2 .

We have also continued to investigate the calculation of planar cross-sections of the workspace for CDPR with sagging cables. We have shown in a previous paper that the border of this workspace was either determined by cable length limits but also by the singularity of the kinematics equations. Hence these singularities play an important role for the design of a CDPR. We have started a preliminary investigation on this topic [20]. We have shown that these singularities may be classified in two categories:

- *classical singularity* which corresponds to the singularity of parallel robots with rigid legs which basically implies that the mechanical equilibrium of the system cannot be obtained, leading to a motion of the platform even if the actuators are locked
- *full singularity* which are singularity of the kinematics equations but are not classical singularity. In this case mechanical equilibrium is obtained but the CDPR is unable to move in a given direction

We have also developed an algorithm that check if a full singularity exists in the neighborhood of a given pose and to locate it with an arbitrary accuracy.

6.1.2. Cable-Driven Parallel Robots for large scale additive manufacturing

Participants: Jean-Pierre Merlet, Yves Papegay [correspondant].

Easy to deploy and to reconfigure, dynamically efficient in large workspaces even with payloads, cable-driven parallel robots are very attractive for solving displacement and positioning problems in architectural building at large scale seems to be a good alternative to crane and industrial manipulators in the area of additive manufacturing. We have co-founded in 2015 years ago the XtreeE (www.xtreee.eu) start-up company that is currently one of the leading international actors in large-scale 3D concrete printing.

We have been contacted in 2018 by the artist Anne-Valérie Gasc that is interested in mimicking the 3D additive manufacturing process on large scale for a live art performance. She was interested in a mean for widespreading glass micro-beads on a given trajectory over a 21 $\times$ 9m large platform located at the contemporary art center *Les Tanneries* (figure 1), located close to Montargis. She was especially interested in using a CDPR for that purpose because of the low visual intrusivity of the cables and its ability to move large load. After a few month of discussions we agree to recycle our old MARIONET-CRANE prototype (2009) for this exhibition although the place was not the most appropriate for the CDPR as the height of the location was only 3 meters. We design as load a 80 liters drum of weight 55 kg with 40 kg of powder that was sufficient for printing one trajectory (figure 1). An on-board computer connected through wifi to a master computer was managing the lidar measurement and the opening/closing of the servo-valve controlling the powder flow. The drum was supported by 4 Dyneema cables of diameter 3mm whose output points were located at the corners of the platform and whose lengths were varying between 3 and 26 meters. The master computer was controlling the CDPR and the parameters of the system were recorded every second in log files. The development was very fast and we were not able to test a full scale installation in our laboratory for lack of the appropriate space. The on-site deployment was difficult because it has to be done in a record time, far away from our home base. The lack of height has especially a strong influence on the positioning errors of the drum that drastically increase if the cables are close to the horizontal. We solve on-site this problem by adding 3 low-cost lidars that were providing partial measurement on the drum pose. The system was fully operational a few days after the official opening of the exhibition and was at the heart of the artistic exhibition "Les Larmes du Prince - Vitrifications" (http://www.lestanneries.fr/exposition/larmes-prince-vitrifications), that was run during July and August under the control of a local student. The exhibition was scheduled to run 5 days per week until the end of August. During this period the CDPR has worked 174 hours (4h15mn/day), has traveled 4757 meters and has dispersed about 1.5 tons of powder. We get two failures: one of the cables has broken but without any consequence because of the redundancy of the robot and a failure of the reduction gear of one of the winch on

the exhibition closing day, which has been immediately repaired. From a scientific viewpoint we have been able to test, in this quasi-industrial context, the efficiency of a control law using external measurements of the pose and the logs, still being processed has allowed us to identify possible improvements and scientific issues regarding the modeling of the system. An unexpected benefit of using the lidars was to allow to record a profile of the powder wall at each trajectory, showing its life over time as it was always evolving because of the powder particle motion after a printing.

Figure 1. The exhibition place and the drum. Photos copyrighted Anne-Valérie Gasc, "Vitrifications", Photograph: Aurélien Mole

6.1.3. Killing robots

Participant: Jean-Pierre Merlet [correspondant].

The director, Linda Blanchet, of a theater company has contacted us for helping organizing a theater event, *Killing Robots*, centered on the story of *Hitchbot*, a passive 70cm high mannequin designed by Canadian colleagues, that was put on the side-way of roads in Canada so that people may transport it, the purpose being to study the human interaction with people during a travel from the east to the west cost of Canada. The mannequin was located through a GPS and has taken a picture of its surrounding every 20 minutes while it was active. This mannequin indeed performs this travel in 15 days and a similar experiment was then scheduled in the US, the purpose being to go from Boston to San Francisco. Unfortunately after 5 days of travel the mannequin was discovered completely dismantled in Philadelphia. The idea of Linda Blanchet's performance was to propose a thriller based on the robot data for discovering who has dismantled the robot and in parallel to have the robot interacts with the actors to describe its feeling. For that purpose it was necessary that the robot becomes actuated while keeping its appearance identical to the original model. We have therefore retrieved a clone of the original Hitchbot and we have actuated the arms and head, so that the robot was able to move them, adding a lidar on top of the head so that it was able to locate the actors on stage (figure 2).

The Canadian colleague have also provided a conversational agent so that the robot was able to speak with a learning process. The opening of the performance was done on November 6 at the National theater of Nice and it is now performing in various places in France. We have been present at several of them to interact with the public at the end of the performance. From a scientific viewpoint our interest in this exhibition was to better understand why adding motion to a mannequin modify drastically the perception of the robot by the public. These understanding will help to work on the factors that increase the acceptance of a technological object by the public, which is clearly a major factor for the efficiency of our assistance devices.

Figure 2. The transformed Hitchbot robot

6.2. Smart Environment for Human Behaviour Recognition

Participants: Jean-Pierre Merlet, Yves Papegay, Odile Pourtallier [correspondant], Eric Wajnberg.

The general aim of this research activity focuses on long term indoor monitoring of frail persons. In particular we are interested in early detection of daily routine and activity modifications. These modifications may indicate health condition alteration of the person and may require further medical or family care. Note that t our work does not aim at detecting brutal modifications such as faintness or fall.

In our research we envisage both individual and collective housing such as rehabilitation center or retirement home.

Our work relies on the following leading ideas :

- We do not base our monitoring system on wearable devices since it appears that they may not be well accepted and worn regularly,
- Privacy advocates adequacy between the monitoring level needed by a person and the detail level of the data collected. We therefore strive to design a system fitted to the need of monitoring of the person.
- In addition to privacy concern, intrusive feature of video led us not to use it.

The main aspect that grounds this work is the ability to locate a person or a group in their indoor environment. We focus our attention to the case where several persons are present in the environment. As a matter of fact the single person case is less difficult.

6.2.1. Tools and data analysis for experimental systems

Two experimental systems are installed in two areas (a consultation center (Institut Claude Pompidou, ICP, Nice), and a retirement home (EHPAD Valrose, Nice)) were several types of persons (residents, visitors, staff) evolve. They are made up of virtual barriers (constituted of distance and motion sensors) displayed in the environment and connected to a PC that collects and stores the measurements of the barriers. Each crossing

of a barriers hence corresponds to a specific signal of a set of sensors. We develop a set of codes that aim to analyze the data collected to construct information on the moves of the persons in the experiment areas [23].

This year we have improved the code that yields the barrier events (time and direction of crossing of barriers) from the raw data. This allowed us to use this first step to reconstruct the individual trajectories of the users.

Although the filtering technics do not use external information (such as specific use of a zone bounded by barriers, habit of users according to time....) we can determine most of the individual trajectories of the users, even when several users evolve simultaneously in the area. Although some uncertainties remain (and could probably be improved using external knowledge), we can use the results obtained to perform a statistical analysis.

The aim on the main scientific efforts this year was to develop a detailed statistical treatment chain to extract and to visualize the events information coming from the set of movement activity detectors installed at ICP. All the (statistical and graphical) development were performed in the R software environment. Globally, two sets of information were collected, for the recorded data. The first provides a kinematic view of the presence of individuals on the mass plan of ICP during a chosen time interval. The following graph gives a static example of the kinematic graph obtained. Such a dynamic information points, for example, to specific movement activities in the medical center, at given time intervals. Figure 3 shows the presence of individuals in the corridors and consultation rooms at ICP at different times.

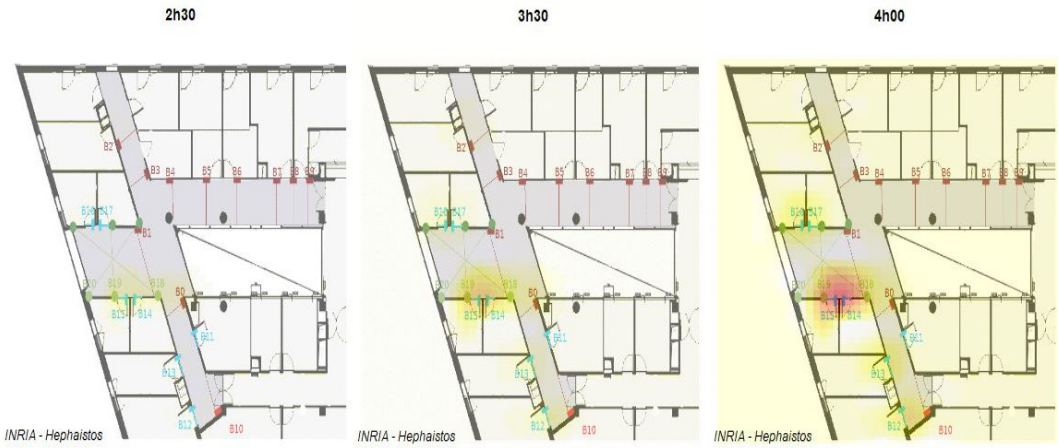


Figure 3. Three photograms on a kinematic view of the presence of individuals on the mass plan of ICP

Such a graph is only descriptive. Hence, it does not provide a functional analysis of the displacements of individuals in the medical center. In order to understand this better, the chronological movement patterns were functionally described by building, for every time interval, the transition matrix between all zones present in the analyzed medical center. After proper algebraic manipulation, the obtained transition matrices were analyzed using a factorial correspondence analysis, a multivariate method that - in this case and among other features – built graphs describing the functional movement patterns between zones. The graph presented in figure 4 gives an example of the obtained results.

The next step will be to statistically compare such results, e.g., between morning or afternoon activity, between days with or without medical consultation, etc. Results obtained might lead to a better organization of the medical activities at ICP.

6.3. Other medical activities

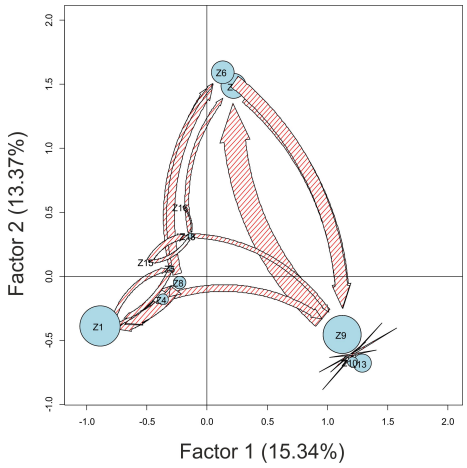


Figure 4. Example of the best factorial plan (explaining almost 30% of all the information contained in the data) obtained from a factorial correspondence analysis used to describe the functional movement patterns of individual between zones in the followed medical center during a full day of activity. Each blue circle represents a zone, with a radius proportional of its frequentation frequency. Arrows between zones (in red) are proportional to the observed flux of individual movements between zones. Only the most important arrows are presented.

Participants: Jean-Pierre Merlet [correspondant], Sylvain Guénon.

Eric Sejor, a surgeon at Nice hospital, has contacted us about developing a robotized system for realizing sutures in an autonomous way. Suturing is a lengthy process while in many cases this is not a complex operation. Eric Sejor mentions that developing an autonomous system allowing to manage standard wounds may be extremely interesting, especially for emergency service that are under-staffed. Instead of developing a new robot dedicated to this purpose we have proposed to Eric Sejor to build a system based on the existing manual tools that require to put the instrument in place and then simply squeezing a trigger. The placement will be realized by one of our small parallel robot, with the help of vision system to locate the edge of the wound, while the trigger squeezing will be performed by an actuator. We have obtained an Idex funding (one year for an engineer) to develop a proof-of-concept prototype that will perform the operation on silicone mockups that are used for the surgeon training.

We have had also a contact with the ergotherapist Nicolas Ciai from Nice hospital for the evaluation of patient motricity before an operation. For this evaluation the ergotherapist performs muscular testing before the operation, right after the operation and 6 months later. The exercise consists in opposing the ergotherapist palm against the musculo group that has to be tested until a force equilibrium is reached. Then the ergotherapist ranks the tonicity of the muscles on a discrete scale between 0 and 6 according to his muscular feeling. As numerous muscles have to be tested, the process is quite lengthy. Clearly this process is quite subjective and we have proposed an objectification of the process by developing a glove prototype that includes pression sensors for measuring accurately the pressure exerted by the patient. These sensors are used by a micro-computer the size of a large watch located on the wrist of the ergotherapist. This computer determines when the pressure becomes stable, in which case this pressure is displayed and recorded. A companion software will then exploit the recorded data to provide an evaluation report. Beside the objectification of the ranking, the purpose is also to speed-up the tests. Although this project is quite advanced, we are lacking of manpower to complete it so that we have presented a project to Nice hospital for funding an engineer that may complete the second version of the glove.

LARSEN Project-Team

7. New Results

7.1. Lifelong autonomy

7.1.1. Motion planning for robot audition

Participants: François Charpillet, Francis Colas, Van Quan Nguyen.

We collaborated on this subject with Emmanuel Vincent from the Multispeech team (Inria Nancy – Grand Est).

Robot audition refers to a range of hearing capabilities which help robots explore and understand their environment. Among them, sound source localization is the problem of estimating the location of a sound source given measurements of its angle of arrival with respect to a microphone array mounted on the robot. In addition, robot motion can help quickly solve the front-back ambiguity existing in a linear microphone array. In this work, we focus on the problem of exploiting robot motion to improve the estimation of the location of an intermittent and possibly moving source in a noisy and reverberant environment. We first propose a robust extended mixture Kalman filtering framework for jointly estimating the source location and its activity over time. Building on this framework, we then propose a long-term robot motion planning algorithm based on Monte Carlo tree search to find an optimal robot trajectory according to two alternative criteria: the Shannon entropy or the standard deviation of the estimated belief on the source location. Experimental results show the robustness of the proposed estimation framework to false angle of arrival measurements within $\pm 20°$ and 10% false source activity detection rate. The proposed robot motion planning technique achieves an average localization error 48.7% smaller than a one-step-ahead method.

Publication: [10]

7.1.2. Addressing Active Sensing Problems through Monte-Carlo Tree Search (MCTS)

Participants: Vincent Thomas, Gabriel Belouze, Sylvain Geiser, Olivier Buffet.

The problem of active sensing is of paramount interest for building self awareness in robotic systems. It consists in planning actions in a view to gather information (*e.g.*, measured through the entropy over certain state variables) in an optimal way. In the past, we have proposed an original formalism, ρ-POMDPs, and new algorithms for representing and solving such active sensing problems [24] by using point-based algorithms, assuming either convex or Lipschitz-continuous criteria. More recently, we have developed new approaches based on Monte-Carlo Tree Search (MCTS), and in particular Partially Observable Monte-Carlo Planning (POMCP), which provably converge only assuming the continuity of the criterion. We are now going towards algorithms more suitable to certain robotic tasks by allowing for continuous state and observation spaces.

Publication: [20]

7.1.3. Heuristic Search for (Partially Observable) Stochastic Games

Participants: Olivier Buffet, Vincent Thomas.

Collaboration with Jilles Dibangoye (INSA-Lyon, Inria team CHROMA) and Abdallah Saffidine (University of New South Wales (UNSW), Sydney, Australia).

Many robotic scenarios involve multiple interacting agents, robots or humans, *e.g.*, security robots in public areas. We have mainly worked in the past on the collaborative setting, all agents sharing one objective, in particular through solving Dec-POMDPs by (i) turning them into occupancy MDPs and (ii) using heuristic search techniques and value function approximation [2]. A key idea is to take the point of view of a central planner and reason on a sufficient statistic called *occupancy state*. We are now working on applying similar approaches in the important 2-player zero-sum setting, *i.e.*, with two competing agents. As a preliminary step, we have proposed and evaluated an algorithm for (fully observable) stochastic games, which does not require any problem transformation. Then we have proposed an algorithm for partially observable stochastic games, here turning the problem into an occupancy Markov game.

[This line of research will be pursued through Jilles Dibangoye's ANR JCJC PLASMA.]

7.1.4. Interpretable Action Policies

Participant: Olivier Buffet.

Collaboration with Iadine Chadès and Jonathan Ferrer Mestres (CSIRO, Brisbane, Australia), and Thomas G. Dietterich (Oregon State University, USA).

Computer-aided task planning requires providing user-friendly plans, in particular, plans that make sense to the user. In probabilistic planning (in the MDP formalism), such interpretable plans can be derived by constraining action policies (if X happens, do Y) to depend on a reduced subset of (abstract) states or state variables. We have (i) formalized the problem of finding a set of at most K abstract states (forming a partition of the original state space) such that any optimal policy of the induced abstract MDP is as close as possible to optimal policies of the original MDP, and (ii) proposed 3 solution algorithms with theoretical and empirical evaluations.

7.1.5. Perspective: hierarchical quality diversity, from materials to machines

Participant: Jean-Baptiste Mouret.

Collaboration with CSIRO (Australia) and Vrije Universiteit Amsterdam (Netherlands).

Natural lifeforms specialize to their environmental niches across many levels, from low-level features such as DNA and proteins, through to higher-level artefacts including eyes, limbs and overarching body plans. We propose 'multi-level evolution', a bottom-up automatic process that designs robots across multiple levels and niches them to tasks and environmental conditions. Multi-level evolution concurrently explores constituent molecular and material building blocks, as well as their possible assemblies into specialized morphological and sensorimotor configurations. Multi-level evolution provides a route to fully harness a recent explosion in available candidate materials and ongoing advances in rapid manufacturing processes. We outline a feasible architecture that realizes this vision, highlight the main roadblocks and how they may be overcome, and show robotic applications to which multi-level evolution is particularly suited. By forming a research agenda to stimulate discussion between researchers in related fields, we hope to inspire the pursuit of multi-level robotic design all the way from material to machine.

Publication: [5]

7.1.6. Improving Embodied Evolutionary Robotics

Participant: Amine Boumaza.

Multi-robots learning is a hard still unsolved problem. When framed into the machine learning theoretical setting, it suffers from a high complexity when seeking optimal solutions. On the other hand, when sub-optimal solutions are acceptable Embodied Evolutionary Robotics, can provide solutions that perform well in practice. Improving these algorithms in terms of run-time or solution quality is an important research question.

It has been long known from the theoretical work on evolution strategies, that recombination improves convergence towards better solution and improves robustness against selection error in noisy environment. We propose to investigate the effect of recombination in online embodied evolutionary robotics, where evolution is decentralized on a swarm of agents. We hypothesize that these properties can also be observed in these algorithms and thus could improve their performance. We introduce the $(\mu/\mu, 1)$-On-line Embedded Evolutionary Algorithm (EEA) which uses a recombination operator inspired from evolution strategies and apply it to learn three different collective robotics tasks, locomotion, item collection and item foraging. Different recombination operators are investigated and compared against a purely mutative version of the algorithm. The experiments show that, when correctly designed, recombination improves significantly the adaptation of the swarm in all scenarios.

Publication: [13] [12]

7.1.7. Multi-robot exploration of an unknown environment

Participants: Nicolas Gauville, François Charpillet.

Different approaches exist for multi-robot autonomous exploration. These include frontier approaches, where robots are assigned to unexplored areas of the map, which provide good performance but require sharing the map and centralizing decision-making. The Brick and Mortar approaches, on the other hand, use a ground marking with local decision-making, but give much lower performance. The algorithm developped by Nicolas Gauville during his pre-thesis period is a trade-off between these two approaches, allowing local decision-making and, surprisingly, performances are closed to centralized frontier approaches. We also propose a comparative study of the performance of the three different approaches : *Brick & Mortar*, *Global Frontiers* and *Local Frontiers*. Our local algorithm is also complete for the exploration problem and can be easily distributed on robots with a minor loss of performance. This work follows the *Cart-O-Matic* project in which our team participated, which aimed to explore and map a building while recognizing specific objects inside with a team of 5 mobile robots.

Publication: [16]

7.2. Natural Interaction with Robotics Systems

Thanks to the arrival of Pauline Maurice and the AnDy H2020 project, our activities about interaction are currently focused on ergnonomic interaction, which requires good foundations in motion analysis.

7.2.1. Digital human modeling for collaborative robotics

Participant: Pauline Maurice.

Collaboration with Vincent Padois (Inria Bordeaux and Sorbonne Université), Yvan Measson (CEA-LIST) and Philippe Bidaud (ONERA and Sorbonne Université).

Work-related musculoskeletal disorders in industry represent a major and growing health problem in many developed countries. Collaborative robotics, which allows the joint manipulation of objects by both a robot and a person, is a possible solution provided that it is possible to assess the ergonomic benefit they offer. Using a digital human model (DHM) can cut down the development cost and time by replacing the physical mock-up by a virtual one easier to modify. The first part of this work details the challenges of digital ergonomic assessment for collaborative robotics. State-of-the-art work on DHM simulations with collaborative robots is reviewed to identify which questions currently remain open. The second part of this work focuses on a specific use case and presents a DHM-based method to optimize design parameters of a collaborative robot for an industrial task.

Publication: [21]

7.2.2. Probabilistic decision making for collaborative robotics

Participants: Yang You, Vincent Thomas, Olivier Buffet, François Charpillet, Francis Colas.

Collaboration with Rachid Alami (LAAS, France).

This work is part of the ANR Flying Co-Worker project and focuses on high-level decision making for collaborative robotics. When a robot has to assist a human worker, it does not have direct access to his current intention or his preferences but has to adapt its behaviour to help the human completing his task. To achieve this, we followed what has been proposed by [31] to model a situation of interaction as a Partially Observable Markov Decision Process (POMDP) by assuming that (i) the robot and the human act sequentially, one after another, and that (ii) the human is rational and makes his decision without considering the future robot's action.

7.2.3. Ativity recognition and prediction

Participants: François Charpillet, Francis Colas, Serena Ivaldi, Niyati Rawal, Vincent Thomas.

This work is part of the ANR Flying Co-Worker project and focuses on activity recognition and long-term prediction for collaborative robotics. Recognizing and predicting human activities is fundamental for a robot to help a human. Previous work in the team on activity recognition [6] rely on Hidden Markov Models (HMM) with, in particular, the Markov assumption stating that the distribution on the next state is independent from former states given the current state. This assumption, at the heart of the recurrent expression of the inference in HMM, has the unfortunate consequence to constrain the a priori distribution on the duration in each state to exponential distributions. However, it can be observed in datasets that this is not the case for many activities, which have a typical duration. This discrepancy is negligeable for recognition where HMM models achieve good performance thanks to the observations, but prevents longer-term activity prediction.

In the master project of Niyati Rawal, we investigated a slightly different model, Explicit Duration Hidden Markov Model (EDHMM), in which the duration of the activity can be modeled more finely. Preliminary results show that the recognition performance was similar to HMM but with a better prediction performance.

7.2.4. Humanoid Whole-Body Movement Optimization from Retargeted Human Motions

Participants: Waldez Azevedo Gomes Junior, Vishnu Radhakrishnan, Luigi Penco, Valerio Modugno, Jean-Baptiste Mouret, Serena Ivaldi.

Motion retargeting and teleoperation are powerful tools to demonstrate complex whole-body movements to humanoid robots: in a sense, they are the equivalent of kinesthetic teaching for manipulators. However, retargeted motions may not be optimal for the robot: because of different kinematics and dynamics, there could be other robot trajectories that perform the same task more efficiently, for example with less power consumption. We propose to use the retargeted trajectories to bootstrap a learning process aimed at optimizing the whole-body trajectories w.r.t. a specified cost function. To ensure that the optimized motions are safe, i.e., they do not violate system constraints, we used constrained optimization algorithms. We compared both global and local optimization approaches, since the optimized robot solution may not be close to the demonstrated one. We evaluated our framework with the humanoid robot iCub on an object lifting scenario, initially demonstrated by a human operator wearing a motion-tracking suit. By optimizing the initial retargeted movements, we can improve robot performance by over 40%.

Publication: [14]

7.2.5. Tele-operation of Humanoids

Participants: Luigi Penco, Waldez Gomes, Valerio Modugno, Serena Ivaldi.

We envision a world where robots can act as physical avatars and effectively replace humans in hazardous scenarios by means of teleoperation, which we see as a particular way of interacting with a robot. However, teleoperating humanoids is a challenging task because of differences in kinematics (e.g., structure and joint limits) and dynamics (e.g., mass distribution, inertia) are still significant. Another crucial issue is ensuring the dynamic balance of the robot while trying to imitate the human motion. We propose a multi-mode teleoperation framework for controlling humanoid robots for loco-manipulation tasks that address the aforementioned challenges by using two levels of teleoperation: a low-level for manipulation, realized via whole-body teleoperation, and a high-level for locomotion, based on the generation of reference velocities that are then tracked by the humanoid. We believe that this combination of different modes of teleoperation will considerably ease the burden of controlling humanoids, ultimately increasing their adaptability to complex situations which cannot be handled satisfactorily by fully autonomous systems.

Publication: [11]

7.2.6. Activity Recognition for Ergonomics Assessment of Industrial Tasks with Automatic Feature Selection

Participants: Adrien Malaisé, Pauline Maurice, Francis Colas, Serena Ivaldi.

In industry, ergonomic assessment is currently performed manually based on the identification of postures and actions by experts. We aim at proposing a system for automatic ergonomic assessment based on activity recognition. In this work, we define a taxonomy of activities, composed of four levels, compatible with items evaluated in standard ergonomic worksheets. The proposed taxonomy is applied to learn activity recognition models based on Hidden Markov Models. We also identify dedicated sets of features to be used as input of the recognition models so as to maximize the recognition performance for each level of our taxonomy. We compare three feature selection methods to obtain these subsets. Data from 13 participants performing a series of tasks mimicking industrial tasks are collected to train and test the recognition module. Results show that the selected subsets allow us to successfully infer ergonomically relevant postures and actions.

Publication: [6]

7.2.7. Human movement and ergonomics: An industry-oriented dataset for collaborative robotics

Participants: Pauline Maurice, Adrien Malaisé, Serena Ivaldi.

With the participation of Clélie Amiot, Nicolas Paris and Guy-Junior Richard, interns from Université de Lorraine during the summer 2018.

Improving work conditions in industry is a major challenge that can be addressed with new emerging technologies such as collaborative robots. Machine learning techniques can improve the performance of those robots, by endowing them with a degree of awareness of the human state and ergonomics condition. The availability of appropriate datasets to learn models and test prediction and control algorithms, however, remains an issue. This work presents a dataset of human motions in industry-like activities, fully labeled according to the ergonomics assessment worksheet EAWS, widely used in industries such as car manufacturing. Thirteen participants performed several series of activities, such as screwing and manipulating loads under different conditions, resulting in more than 5 hours of data. The dataset contains the participants' whole-body kinematics recorded both with wearable inertial sensors and marker-based optical motion capture, finger pressure force, video recordings, and annotations by three independent annotators of the performed action and the adopted posture following the EAWS postural grid. Sensor data are available in different formats to facilitate their reuse. The dataset is intended for use by researchers developing algorithms for classifying, predicting, or evaluating human motion in industrial settings, as well as researchers developing collaborative robotics solutions that aim at improving the workers' ergonomics. The annotation of the whole dataset following an ergonomics standard makes it valuable for ergonomics-related applications, but we expect its use to be broader in the robotics, machine learning, and human movement communities.

Publication: [8]

7.2.8. Objective and Subjective Effects of a Passive Exoskeleton on Overhead Work

Participants: Pauline Maurice, Serena Ivaldi.

Collaboration with Jernej Čamernik, Daša Gorjan and Jan Babič (Jozef Stefan Institute, Ljubljana, Slovenia), with Benjamin Schirrmeister and Jonas Bornmann (Otto Bock SE & Co. KGaA, Duderstadt, Germany), with Luca Tagliapietra, Claudia Latella and Daniele Pucci (Istituto Italiano di Tecnologia, Genova, Italy), and with Lars Fritzsche (IMK Automotive, Chemitz, Germany).

Overhead work is a frequent cause of shoulder work-related musculoskeletal disorders. Exoskeletons offering arm support have the potential to reduce shoulder strain, without requiring large scale reorganization of the workspace. Assessment of such systems however requires to take multiple factors into consideration. This work presents a thorough in-lab assessment of PAEXO, a novel passive exoskeleton for arm support during overhead work. A list of evaluation criteria and associated performance metrics is proposed to cover both objective and subjective effects of the exoskeleton, on the user and on the task being performed. These metrics are measured during a lab study, where 12 participants perform an overhead pointing task with and without the exoskeleton, while their physical, physiological and psychological states are monitored. Results show that using PAEXO reduces shoulder physical strain as well as global physiological strain, without increasing low back strain nor degrading balance. These positive effects are achieved without degrading task performance.

Importantly, participant' opinions of PAEXO are positive, in agreement with the objective measures. Thus, PAEXO seems a promising solution to help prevent shoulder injuries and diseases among overhead workers, without negatively impacting productivity.

Publication: [7], [19]

7.2.9. Assessing and improving human movements using sensitivity analysis and digital human simulation

Participant: Pauline Maurice.

Collaboration with Vincent Padois (Inria Bordeaux and Sorbonne Université), Yvan Measson (CEA-LIST) and Philippe Bidaud (ONERA and Sorbonne Université).

Enhancing the performance of technical movements aims both at improving operational results and at reducing biomechanical demands. Advances in human biomechanics and modeling tools allow to evaluate human performance with more and more details. Finding the right modifications to improve the performance is, however, still addressed with extensive time consuming trial-and-error processes. This work presents a framework for easily assessing human movements and automatically providing recommendations to improve their performances. An optimization-based whole-body controller is used to dynamically replay human movements from motion capture data, to evaluate existing movements. Automatic digital human simulations are then run to estimate performance indicators when the movement is performed in many different ways. Sensitivity indices are thereby computed to quantify the influence of postural parameters on the performance. Based on the results of the sensitivity analysis, recommendations for posture improvement are provided. The method is successfully validated on a drilling activity.

Publication: [9]

7.2.10. Human Motion analysis for assistance

Participants: François Charpillet, Jessica Colombel.

Collaboration with David Daney (Inria Bordeaux, Auctus Team)

Different sort of sensors can be used for rehabilitation at home. This year we have evaluated the usabily of a Kinect 2. The proposed approach is to improve joint angle estimates. It is based on a constrained extended Kalman Filter that tracks inputted measured joint centers. Since the proposed approach uses a biomechanical model, it allows to obtain physically consistent constrained joint angles and constant segment lengths. A practical method, that is not sensor specific, for the optimal tuning of the extended Kalman filter covariance matrices is provided. It uses reference data obtained from a stereophotogrammetric system but it has to be tuned only once since it is task specific only. The improvement of optimal tuning over classical methods for setting the covariance matrices is shown with a statistical parametric mapping analysis. The proposed approach was tested with six healthy subjects performing 4 rehabilitation tasks. Joint estimates accuracy was assessed with a reference stereophotogrammetric system. Even if some joints such as the internal/external rotations were not well estimated, the proposed optimized algorithm reached a satisfactory average root mean square difference of 9.7deg and a correlation coefficient 0.86 of for all joints. Our results show that affordable RGB-D sensor can be used for simple in-home rehabilitation when using a constrained biomechanical model.

A work carried out this year, takes the search for a sensor for personal assistance a step further with the study of the new Kinect Azure. Human-robot interaction requires a robust estimate of human motion in real-time. This work presents a fusion algorithm for joint center positions tracking from multiple depth cameras to improve human motion analysis accuracy. The proposed algorithm is based on body tracking measurements fusion with an extended Kalman filter and anthropomorphic constraints. However, the effectiveness and robustness of such algorithm depends on the A direct comparison of joint center positions estimated with a reference stereophotogrammetric system and the ones estimated with the new Kinect 3 (Azure Kinect) sensor and its older version the Kinect 2 (Kinect for Windows) has been made. The proposed approach improves body tracker data even for Kinect 3 which has not the same characteristics than Kinect 2. This study shows also the importance of defining good heuristics to merge data depending on how the body tracking works. Thus, with

proper heuristics, the joint center position estimates are improved by at least 14.6 %. Finally, we propose an additional comparison between Kinect 2 and Kinect 3 exhibiting the pros and cons of the two sensors. This study is now in submission for an international conference.

Finally, a state of the art on biological motion was realized. The purpose of this study is to understand and develop methods for decomposing motion. The EWalk dataset (http://gamma.cs.unc.edu/GAIT/#EWalk) will allow us to test emotion recognition from simple decompositions and classifiers. Then, we will extend the methods to other cognitive parameters.

7.2.11. Reliable localization of pedestrians in a smart home using multi-sensor data fusion

Participants: François Charpillet, Lina Achaji.

Collaboration with Maan Badaoui EL Najjar(Cristal Laboratory Lille, DiCOT Team), Mohamad Daher (the Lebanese University Faculty of technology, Tripoli)

One objective of the Larsen team is to develop technologies allowing older people to live independently as long as possible in their own homes instead of in specialized institutions. However, elderly people face physical problems that reduce their autonomy, and consequently their capacity to achieve daily activities. The integration of environmental or body sensors in what is called nowadays smart habitats is a solution that is appealing to provide a better quality of life with safer conditions. Localization and tracking of people in indoor environments are one of the primary services to be developed to follow them up at home, permitting to evaluate their physical states through the observation of their Activities of Daily Living (ADL). We proposed during the internship of Lina Achaji to localize and track the center of pressure (CoP) of people (one or two) in a smart home using a load sensing floor equipped with around 400 load sensors as well as wearable sensors. The data fusion is made using an informational filter where an inverted pendulum bio-mechanical model is introduced. The obtained results are very promising and were validated using a motion tracking system and force plates.

Publication: [4]

7.2.12. Ambient assisting living

Participants: François Charpillet, Yassine El Khadiri.

Collaboration with Cedric Rose from Diatelic compagny.

The ageing of the population confronts modern societies with an unprecedented demographic transformation. These include the imbalance in our pension systems and the cost of caring for the elderly. On this last point, apart from the economic aspects, the placement of elderly people is often only a choice of reason and can be quite badly experienced by people. One response to this societal problem is the development of technologies that make it easier to keep elderly people at home. The state of the art in this field abounds with upstream projects that are moving in this direction. Many of them are seeking to develop home monitoring systems. Their objectives are to detect and even prevent the occurrence of worrying or critical situations and to assess the physical condition or even fragility of the people being monitored. It is within this framework that this contribution is made. In this work, we have focused on the particular problem of monitoring the quality of sleep as well as the detection of nocturnal waking of a person living alone at home. The home is equipped with simple ambient sensors such as binary motion detectors. We have developed a Bayesian inference method that allows our solution to be flexible and robust enough for different types of installations and apartment configurations while maintaining a prediction accuracy of 0.94. This solution is currently being deployed on several dozen apartments in Lorraine by Diatelic and Pharmagest compagnies.

Publication: [15]

PERVASIVE Project-Team

6. New Results

6.1. Observing and Modelling Expertise and Awareness from Eye-gaze and Emotion

Participants: Thomas Guntz, James Crowley, Dominique Vaufreydaz, Philippe Dessus, Raffaella Balzarini.

We have constructed an instrument for capturing and interpreting multimodal signals of humans engaged in solving challenging problems. Our instrument captures eye gaze, fixations, body postures, and facial expressions signals from humans engaged in interactive tasks on a touch screen. We use a 23 inch Touch-Screen computer, a Kinect 2.0 mounted 35 cm above the screen to observe the subject, a 1080p Webcam for a frontal view, a Tobii Eye-Tracking bar (Pro X2-60 screen-based) and two adjustable USB-LED for lighting condition control. A wooden structure is used to rigidly mount the measuring equipment in order to assure identical sensor placement and orientation for all recordings.

As a pilot study, we observed expert chess players engaged in solving problems of increasing difficulty]. Our initial hypothesis was that we could directly detect awareness of significant configurations of chess pieces (chunks) from eye-scan and physiological measurements of emotion in reaction to game situation. The pilot experiment demonstrated that this initial hypothesis was overly simplistic.

In order to better understand the phenomena observed in our pilot experiment, we have constructed a model of the cognitive processes involved, using theories from cognitive science and classic (symbolic) artificial intelligence. This model is a very partial description that allows us to ask questions and make predictions to guide future experiments. Our model posits that experts reason with a situation model that is strongly constrained by limits to the number of entities and relations that may be considered at a time. This limitation forces subjects to construct abstract concepts (chunks) to describe game play, in order to explore alternative moves. Expert players retain associations of situations with emotions in long-term memory. The rapid changes in emotion correspond to recognition of previously encountered situations during exploration of the game tree. Recalled emotions guide selection of situation models for reasoning. This hypothesis is in accordance with Damasio's Somatic Marker hypothesis, which posits that emotions guide behavior, particularly when cognitive processes are overloaded.

Our hypothesis is that the subject uses the evoked emotions to select from the many possible situations for reasoning about moves during orientation and exploration. With this interpretation, the player rapidly considers partial descriptions as situations composed of a limited number of perceived chunks. Recognition of situations from experience evokes emotions that are displayed as face expressions and body posture.

With this hypothesis, valence, arousal and dominance are learned from experience and associated with chess situations in long-term memory to guide reasoning in chess. Dominance corresponds to the degree of experience with the recognized situation. As players gain experience with alternate outcomes for a situation, they become more assured in their ability to spot opportunities and avoid dangers. Valence corresponds to whether the situation is recognized as favorable (providing opportunities) or unfavorable (creating threats). Arousal corresponds to the imminence of a threat or opportunity. A defensive player will give priority to reasoning about unfavorable situations and associated dangers. An aggressive player will seek out high valence situations. All players will give priority to situations that evoke strong arousal. The amount of effort that player will expend exploring a situation can determined by dominance.

6.2. Recognition, Modelling and Description of Manipulation Actions

Participants: Nachwa Abou Bakr, James Crowley.

A full understanding of human actions requires: recognizing what action has been performed, predicting how it will affect the surrounding environment, explaining why this action has been performed, and who is performing it. Classic approaches to action recognition interpret a spatio-temporal pattern in a video sequence to tell what action has been performed, and perhaps how and where it was performed. A more complete understanding requires information about why the action was performed, and how it affects the environment. This face of understanding can be provided by explaining the action as part of a narrative.

We have addressed the problem of recognition, modelling and description of human activities, with results on three problems: (1) the use of transfer learning for simultaneous visual recognition of objects and object states, (2) the recognition of manipulation actions from state transitions, and (3) the interpretation of a series of actions and states as events in a predefined story to construct a narrative description.

These results have been developed using food preparation activities as an experimental domain. We start by recognizing food classes such as tomatoes and lettuce and food states, such as sliced and diced, during meal preparation. We adapt the VGG network architecture to jointly learn the representations of food items and food states using transfer learning. We model actions as the transformation of object states. We use recognised object properties (state and type) to detect corresponding manipulation actions by tracking object transformations in the video. Experimental performance evaluation for this approach is provided using the 50 salads and EPIC-Kitchen datasets. We use the resulting action descriptions to construct narrative descriptions for complex activities observed in videos of 50 salads dataset.

RAINBOW Project-Team

6. New Results

6.1. Optimal and Uncertainty-Aware Sensing

6.1.1. Tracking of Rigid Objects of Complex Shapes with a RDB-D Camera

Participants: Agniva Sengupta, Alexandre Krupa, Eric Marchand.

In the context of the iProcess project (see Section 8.3.8), we developed a method for accurately tracking the pose of rigid objects of complex shapes using a RGB-D camera [52]. This method only needs a coarse 3D geometric model of the object of interest represented as a 3D mesh. The tracking of the object is based on a joint minimization of geometric and photometric criteria and more particularly on a combination of point-to-plane distance minimization and photometric error minimization. The concept of successive "keyframes" was also used in this approach for minimizing possible drift of the tracking. The proposed approach was validated on both simulated and real data and the results experimentally demonstrated a better tracking accuracy than existing state-of-the-art 6-DoF object tracking methods, especially when dealing with low-textured objects, multiple coplanar faces, occlusions and partial specularities of the scene.

6.1.2. Deformable Object 3D Tracking based on Depth Information and Coarse Physical Model

Participants: Agniva Sengupta, Alexandre Krupa, Eric Marchand.

This research activity was also carried out in the context of the iProcess project (see Section 8.3.8) and will continue with the recent starting GentleMAN project (see Section 8.3.9). It focusses on the elaboration of approaches able to accurately track in real-time the deformation of soft objects using a RGB-D camera. The state-of-the-art approaches are currently relying on the use of Finite Element Model (FEM) to simulate the physics (mechanical behavior) of the deformable object. However, they suffer from the drawback of being excessively dependent on the accurate knowledge of the physical properties of the object being tracked (Young Modulus, Poisson's ratio, etc). This year, we proposed a first method that only required a coarse physical model of the object based on FEM whose parameters do not need to be precise [53]. The method consists in applying a set of virtual forces on the surface mesh of our coarse FEM model in such a way that it deforms to fit the current shape of the object. A point-to-plane distance error between the point cloud provided by the depth camera and the model mesh is iteratively minimized with respect to these virtual forces. The point of application of force is determined by an analysis of the error obtained from rigid tracking, which is done in parallel with the non-rigid tracking. The approach has been validated on simulated objects with ground-truth, as well on real objects of unknown physical properties and experimentally demonstrated that accurate tracking of deformable objects can be achieved without the need of a precise physical model.

6.1.3. Trajectory Generation for Optimal State Estimation

Participants: Marco Cognetti, Paolo Robuffo Giordano.

This activity addresses the general problem of *active sensing* where the goal is to analyze and synthesize optimal trajectories for a robotic system that can maximize the amount of information gathered by the (few) noisy outputs (i.e., sensor readings) while at the same time reducing the negative effects of the process/actuation noise. Over the last years we have developed a general framework for solving *online* the active sensing problem by continuously replanning an optimal trajectory that maximizes a suitable norm of the Constructibility Gramian (CG), while also coping with a number of constraints including limited energy and feasibility. The results obtained so far have been generalized and summarized in [27], where the online trajectory replanning for CG maximization has been applied to two relevant case studies (unicycle and quadrotor) and validated via a large statistical campaign. We are actually working towards the extension of this machinery to the case of realization of a robot task (e.g., reaching and grasping for a mobile manipulator), and to the mutual localization problem for a multi-robot group.

6.1.4. Robotic manipulators in Physical Interaction with the Environment

Participant: Claudio Pacchierotti.

As robotic systems become more flexible and intelligent, they must be able to move into environments with a high degree of uncertainty or clutter, such as our homes, workplaces, and the outdoors. In these unstructured scenarios, it is possible that the body of the robot collides with its surroundings. As such, it would be desirable to characterise these contacts in terms of their location and interaction forces. We worked to address the problem of detecting and isolating collisions between a robotic manipulator and its environment, using only on-board joint torque and position sensing [37]. We presented an algorithm based on a particle filter that, under some assumptions, is able to identify the contact location anywhere on the robot body. It requires the robot to perform small exploratory movements, progressively integrating the new sensing information through a Bayesian framework. The method assumes negligible friction forces, convex contact surfaces, and linear contact stiffness. Compared to existing approaches, it allows this detection to be carried in almost all the surface of the robot's body. We tested the proposed approach both in simulation and in a real environment. Experiments in simulation showed that our approach outperformed two other methods that made simpler assumptions. Experiments in a real environment using a robot with joint torque sensors showed the applicability of the method to real world scenarios and its ability to cope with situations where the algorithm's assumptions did not hold.

6.1.5. Cooperative Localization using Interval Analysis

Participants: Ide Flore Kenmogne Fokam, Vincent Drevelle, Eric Marchand.

In the context of multi-robot fleets, cooperative localization consists in gaining better position estimate through measurements and data exchange with neighboring robots. Positioning integrity (i.e., providing reliable position uncertainty information) is also a key point for mission-critical tasks, like collision avoidance. The goal of this work is to compute position uncertainty volumes for each robot of the fleet, using a decentralized method (i.e., using only local communication with the neighbors). The problem is addressed in a bounded-error framework, with interval analysis and constraint propagation methods. These methods enable to provide guaranteed position error bounds, assuming bounded-error measurements. They are not affected by over-convergence due to data incest, which makes them a well sound framework for decentralized estimation. Quantifier elimination techniques have been used to consider uncertainty in the landmarks positions without adding pessimism in the computed solution. This work has been applied to cooperative localization of UAVs, based on image and range measurements [20].

6.2. Advanced Sensor-Based Control

6.2.1. Sensor-based Trajectory Planning for quadrotor UAVs

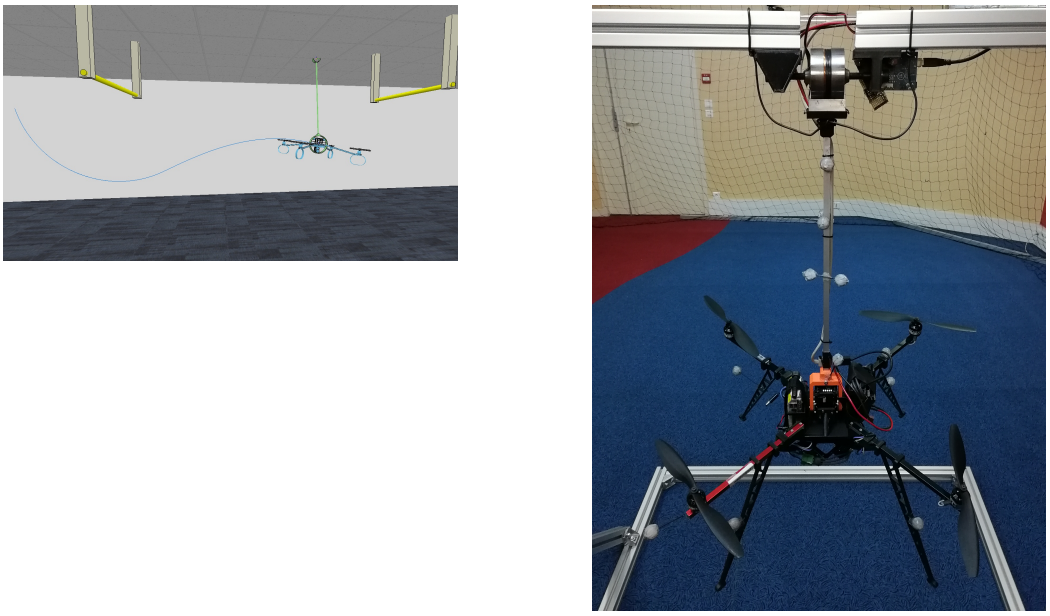
Participants: François Chaumette, Paolo Robuffo Giordano.

In the context of developing robust navigation strategies for quadrotor UAVs with onboard cameras and IMUs, we considered the problem of planning minimum-time trajectors in a cluttered environment for reaching a goal while coping with actuation and sensing constraints [25]. In particular, we considered a realistic model for the onboard camera that considers limited fov and possible occlusions due to obstructed visibility (e.g., presence of obstacles). Whenever the camera can detect landmarks in the environment, the visual cues can be used to drive a state estimation algorithm (a EKF) for updating the current estimation of the UAV state (its pose and velocity). However, beacause of the sensing constraints, the possibility of detecting and tracking the landmarks may be lost while moving in the environment. Therefore, we proposed a robust "perception-aware" planning strategy, based on the bi-directional A^* planner,

6.2.2. UAVs in Physical Interaction with the Environment

Participants: Quentin Delamare, Paolo Robuffo Giordano.

Most research in UAVs deals with either contact-free cases (the UAVs must avoid any contact with the environment), or "static" contact cases (the UAVs need to exert some forces on the environment in quasi-static conditions, reminiscent of what has been done with manipulator arms). Inspired by the vast literature on robot locomotion (from, e.g., the humanoid community), in this research topic we aim at exploiting the contact with the environment for helping a UAV maneuvering in the environment, in the same spirit in which we humans (and, supposedly, humanoid robots) use our legs and arms when navigating in cluttered environments for helping in keeping balance, or perform maneuvers that would be, otherwise, impossible. During last year we have considered the modeling, control and trajectory planning problem for a planar UAV equipped with a 1 DoF actuated arm capable of hooking at some pivots in the environment. This UAV (named MonkeyRotor) needs to "jump" from one pivot to the next one by exploiting the forces exchanged with the environment (the pivot) and its own actuation system (the propellers), see Fig. 8 (a). We are currently finalizing a real prototype (Fig. 8 (b)) for obtaining an experimental validation of the whole approach [1].



(a) (b)

Figure 8. UAVs in Physical Interaction with the Environment. a) The simulated MonkeyRotor performing a hook-to-hook maneuver. b) The prototype currently under finalization.

6.2.3. Trajectory Generation for Minimum Closed-Loop State Sensitivity

Participants: Pascal Brault, Quentin Delamare, Paolo Robuffo Giordano.

The goal of this research activity is to propose a new point of view in addressing the control of robots under parametric uncertainties: rather than striving to design a sophisticated controller with some robustness guarantees for a specific system, we propose to attain robustness (for any choice of the control action) by suitably shaping the reference motion trajectory so as to minimize the *state sensitivity* to parameter uncertainty of the resulting closed-loop system. During this year, we have extended the existing minimization framework to also include the notion of "input sensitivity", which allows to obtain trajectories whose realization (in perturbed conditions) leaves the control inputs unchanged to the largest extent. Such a feature is relevant whenever dealing with, e.g., limited actuation since it guarantees that, even under model perturbations, the inputs do not deviate too much from their nominal values. This novel input sensitivity has been combined

with the previously introduced notion of state sensitivity and validated both via monte-carlo simulations and experimentally with a unicycle robot in a large number of tests [1].

6.2.4. Visual Servoing for Steering Simulation Agents

Participants: Axel Lopez Gandia, Eric Marchand, François Chaumette, Julien Pettré.

This research activity is dedicated to the simulation of human locomotion, and more especially to the simulation of the visuomotor loop that controls human locomotion in interaction with the static and moving obstacles of its environment. Our approach is based on the principles of visual servoing for robots. To simulate visual perception, an agent perceives its environment through a virtual camera located in the position of its head. The visual input is processed by each agent in order to extract the relevant information for controlling its motion. In particular, the optical flow is computed to give the agent access to the relative motion of visible objects around it. Some features of the optical flow are finally computed to estimate the risk of collision with obstacle. We have established the mathematical relations between those visual features and the agent's self motion. Therefore, when necessary, the agent motion is controlled and adjusted so as to cancel the visual features indicating a risk of future collision [22], [46].

6.2.5. Strategies for Crowd Simulation Agents

Participants: Wouter Van Toll, Julien Pettré.

This research activity is dedicated to the simulation of crowds based on microscopic approaches. In such approaches, agents move according to local models of interactions that give them the capacity to adjust to the motion of neighbor agents. These purely local rules are not sufficient to produce high-quality long term trajectories through their environment. We provide agents with the capacity to establish mid-term strategies to move through their environment, by establishing a local plan based on their prediction of their surroundings and by verifying regularly this prediction remains valid. In the case validity is not checked, planning a new strategy is triggered [55].

6.2.6. Study of human locomotion to improve robot navigation

Participants: Florian Berton, Julien Bruneau, Julien Pettré.

This research activity is dedicated to the study of human gaze behaviour during locomotion. This activity is directly linked to the previous one on simulation, as human locomotion study results will serve as an input for the design of novel models for simulation. We are interested in the study of the activity of the gaze during locomotion that, in addition to the classical study of kinematics motion parameters, provides information on the nature of visual information acquired by humans to move, and the relative importance of visual elements in their surroundings [36].

6.2.7. Robot-Human Interactions during Locomotion

Participants: Javad Amirian, Fabien Grzeskowiak, Marie Babel, Julien Pettré.

This research activity is dedicated to the design of robot navigation techniques to make them capable of safely moving through a crowd of people. We are following two main research paths. The first one is dedicated to the prediction of crowd motion based on the state of the crowd as sensed by a robot. The second one is dedicated to the creation of a virtual reality platform that enables robots and humans to share a common virtual space where robot control techniques can be tested with no physical risk of harming people, as they remain separated in the physical space. This year, we have delivered techniques for the short term prediction of human locomotion trajectories [34], [35] and robot-human collision avoidance [39].

6.2.8. Visual Servoing for Cable-Driven Parallel Robots

Participant: François Chaumette.

This study is done in collaboration with IRT Jules Verne (Zane Zake, Nicolo Pedemonte) and LS2N (Stéphane Caro) in Nantes (see Section 7.2.2). It is devoted to the analysis of the robustness of visual servoing to modeling and calibration errors for cable-driven parallel robots. The modeling of the closed loop system has been derived, from which a Lyapunov-based stability analysis allowed exhibiting sufficient conditions for ensuring its stability. Experimental results have validated the theoretical results obtained and shown the high robustness of visual servoing for this sort of robots [30], [56].

6.2.9. Visual Exploration of an Indoor Environment

Participants: Benoît Antoniotti, Eric Marchand, François Chaumette.

This study is done in collaboration with the Creative company in Rennes (see Section 6.2.9). It is devoted to the exploration of indoor environments by a mobile robot, Pepper typically (see Section 5.4.2) for a complete and accurate reconstruction of the environment. The exploration strategy we are currently developing is based on maximizing the entropy generated by a robot motion.

6.2.10. Deformation Servoing of Soft Objects

Participant: Alexandre Krupa.

Nowadays robots are mostly used to manipulate rigid objects. Manipulating deformable objects remains challenging due to the difficulty of accurately predicting the object deformations. This year, we developed a model-free deformation servoing method able to do an online estimation of the deformation Jacobian that relates the motion of the robot end-effector to the deformation of a manipulated soft object. The first experimental results are encouraging since they showed that our model-free visual servoing approach based on online estimation provides similar results than a model-based approach based on physics simulation that requires accurate knowledge of the physical properties of the object to deform. This approach has been recently submitted to the ICRA'20 conference.

6.2.11. Multi-Robot Formation Control

Participant: Paolo Robuffo Giordano.

Most multi-robot applications must rely on relative sensing among the robot pairs (rather than absolute/external sensing such as, e.g., GPS). For these systems, the concept of rigidity provides the correct framework for defining an appropriate sensing and communication topology architecture. In several previous works we have addressed the problem of coordinating a team of quadrotor UAVs equipped with onboard sensors (such as distance sensors or cameras) for cooperative localization and formation control under the rigidity framework. In [9] an interesting interplay between the rigidity formalism and notions of parallel robotics has been studied, showing how well-known tools from the parallel robotics community can be applied to the multi-robot case, and how these tools can be used for characterizing the stability and singularities of the typical formation control/localization algorithms.

In [17], the problem of distributed leader selection has been addressed by considering agents with a second-order dynamics, thus closer to physical robots that have some unavoidable inertia when moving. This work has extended a previous strategy developed for a first-order case and ported it to the second-order: the proposed algorithm is able to periodically select at runtime the 'best' leader (among the neighbors of the current leader) for maximizing the tracking performance of an external trajectory reference while maintaining a desired formation for the group. The approach has been validated via numerical simulations.

6.2.12. Coupling Force and Vision for Controlling Robot Manipulators

Participants: Alexander Oliva, François Chaumette, Paolo Robuffo Giordano.

The goal of this activity is about coupling visual and force information for advanced manipulation tasks. To this end, we plan to exploit the recently acquired Panda robot (see Sect. 5.4.4), a state-of-the-art 7-dof manipulator arm with torque sensing in the joints, and the possibility to command torques at the joints or forces at the end-effector. Thanks to this new robot, we plan to study how to optimally combine the torque sensing and control strategies that have been developed over the years to also include in the loop the feedback from a vision sensor (a camera). In fact, the use of vision in torque-controlled robot is quite limited because of many issues, among which the difficulty of fusing low-rate images (about 30 Hz) with high-rate torque commands (about 1 kHz), the delays caused by any image processing and tracking algorithms, and the unavoidable occlusions that arise when the end-effector needs to approach an object to be grasped.

Towards this goal, this year we have considered the problem of identification of the dynamical model for the Panda robot [18], by suitably exploiting tools from identification theory. The identified model has been validated in numerous tests on the real robot with very good results and accuracy. A special feature of the model is the inclusion of a (realistic) friction term that accounts well for joint friction (a term that is usually neglected in dynamical model identification).

6.2.13. Subspace-based visual servoing

Participant: Eric Marchand.

To date most of visual servoing approaches have relied on geometric features that have to be tracked and matched in the image. Recent works have highlighted the importance of taking into account the photometric information of the entire images. This leads to direct visual servoing (DVS) approaches. The main disadvantage of DVS is its small convergence domain compared to conventional techniques, which is due to the high non-linearities of the cost function to be minimized. We proposed to project the image on an orthogonal basis (PCA) and then servo on either images reconstructed from this new compact set of coordinates or directly on these coordinates used as visual features [23]. In both cases we derived the analytical formulation of the interaction matrix. We show that these approaches feature a better behavior than the classical photometric visual servoing scheme allowing larger displacements and a satisfactory decrease of the error norm thanks to a well modelled interaction matrix.

6.2.14. Wheelchair Autonomous Navigation for Fall Prevention

Participants: Solenne Fortun, Marie Babel.

The Prisme project (see Section 8.1.4) is devoted to fall prevention and detection of inpatients with disabilities. For wheelchair users, falls typically occur during transfer between the bed and the wheelchair and are mainly due to a bad positioning of the wheelchair. In this context, the Prisme project addresses both fall prevention and detection issues by means of a collaborative sensing framework. Ultrasonic sensors are embedded onto both a robotized wheelchair and a medical bed. The measured signals are used to detect fall and to automatically drive the wheelchair near the bed at an optimal position determined by occupational therapists. This year, we finalized the related control framework based on sensor-based servoing principles. We validated the proposed solution through usage tests within the Rehabilitation Center of Pôle Saint Hélier (Rennes).

6.3. Haptic Cueing for Robotic Applications

6.3.1. Wearable Haptics

Participants: Marco Aggravi, Claudio Pacchierotti.

We worked on developing a novel modular wearable finger interface for cutaneous and kinesthetic interaction [11], shown in Fig. 9 . It is composed of a 3-DoF fingertip cutaneous device and a 1-DoF finger kinesthetic exoskeleton, which can be either used together as a single device or separately as two different devices. The 3-DoF fingertip device is composed of a static body and a mobile platform. The mobile platform is capable of making and breaking contact with the finger pulp and re-angle to replicate contacts with arbitrarily oriented surfaces.

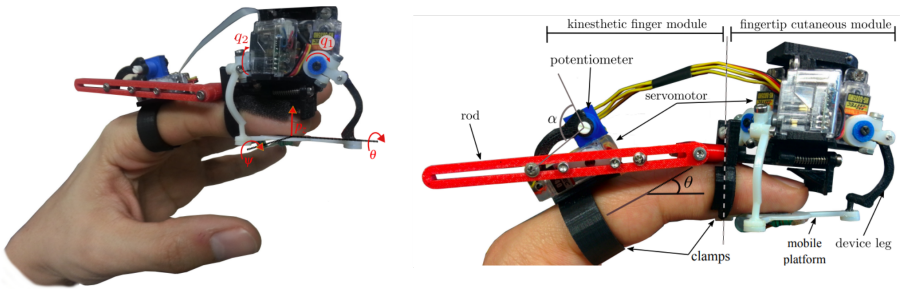


Figure 9. The proposed wearable device. It can provide both cutaneous feedback at the fingertip and kinesthetic feedback at the finger.

The 1-DoF finger exoskeleton provides kinesthetic force to the proximal and distal interphalangeal finger articulations using one servo motor grounded on the proximal phalanx. Together with the wearable device, we designed three different position, force, and compliance control schemes. We also carried out three human subjects experiments, enrolling a total of 40 different participants: the first experiment considered a curvature discrimination task, the second one a robotassisted palpation task, and the third one an immersive experience in Virtual Reality. Results showed that providing cutaneous and kinesthetic feedback through our device significantly improved the performance of all the considered tasks. Moreover, although cutaneous-only feedback showed promising performance, adding kinesthetic feedback improved most metrics. Finally, subjects ranked our device as highly wearable, comfortable, and effective.

On the same line of research, this year we guest edited a Special Issue on the IEEE Transactions on Haptics [26]. Thirteen papers on the topic have been published.

6.3.2. Mid-Air Haptic Feedback

Participants: Claudio Pacchierotti, Thomas Howard.

GUIs have been the gold standard for more than 25 years. However, they only support interaction with digital information indirectly (typically using a mouse or pen) and input and output are always separated. Furthermore, GUIs do not leverage our innate human abilities to manipulate and reason with 3D objects. Recently, 3D interfaces and VR headsets use physical objects as surrogates for tangible information, offering limited malleability and haptic feedback (e.g., rumble effects). In the framework of project H-Reality (Sect. 8.3.5), we are working to develop novel mid-air haptics paradigm that can convey the information spectrum of touch sensations in the real world, motivating the need to develop new, natural interaction techniques.

In this respect, we started working on investigating the recognition of local shapes using mid-air ultrasound haptics [45]. We have presented a series of human subject experiments investigating important perceptual aspects related to the rendering of 2D shapes by an ultrasound haptic interface (the Ultrahaptics STRATOS platform). We carried out four user studies aiming at evaluating (i) the absolute detection threshold for a static focal point rendered via amplitude modulation, (ii) the absolute detection and identification thresholds for line patterns rendered via spatiotemporal modulation, (iii) the ability to discriminate different line orientations, and (iv) the ability to perceive virtual bumps and holes.

Our results show that focal point detection thresholds are situated around 560Pa peak acoustic radiation pressure, with no evidence of effects of hand movement on detection. Line patterns rendered through spatiotemporal modulation were detectable at lower pressures, however their shape was generally not recognized as a line below a similar threshold of approx. 540Pa peak acoustic radiation pressure. We did not find any significant effect of line orientation relative to the hand both in terms of detection thresholds and in terms of correct identification of line orientation.

6.3.3. Tangible objects in VR and AR

Participant: Claudio Pacchierotti.

Still in the framework of the H-Reality project (Sect. 8.3.5), we studied the role of employing simple tangible objects in VR and AR scenarios, to improve the illusion of telepresence in these environments. We started by investigating the role of haptic sensations when interacting with tangible objects. Tangible objects are used in Virtual Reality to provide human users with distributed haptic sensations when grasping virtual objects. To achieve a compelling illusion, there should be a good correspondence between the haptic features of the tangible object and those of the corresponding virtual one, i.e., what users see in the virtual environment should match as much as possible what they touch in the real world. For this reason, we aimed at quantifying how similar tangible and virtual objects need to be, in terms of haptic perception, to still feel the same [40]. As it is often not possible to create tangible replicas of all the virtual objects in the scene, it is indeed important to understand how different tangible and virtual objects can be without the user noticing. Of course, the visuohaptic perception of objects encompasses several different dimensions, including the object's size, shape, mass, texture, and temperature. We started by addressing three representative haptic features - width, local orientation, and curvature, - which are particularly relevant for grasping. We evaluated the just-noticeable difference (JND) when grasping, with a thumb-index pinch, a tangible object which differ from a seen virtual one on the above three important haptic features. Results show JND values of 5.75%, 43.8%, and 66.66% of the reference shape for the width, local orientation, and local curvature features, respectively.

As we mentioned above, for achieving a compelling illusion during interaction in VR, there should be a good correspondence between what users see in the virtual environment and what they touch in the real world. The haptic features of the tangible object should – up to a certain extent – match those of the corresponding virtual one. We worked on an innovative approach enabling the use of few tangible objects to render many virtual ones [41]. Toward this objective, we present an algorithm which analyses different tangible and virtual objects to find the grasping strategy best matching the resultant haptic pinching sensation. Starting from the meshes of the considered objects, the algorithm guides users towards the grasping pose which best matches what they see in the virtual scene with what they feel when touching the tangible object. By selecting different grasping positions according to the virtual object to render, it is possible to use few tangible objects to render multiple virtual ones. We tested our approach in a user study. Twelve participants were asked to grasp different virtual objects, all rendered by the same tangible one. For every virtual object, our algorithm found the best pinching match on the tangible one, and guided the participant toward that grasp. Results show that our algorithm was able to well combine several haptically-salient object features to find corresponding pinches between the given tangible and virtual objects. At the end of the experiment, participants were also asked to guess how many tangible objects were used during the experiment. No one guessed that we used only one, proof of a convincing experience.

6.3.4. Wearable haptics for an Augmented Wheelchair Driving Experience

Participants: Louise Devigne, François Pasteau, Marco Aggravi, Claudio Pacchierotti, Marie Babel.

Smart powered wheelchairs can increase mobility and independence for people with disability by providing navigation support. For rehabilitation or learning purposes, it would be of great benefit for wheelchair users to have a better understanding of the surrounding environment while driving. Therefore, a way of providing navigation support is to communicate information through a dedicated and adapted feedback interface.

We then envisaged the use of wearable vibrotactile haptics, i.e. two haptic armbands, each composed of four evenly-spaced vibrotactile actuators. With respect to other available solutions, our approach provides rich navigation information while always leaving the patient in control of the wheelchair motion. We then conducted experiments with volunteers who experienced wheelchair driving in conjunction with the use of the armbands to provide drivers with information either on the presence of obstacles. Results show that providing information on closest obstacle position improved significantly the safety of the driving task (least number of collisions). This work is jointly conducted in the context of ADAPT project (Sect. 8.3.6) and ISI4NAVE associate team (Sect. 8.4.1.1).

6.4. Shared Control Architectures

6.4.1. Shared Control for Remote Manipulation

Participants: Firas Abi Farraj, Paolo Robuffo Giordano, Claudio Pacchierotti, Rahaf Rahal.

As teleoperation systems become more sophisticated and flexible, the environments and applications where they can be employed become less structured and predictable. This desirable evolution toward more challenging robotic tasks requires an increasing degree of training, skills, and concentration from the human operator. For this reason, researchers started to devise innovative approaches to make the control of such systems more effective and intuitive. In this respect, shared control algorithms have been investigated as one the main tools to design complex but intuitive robotic teleoperation system, helping operators in carrying out several increasingly difficult robotic applications, such as assisted vehicle navigation, surgical robotics, brain-computer interface manipulation, rehabilitation. This approach makes it possible to share the available degrees of freedom of the robotic system between the operator and an autonomous controller. The human operator is in charge of imparting high level, intuitive goals to the robotic system; while the autonomous controller translates them into inputs the robotic system can understand. How to implement such division of roles between the human operator and the autonomous controller highly depends on the task, robotic system, and application. Haptic feedback and guidance have been shown to play a significant and promising role in shared control applications. For example, haptic cues can provide the user with information about what the autonomous controller is doing or is planning to do; or haptic force can be used to gradually limit the degrees of freedom available to the human operator, according to the difficulty of the task or the experience of the user. The dynamic nature of haptic guidance enables us to design very flexible robotic system, which can easily and rapidly change the division of roles between the user and autonomous controller.

Along this general line of research, during this year we gave the following contributions:

- in [51] we proposed a shared control algorithm for remote telemanipulation of redundant robots able to fuse the task-prioritized control architecture (for handling the concurrent realization of multiple tasks) with haptic guidance techniques. In particular, we developed a suitable passivity-preserving strategy based on energy tanks for always guaranteeing stability despite the possible presence of autonomous tasks that could generate an increase of energy during operation. The approach has been validated with extensive simulative results in a realistic environment.
- in [6] we have considered a shared control algorithm for telemanipulation that embeds the presence of a grasping planner for guiding the operator towards suitable grasping poses. The operator retains control of the end-effector motion and eventual grasping location, but she/he is assisted by the autonomy (via force cues) in navigating towards good grasps, as classified by the grasping planner that takes as input a RGBD image of the scene and computes a set of grasping poses along the object contour.
- in [50] we have presented two haptic shared-control approaches for robotic cutting. They are designed to assist the human operator by enforcing different nonholonomic-like constraints representative of the cutting kinematics. To validate this approach, we carried out a human-subject experiment in a real cutting scenario. We compared our shared-control techniques with each other and with a standard haptic teleoperation scheme. Results show the usefulness of assisted control schemes in complex applications such as cutting.

6.4.2. Teleoperation of Flexible Needle with Haptic Feedback and Ultrasound Guidance

Participants: Jason Chevrie, Alexandre Krupa, Marie Babel.

Needle insertion procedures under ultrasound guidance are commonly used for diagnosis and therapy. This kind of intervention can greatly benefit from robotic systems to improve their accuracy and success rate. In the past years, we have developed a robotic framework dedicated to 3D steering of beveled-tip flexible needle in order to autonomously reach a desired target in the tissues by ultrasound visual servoing using a 3D ultrasound probe. This year we have proposed a real-time semi-automatic teleoperation framework that enables the user to directly control the trajectory of the needle tip during its insertion via a haptic interface [38]. The framework

enables the user to intuitively guide the trajectory of the needle tip in the ultrasound 3D volume while the controller handles the complexity of the 6D motion that needs to be applied to the needle base. A mean targeting accuracy of 2.5 mm has been achieved in gelatin phantoms and different ways to provide the haptic feedback as well as different levels of control given to the user on the tip trajectory have been compared. Limiting the user input to the insertion speed while automatically controlling the trajectory of the needle tip seems to provide a safer insertion process, however it may be too constraining and can not handle situations where more control over the tip trajectory is required, for example if unpredicted obstacles need to be avoided. On the contrary, giving the full control of the 3D tip velocity to the user and applying a haptic feedback to guide the user toward the target proved to maintain a low level of needle bending and tissue deformation.

6.4.3. Needle Comanipulation with Haptic Guidance

Participants: Hadrien Gurnel, Alexandre Krupa.

The objective of this work is to provide assistance during manual needle steering for biopsies or therapy purposes (see Section 7.2.3). At the difference of our work presented in Section 6.4.2 where a robotic system is used to steer the needle, we propose in this study another way of assistance where the needle is collaboratively manipulated by the physician and a haptic device. The principle of our approach is to provide haptic cue feedback to the clinician in order to help him during his manual gesture [43]. We elaborated 5 different haptic-guidance strategies to assist the needle pre-positioning and pre-orienting on a pre-defined insertion point, and with a pre-planned desired incidence angle. The haptic guides rely on the position and orientation errors between the needle, the entry point and the desired angle of incidence toward the target, which are computed from the measurements provided by an electromagnetic tracker. Each of the guide implements a different Guiding Virtual Fixture producing haptic cues that attract the needle towards a point or a trajectory in space with different force feedback applied on the user's hand manipulating the needle. A two-step evaluation was conducted to assess the performance and ergonomy of each haptic guide, and compare them to the unassisted reference gesture. The first evaluation stage [44] involved two physicians both experts in needle manipulation at Rennes University Hospital. The performance results showed that, compared to the unassisted gesture, the positioning accuracy was enhanced with haptic guidance. The second evaluation stage [43] was a user study with twelve participants. From this second study it results that the most constraining guide allows to perform the gesture with the best accuracy, lower time duration and highest level of ergonomy.

6.4.4. Shared Control of a Wheelchair for Navigation Assistance

Participants: Louise Devigne, Marie Babel.

Power wheelchairs allow people with motor disabilities to have more mobility and independence. However, driving safely such a vehicle is a daily challenge particularly in urban environments while navigating on sidewalks, negotiating curbs or dealing with uneven grounds. Indeed, differences of elevation have been reported to be one of the most challenging environmental barrier to negotiate, with tipping and falling being the most common accidents power wheelchair users encounter. It is thus our challenge to design assistive solutions for power wheelchair navigation in order to improve safety while navigating in such environments. To this aim, we proposed a shared-control algorithm which provides assistance while navigating with a wheelchair in an environment consisting of negative obstacles. We designed a dedicated sensor-based control law allowing trajectory correction while approaching negative obstacles e.g. steps, curbs, descending slopes. This shared control method takes into account the humanin-the loop factor. In this study, our solution the ability of our system to ensure a safe trajectory while navigating on a sidewalk is demonstrated through simulation, thus providing a proof-of-concept of our method [42].

6.4.5. Wheelchair-Human Interactions during crossing situations

Participants: Marie Babel, Julien Pettré.

Designing smart powered wheelchairs requires to better understand interactions between walkers and such vehicles. We focus on collision avoidance task between a power wheelchair (fully operated by a human) and a walker, where the difference in the nature of the agents (weight, maximal speed, acceleration profiles) results into asymmetrical physical risk in case of a collision, for example due to the protection power wheelchair provides to its driver, or the higher energy transferred to the walker during head-on collision.

We then conducted experiments with Results show that walkers set more conservative strategies when interacting with a power wheelchair. These results can then be linked to the difference in the physical characteristics of the walkers and power wheelchairs where asymmetry in the physical risks raised by collisions influence the strategies performed by the walkers in comparison with a similar walker-walker situation. This gives interesting insights in the task of modeling such interactions, indicating that geometrical terms are not sufficient to explain behaviours, physical terms linked to collision momentum should also be considered [49][62].

6.4.6. Multisensory power wheelchair simulator

Participants: Guillaume Vailland, Louise Devigne, François Pasteau, Marie Babel.

Power wheelchair driving is a challenging task which requires good visual, cognitive and visuo-spatial abilities. Besides, a power wheelchair can cause material damage or represent a danger of injury for others or oneself if not operated safely. Therefore, training and repeated practice are mandatory to acquire safe driving skills to obtain power wheelchair prescription from therapists. However, conventional training programs may reveal themselves insufficient for some people with severe impairments. In this context, Virtual Reality offers the opportunity to design innovative learning and training programs while providing realistic wheelchair driving experience within a virtual environment. We then proposed a user-centered design of a multisensory power wheelchair simulator [59][58]. This simulator addresses classical virtual experience drawbacks such as cybersickness and sense of presence by combining 3D visual rendering, haptic and vestibular feedback. It relies on a modular and versatile workflow enabling not only easy interfacing with any virtual display, but also with any user interface such as wheelchair controllers or feedback devices. First experiments with able-bodied people shown that vestibular feedback activation increases the Sense of Presence and decreases cybersickness [54].

RITS Project-Team

6. New Results

6.1. Multi-Task Cross-Modality Deep Learning for Pedestrian Risk Estimation

Participants: Danut Ovidiu Pop, Fawzi Nashashibi.

We want to solve the problem of multi-task pedestrian protection system (PPS) including not only pedestrian classification, detection and tracking,but also pedestrian action-unit classification and prediction, and finally pedestrian risk estimation. The goal of our research work is to develop an intelligent pedestrian protection component based only on single stereo vision system using an optimal cross-modality deep learning architecture in order to fulfill the prior requirements.

The system has to be able not only to detect all the pedestrians with high precision but also to track all the pedestrian paths, to classify the current pedestrian action and to predict their next actions and, finally, to estimate the pedestrian risk by the time to crossing for each pedestrian.

First, we investigate the classification component where we analyzed how learning representations from one modality would enable recognition for other modalitie(s) within various deep learning, which one terms as cross-modality learning. Second, we study how the cross modality learning improves an end-to-end the pedestrian action detection. Third, we analyze the pedestrian action prediction and the estimation of time to cross the street.

This work has been done in collaboration with Alexandrina Rogozan and Abdelaziz Bensrhair of INSA Rouen. More detail can be fund in [12], [13], [20], [11] and in the PhD manuscript of Danut Ovidiu Pop [6].

6.2. Study on the effect of rain on computer vision

Participants: Raoul de Charette, Fabio Pizzati.

Following the works initiated in past years, we have emphasized the need of developing for outdoor-applications to be robust to adverse weather conditions.

Three works were developed this year: two in the the context of the Samuel de Champlain Québec-France collaboration with Jean-François Lalonde from Univ. Laval (Canada) and another in the context of the new co-tutelle PhD thesis of Fabio Pizzati.

- We have first proposed a physically-based rain rendering pipeline for realistically inserting rain into clear weather images. Our research [16] was published at ICCV'19 and relies on a physical particle simulator, an estimation of the scene lighting and an accurate rain photometric modeling to augment images with arbitrary amount of realistic rain or fog. We validate our rendering with a user study, proving our rain is judged 40% more realistic that state-of-the-art. Using our generated weather augmented Kitti and Cityscapes dataset, we conduct a thorough evaluation of deep object detection and semantic segmentation algorithms and show that their performance decreases in degraded weather, on the order of 15% for object detection and 60% for semantic segmentation. Furthermore, we show refining existing networks with our augmented images improves the robustness of both object detection and semantic segmentation algorithms. We experiment on the popular nuScenes dataset and measure an improvement of 15% for object detection and 35% for semantic segmentation compared to original rainy performance.

Along with the research we have released the full augmented dataset on our project page [0] and the source code will be soon released.

- An alternative proposal is to use generative networks (GANs) to learn the translation of clear weather images to rainy images. This was achieved in the thesis of Fabio Pizzati and led to an accepted conference paper at WACV'20. To overcome the limitation of publicly available annotated datasets, we propose to learn the clear to rain mapping from datasets of different sources. Standard image-to-image translation architectures have limited effectiveness in such case due to the large source / target domain gap and usually fail to model typical traits of rain as water drops, which ultimately impacts the synthetic images realism. We proposed here a new type of domain bridge, that benefits from web-crawled data to reduce the domain gap.
- To circumvent the limitation of physics-based rendering and GANs rendering, we are currently working on extensions of [16] with Maxime Tremblay, PhD student at Univ. Laval. In this work, we are combining data-driven GAN approaches and physics-based driven learning.

6.3. Unsupervised Domain Adaptation

Participants: Raoul de Charette, Maximilian Jaritz, Fawzi Nashashibi, Fabio Pizzati.

There is an evident dead end to the paradigm of supervised learning, as it requires costly human labeling of millions of data frames to learn the appearance models of objects. As of today, the databases are recorded in very narrow conditions (e.g. only clear weather, only USA, only daytime). Adjusting to unseen conditions such as snow, hail, nighttime or unseen cities, require supervised algorithms to be retrained. Conversely, as humans we're capable of generalizing prior knowledge to new tasks. During this year, we initiated two works on transfer learning, typically Unsupervised Domain Adaptation (UDA) which is crucial to tackle the lack of annotations in a new domain. We have conducted two parallel projects on UDA: the first one in the scope of Maximilian Jaritz' thesis [27] (submitted), and the second one in the scope of Fabio Pizzati's work on rainy scenarios:

- **xMUDA**: In the first work, we explore how to learn from multi-modality and proposed cross-modal UDA (xMUDA) where we assumed the presence of 2D images and 3D point clouds for 3D semantic segmentation. This is challenging as the two input spaces are heterogeneous and can be impacted differently by domain shift. In xMUDA, modalities learn from each other through mutual mimicking, disentangled from the segmentation objective, to prevent the stronger modality from adopting false predictions from the weaker one. We evaluated on new UDA scenarios including day-to-night, country-to-country and dataset-to-dataset, leveraging recent autonomous driving datasets. xMUDA brings large improvements over uni-modal UDA on all tested scenarios, and is complementary to state-of-the-art UDA techniques.
- **Weighted Pseudo Labels**: The second work focus specifically on semantic segmentation in rainy scenarios. We benefited from our other work on GANs clear to rain translation to apply a self-supervised domain adaptation (aka UDA) that learns from the use of pseudo labels. Using pseudo labels enables the self-supervision of the learning reinforcing the network belief in its own predictions. To circumvent the use of hard-coded threshold, which is a common practice for pseudo labels, we proposed new Weighted Pseudo Labels that actively learn the ad-hoc threshold in a sort of region-growing techniques.

6.4. 3D completion and surface modeling

Participants: Raoul de Charette, Maximilian Jaritz, Manohar Kv.

Depth sensors (LiDARs, Time-of-flight cameras, stereo) gather geometrical knowledge about the scene which are rich and may be beneficial for many tasks. However, the depth information is usually sparse in nature and do not recover volumes and surfaces of objects.

[0] https://team.inria.fr/rits/computer-vision/weather-augment/

This year we have conducted three works on the topic: one work to densify the 3D point clouds generated from LiDAR sensors, another work to fuse 2D images and 3D point clouds, and finalized another work to reconstruct 3D deformable objects.

- The first work is in spirit a 3D point completion and was initiated with intern Manohar Kv. We developed a 3D pipeline to process point cloud and densify existing point clouds. It uses a modified version of the popular PointNet++ and it is thus able to reconstruct highly occluded 3D point clouds. The work is not yet published.
- In [17] we introduce a framework to fuse 2D multi-view images and 3D point clouds in an effective way by computing image features in 2D first, lifting them to 3D, and then fuse complementary geometry and image information in canonical 3D space. This work has been done while Maximilian Jaritz was visiting San Diego University.
- In [24], we propose a new algorithm to reconstruct 3D deformable objects heavily occluded. It uses an automatic registration of multiple depth sensors and Gaussian Mixture Modeling in the radial domain to detect and reconstruct object from their symmetrical properties. This research was applied in the context of pottery wheel for the preservation of the cultural heritage and conducted in collaboration with Mines ParisTech. It resulted that our method enabled reconstruction of challenging deformable objects with an average precision of 7.6mm.

6.5. 3D Surface Reconstruction from Voxel-based Lidar Data

Participants: Luis Roldao, Raoul de Charette, Anne Verroust-Blondet.

To achieve fully autonomous navigation, vehicles need to compute an accurate model of their direct surroundings. In fact, imprecise representations may lead to unexpected situations that could endanger the passengers. This year, we have proposed an algorithm capable to perform a fine and accurate 3D surface reconstruction of the environment from depth sensors. This representation keeps a high level of detail on the reconstruction, while maintaining a high density in the areas close to the vehicle.

Existing methods used for surface reconstruction from 3D data struggle to accommodate to the heterogeneous density of the input data while keeping the reconstruction accuracy. Conversely, our method is capable of handling this variable density by using an adaptive neighborhood kernel that perform local approximations of the data at different levels. This also permit to gain robustness against noise and output a smoother reconstruction. We also introduce a Gaussian confidence function capable to select the most adequate kernel for the local surface estimation. A Truncated Signed Distance Function (TSDF) is then globally estimated from the local surfaces to obtain the final mesh that represents the input scan.

The proposed method was evaluated in both simulated and real data. Reconstruction results show an improvement on the representation when compared with popular methods such as Implicit Moving Least Squares (IMLS), as the average error of our reconstruction is often 50% lower. Furthermore, almost 80% of vertices from our output mesh present an error below $0.2m$, while only 40% of vertices lie below the same threshold for IMLS. Our method is capable to output a higher level of detail on the reconstruction, while keeping a high density in vehicle surroundings, the mesh can be of special interest for both the robotics and the graphics community to perform different tasks, such as terrain traversability assessment or physical modeling.

More details can be found in [21]. This research is partially funded by AKKA Technology.

6.6. Attention mechanisms for vehicle trajectory prediction

Participants: Kaouther Messaoud, Fawzi Nashashibi, Anne Verroust-Blondet, Itheri Yahiaoui.

Scene understanding and future motion prediction of surrounding vehicles are crucial to achieve safe and reliable decision-making and motion planning for autonomous driving in a highway environment. This is a challenging task considering the correlation between the drivers behaviors. Two methods using attention mechanisms have been introduced in this context:

- In [18], we present a new approach based on an LSTM encoder-decoder that uses a social pooling mechanism to model the interactions between all the neighboring vehicles. This social pooling module combines both local and non-local operations: the non-local multi-head attention mechanism captures the relative importance of each vehicle despite the inter-vehicle distances to the target vehicle, while the local blocks represent nearby interactions between vehicles. Evaluations have been performed using two naturalistic driving datasets: Next Generation Simulation (NGSIM) and the highD Dataset [0]. The proposed method outperforms existing ones in terms of RMS values of prediction error, which shows the effectiveness of combining local and non-local operations in such a context.
- In [19] we propose an RRNNs based encoder-decoder architecture where the encoder analyzes the patterns underlying in the past trajectories and the decoder generates the future trajectory sequence. The originality of this network is that it combines the advantages of the LSTM blocks in representing the temporal evolution of trajectories and the attention mechanism to model the relative interactions between vehicles. The proposed method outperforms LSTM encoder decoder in terms of RMSE values of the predicted trajectories on the large scaled naturalistic driving highD dataset.

6.7. A unified framework for robust 2D/3D PML-SLAM

Participants: Kathia Melbouci, Fawzi Nashashibi.

Enhancing the outdoor mapping with SLAM based approaches is still an active research area. The main reason is that a consistent map of the vehicle's surrounding is one of the prerequisites for an effective vehicle interaction with this environment. In this context, and for the VALET project purpose, we have extended the PML-SLAM framework to handle 2D and 3D Lidars by replacing the localization module and designing a sparse pose graph optimizer. The sparse pose graph jointly optimizes the poses of the submaps generated by the local SLAM, which are already used for the mapping task, and the poses of the scans estimated following the scan matching process. This optimization is formulated as a non linear least square problem, and runs online in a background thread. The optimized poses are used to correct the vehicle's trajectory and to update the environment map. Furthermore, the graph-based PML-SLAM can deal with different sensors (IMU, GPS), that is, a sensor fusion "Kalman-filter" based is available to provide a good pose estimate for the local SLAM.

6.8. LIDAR-Based perception For Vehicle Localization in an HD Map

Participants: Farouk Ghallabi, Fawzi Nashashibi.

Self-vehicle localization is one of the fundamental tasks for autonomous driving. Most of current techniques for global positioning are based on the use of GNSS (Global Navigation Satellite Systems). However, these solutions do not provide a localization accuracy that is better than 2-3 m in open sky environments. Alternatively, the use of maps has been widely investigated for localization since maps can be pre-built very accurately. State of the art approaches often use dense maps or feature maps for localization. This year, we tackled to problems:

- In [14] we proposed a road sign perception system for vehicle localization within a third party map. This is challenging since third party maps are usually provided with sparse geometric features, which makes the localization task more difficult in comparison to dense maps. Experiments have been conducted on a Highway-like test track using GNSS/INS with RTK corrections as ground truth (GT).

[0] https://www.highd-dataset.com/

- In [15] High Reflective Landmarks (HRL) - such as lane markings, road signs and guard rail reflectors (GRR) - are detected from a 3D point cloud. A particle filtering algorithm estimates the position of the vehicle by matching observed HRLs with HD map attributes. Experiments have been conducted on a highway-like test track using GNSS/INS with RTK corrections as a ground truth (GT). Error evaluations are given as cross-track (CT) and along-track (AT) errors defined in the curvilinear coordinates related to the map. The obtained accuracies of the localization system is 18 cm for the cross-track error and 32 cm for the along-track error.

6.9. Motion planning in presence of highly dynamic obstacles with uncertain motion

Participants: Pierre de Beaucorps, Renaud Poncelet, Anne Verroust-Blondet, Fawzi Nashashibi.

Safe motion planning in a dynamic environment is of great importance in many robotics applications. This year, we have worked in two directions:

- The work on reachable interaction sets introduced in [37] has been extended to the case of dynamic obstacles with uncertain motions. We consider that the obstacles have stochastic motions and we use a probabilistic formulation to compute the RIS at each time step. Our approach improves existing methods in such a context (cf. Pierre de Beaucorps PhD thesis [7]).
- Focusing on autonomous vehicles, we begun to study scenarios with occluded dynamic obstacles.

6.10. A vehicle dynamic model corrector with side slip estimation for adding safety capabilities in autonomous vehicle

Participants: Imane Mahtout, Fawzi Nashashibi.

The ability to identify malfunctions on autonomous vehicles is critical for their deployment. As a matter of fact, systems able to identify when the positioning systems are not providing accurate data, or the perception algorithms are not properly detecting the environment, are extremely important to assure a certain safety level for automated vehicles. This is especially true since these systems are finally connected to the control module that provides the adequate commands to vehicle's actuators. For control algorithms to work properly, proper inputs are necessary to reduce noise, increase controllability and avoid system's malfunctions and instability. It is also critical for these algorithms to identify/consider vehicle physical limits for determining when is the automated system still capable of handling the vehicle. From the above, it is clear that automated vehicles are in need of proper inputs to control the vehicle, but also it is necessary to detect critical situations where the nominal control behavior is no longer assured, in order to take the vehicle to a safe state. Slide slip state is an example of a critical situation where the vehicle is no longer able to correct its trajectory. Thus, this part of my thesis work consists on developing a module for providing smooth signals to the controller and, at the same time, detect side slip situations. The lateral controller implemented in our automated driving (AD) system is based on the yaw error minimization between the desired yaw rate (obtained from the road layout information in function of the curvature) and the current vehicle yaw rate. From this, the first step is to provide a proper current yaw rate measurement. The proposed device compares the measured yaw rate value (coming from the vehicle sensor) with a model-based estimated yaw rate value. The idea is to identify vehicle model mismatches, correcting the model in real time. This permits to extend the nominal vehicle planar model to a road layout-independent model, where roll and pitch variations are considered. This first stage consists of a vehicle model compensator that includes unmodeled vehicle dynamics and parameter incertitude in real time when the vehicle is operating in autonomous mode. The vehicle lateral controller is then fed by the compensator output to allow robust performance in all road conditions. Once lateral control is fed with proper inputs, the second stage is the one detecting that vehicle handling physical limits are surpassed. The proposed system based on Youla-Kucera parametrization identifies the vehicle physical limits by estimating front and rear lateral forces, using as input the previous corrected model and on-board vehicle info. This permits to provide an accurate identification of slide slip vehicle states without adding any additional sensor to production vehicle's on-board sensors (see Fig. 1).

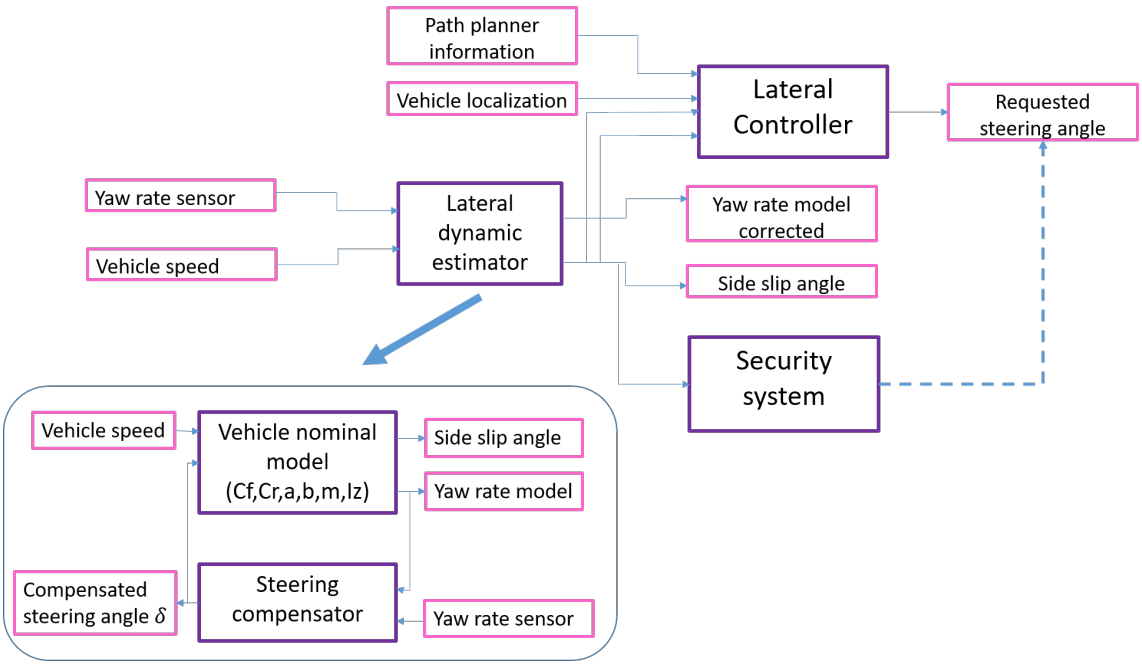


Figure 1. Lateral model compensator

6.11. Perception-adapted controller device for autonomous vehicles

Participants: Imane Mahtout, Fawzi Nashashibi.

Without loss of generality, let us consider a single camera for detecting a preceding vehicle in the road. It is clear that there are two parameters that impact the performance of the perception system:

1) the specific algorithm developed to detect and track the objects providing accurate measurement; and

2) the physical limitation of the sensor itself. For a camera, the number of pixels limits the resolution of the image so the farther away the vehicle is, the lower the accuracy in its detection. This implies a more inaccurate measurement that will degrade the ego-vehicle performance. From the vehicle response point of view, we cannot expect that a single control device can handle for example a camera-based car-following system for all detected vehicle distances.
For the sake of clarity, Figure 2 shows the speed of a preceding vehicle measured from a ground truth (solid blue line); and the measured speed from an on-board perception system. The speed of the preceding vehicle was computer controlled so we can assure a given response for it. In this example, it follows four consecutive reference speed changes from 0 to 5m/s, and finally to 8m/s. The ego-vehicle equipped with the on-board perception system was following that preceding vehicle at a speed dependent distance (i.e. as any on-the-market ACC system), meaning than the higher the speed, the higher the inter-vehicle distance. One can clearly see how the higher the speed, the higher the inaccuracy of the perception system, the more degraded the measurement.

We worked on a novel control device able to adapt its response to the perception system capabilities, modifying vehicle response accordingly to the level of accuracy of the perception system. This novel idea redefines and extends the capabilities of any ADAS or autonomous vehicle technology, not only because it improves vehicle's performance but also because we can, a-priori, understand the limitations of the full vehicle performance with a complete closed loop analysis from perception to vehicle control. As additional

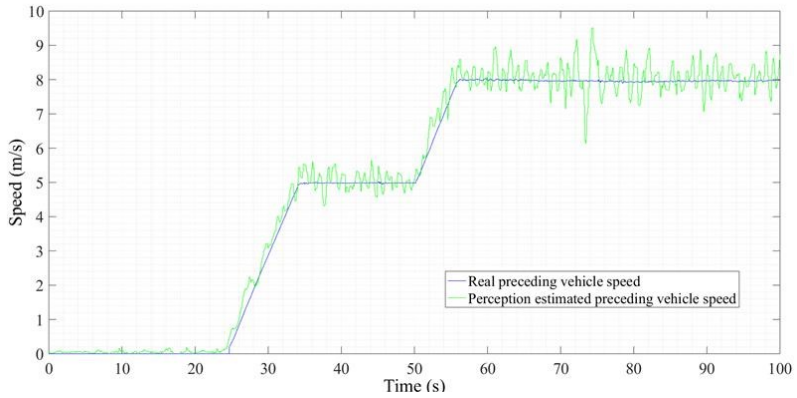


Figure 2. Perception speed profile measurement

remark, there are ways of dealing with these problems by filtering the perception signal. This causes two problems: control response stability is no longer guaranteed, and it smooths the response but it's not possible to link the full system performance to that filtering. Figure 3 presents an overview of the method. The block Perception set-of-sensor represents the specific embedded perception system. It can consist either in a single sensor or a combination of them. The characterization of the specific perception setup can be done offline, calibrating the system performance accordingly to the on-board sensors. The module Offline calibration will contain a 3-D look-up table with object distance, preceding-vehicle speed, and as a third parameter the desired measurement inaccuracy (it can be either the speed as shown in Figure 2 or any other relevant parameter as the distance, yaw rate...). This offline calibration allows defining specific design parameters for the vehicle performance. Current control systems linked to perception don't consider this inaccuracy when designing the vehicle performance. We here include two different control design criteria blocks. Assuming that we keep the regular controller design that considers perfect measurement from the perception system; the First control design criteria block includes the current production system controller. On the contrary side, we have also included a Second control design criteria block that can be adapted in function of the specific interest of each application. Following with the example on Figure 2 , let us assume that we are interested on developing an application between 0 and 10m/s and the inaccuracy of the perception system is the one presented in the plots. Having this in mind, we can design the second controller with the goal of minimizing the impact of that inaccuracy in the vehicle performance. The system also uses as input the real-time perception value coming from the Perception system measurement block (in the case of Figure 2 would provide the speed of the preceding vehicle in real-time). Then, this measurement feds the Perception-adapted controller block and the Performance degradation block. This last, accordingly to the information from the Offline calibration block, determines the status of the on-board perception system. The output of the Performance degradation block with the output of the first and second design criteria blocks fed the core module of this work: the real-time vehicle performance adaptation module. It is composed by two main blocks: the response corrector block to adapt the vehicle performance and the Perception-adapted controller block that merges both designed controller in a single stable structure.

6.12. Cyberphysical constructs and mobile communications for fully automated networked vehicles

Participant: Gérard Le Lann.

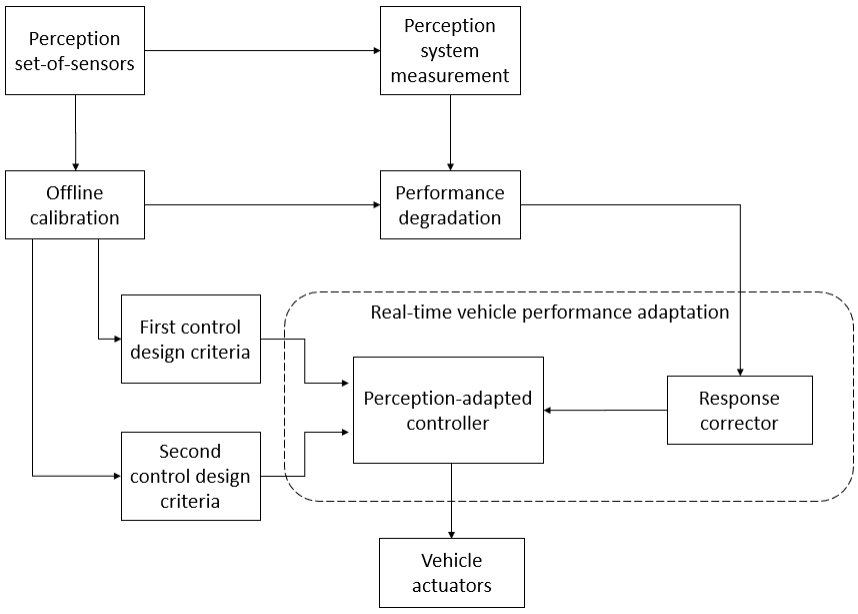


Figure 3. General overview of the block diagram for perception-adapted autonomous vehicle control device.

Safety, privacy, efficiency, and cybersecurity (SPEC) properties are key to the advent of self-forming and self-healing networks of fully automated (driverless) terrestrial vehicles. Such vehicles are referred to as Next-Gen Vehicles (NGVs) in order to avoid confusion with Connected Autonomous Vehicles (CAVs). NGVs prefigure SAE level 5 vehicles. CAVs and NGVs rest on robotics capabilities (sensors, motion control laws, actuators, onboard systems, etc.). CAVs are equipped with V2X (vehicle-to-everything) functionalities based on medium range WiFi radio communications. NGVs will be equipped with CMX (coordinated mobility for X) functionalities, X standing for S, P, E, and C, based on very short range communications (cellular radio and optics).

Work in 2019 has been devoted to defining the CMX framework and to comparing V2X and CMX functionalities. The outputs of this work have been published in [23].

Highest SE (safety and efficiency) is one of the most fundamental goals set to designers of onboard systems. It is surmised that onboard robotics must be supplemented with inter-vehicular communication (IVC) capabilities in order to achieve highest SE properties. Thus the question: which IVC capabilities? In the V2X framework, two distinct sets of IVC capabilities are considered, namely DSRC-V2X (WiFi radio) and C-V2X (4G LTE, 5G, cellular radio). IVC capabilities in the CMX framework encompass cellular radio, VLC and passive optics.

Since V2X functionalities rest on medium range radio communications, they are vulnerable to remote and local cyberattacks (message falsification, masquerading, Sybil attacks, injection of bogus messages, DDoS attacks, etc.). It has been amply demonstrated that such cyberattacks can compromise safety (collisions caused by remote and/or local attackers) as well as efficiency (congested roadways and cities). Furthermore, V2X functionalities break down when radio channels are noisy (messages get lost) or/and jammed (intentional remote and local cyberattacks). Finally, owing to decade-old design decisions, there are no privacy properties with V2X functionalities. For example, every CAV must periodically broadcast messages that carry vehicle-centric characteristics and unencrypted current GNSS space coordinates (referred to as beaconing, frequencies ranging between 1 Hz and 10 Hz. Despite certificate-based pseudonymisation, routes followed by vehicles can

be tracked and communications can be eavesdropped and recorded. Linkage with passengers) personal data is straightforward.

Therefore, in addition to degrading safety and efficiency properties achieved by onboard robotics, V2X functionalities do not meet elementary requirements regarding privacy and cybersecurity. Some proponents of the V2X approach assert that it is impossible to deliver road safety without breaching passengers' privacy. To be valid, that statement should be backed with an impossibility proof. Such a proof has not appeared yet and will never appear for the simple reason that safety and privacy properties can be achieved jointly, by design, proofs given, as demonstrated with the CMX approach.

From a more theoretical perspective, the V2X and the CMX frameworks can be contrasted as follows. Unquestionably, full asynchrony is the appropriate model for representing the vehicular network universe faithfully. Vehicles are started or stopped at arbitrary times, velocities change unpredictably, ditto for lane changes, on-ramp merging, concurrent traversals of intersections and roundabouts, and so on. Onboard processes that are life/safety critical are run in the presence of fortuitous failures, cyberattacks, and concurrency (due to resource sharing). It follows that even if one postulates the existence of finite bounds for process execution durations, it is impossible to assume any a priori knowledge of values taken by those bounds. That is precisely the definition of full asynchrony.

Numerous impossibility results relative to fully asynchronous systems have been published since the late-1970s. For example, problems akin to distributed consensus (terminating reliable broadcast, consistent multi-copied data structures, exact agreement, leader election, etc.) have no solutions in the presence of a single failure, even when communications are assumed to be perfect (no message losses). Since mobile wireless communications are unreliable, those results hold a fortiori in vehicular networks. Obviously, problems that involve termination in computable/predictable time bounds (a real-time property) have no solutions either.

The above-mentioned problems shall be solved in order to provide vehicles and vehicular networks with the SPEC properties. Knowing that solutions exist when considering synchrony models -such as e.g. partial synchrony, timed asynchrony, full synchrony- the challenge is to show how synchrony models could emerge from full asynchrony. This challenge is ignored in the V2X framework. Conclusion: since V2X designs are conducted considering full asynchrony, none of the SPEC properties may hold true.

The CMX framework results from addressing this challenge. NGVs are endowed with CMX functionalities which are based on specific cyberphysical constructs (cells, cohorts, flocks). These constructs serve to instantiate synchrony models within which it is possible to design protocols and algorithms (e.g., deterministic MAC protocols, time-bounded distributed algorithms for message dissemination, approximate agreement, and consensus) that are needed for establishing and proving the SPEC properties, while matching the real vehicular networks universe.

Concepts at the core of the CMX framework (cyberphysical levels, unfalsifiable vehicle profiles, proactive security modules, privacy-preserving naming, etc.) are detailed in [23]. Regarding SE properties, we show how to achieve theoretical absolute safety (no fatalities, no severe injuries) while keeping smallest safe gaps (highest efficiency) in cohort-structured vehicular networks, under assumptions of high coverage. As for PC properties, we show that passengers' privacy cannot be compromised via cyber eavesdropping and/or physical tracking of vehicles. This is due to the fact that messages do not carry vehicle-centric characteristics or GNSS space coordinates. CMX functionalities are shown to be immune to remote cyberattacks. Thanks to optical communications (in addition to very short range cellular radio), they can withstand radio channel jamming. Owing to controlled cohort admission, external local cyberattacks aimed at cohort members are inoperative. Local cyberattacks launched from the inside of a cohort, i.e. by cohort members themselves, can be thwarted. In the unlikely case of success, dishonest members would be involved in those collisions which they create. Conclusion: the only cyberattacks that may compromise safety in cohort-structured vehicular networks are due to irrational attackers.

6.13. Belief propagation inference for traffic prediction

Participant: Jean-Marc Lasgouttes.

This work [36], [35], in collaboration with Cyril Furtlehner (TAU, Inria), deals with real-time prediction of traffic conditions in a urban setting with incomplete data. The main focus is on finding a good way to encode available information (flow, speed, counts,...) in a Markov Random Field, and to decode it in the form of real-time traffic reconstruction and prediction. Our approach relies in particular on the Gaussian belief propagation algorithm.

Through our collaboration with PTV Sistema, we obtained extensive results on large-scale datasets containing 250 to 2000 detectors. The results show very good ability to predict flow variables and a reasonably good performance on speed or occupancy variables. Some element of understanding of the observed performance are given by a careful analysis of the model, allowing to some extent to disentangle modelling bias from intrinsic noise of the traffic phenomena and its measurement process.

This year we worked on code optimization and submitted our work to *Transportation Research: Part C*.

6.14. Stabilization of traffic through cooperative autonomous vehicles

Participants: Guy Fayolle, Carlos Flores, Jean-Marc Lasgouttes.

We investigate in [26] the transfer function emanating from the linearization of a car-following model, when taking into account a driver reaction time. This leads to stability conditions, which are explicitly given. We also show how this reaction time can introduce a *weak string instability*.

This paper is intended as a foundation of a larger work on traffic stabilization by means of a fleet of cooperative automated vehicles. Contrary to some earlier works, our approach is based on a car-following model with reaction-time delay, rather than on a first order fluid model. The continuation of these studies will concern schockwave analysis and adequate traffic-stabilizing control strategies.

6.15. Random walks in orthants and lattice path combinatorics

Participant: Guy Fayolle.

In the second edition of the book [2], original methods were proposed to determine the invariant measure of random walks in the quarter plane with small jumps (size 1), the general solution being obtained via reduction to boundary value problems. In this framework, number of difficult open problems related to lattice path combinatorics are currently being explored, in collaboration with A. Bostan and F. Chyzak (project-team SPECFUN, Inria-Saclay), both from theoretical and computer algebra points of view: concrete computation of the criteria, utilization of differential Galois theory, genus greater than 1 (i.e. when some jumps are of size ≥ 2), etc. A recent topic deals with the connections between simple product-form stochastic networks (so-called *Jackson networks*) and explicit solutions of functional equations for counting lattice walks, see [25].

6.16. Optimization of test case generation for ADAS via Gibbs sampling algorithms

Participant: Guy Fayolle.

Validating Advanced Driver Assistance Systems (ADAS) is a strategic issue, since such systems are becoming increasingly widespread in the automotive field.

But ADAS validation is a complex issue, particularly for camera based systems, because these functions may be facing a very high number of situations that can be considered as infinite. Building at a low cost level a sufficiently detailed campaign is thus very difficult. Indeed, test case generation faces the crucial question of *inherent combinatorial explosion*. An important constraint is to generate *almost all* situations in the most economical way. This task can be considered from two points of view: deterministic via binary search trees, or stochastic via Markov chain Monte Carlo (MCMC) sampling. We choose the latter probabilistic approach described below, which in our opinion seems to be the most efficient one. Typically, the problem is to produce samples of large random vectors, the components of which are possibly dependent and take a finite number of values with some given probabilities.The following flowchart is proposed.

1. In a first step, starting from the simulation graph generated by the toolboxes of MATLAB, we construct a so-called *Markov Random Field (MRF)*. When the parameters are locally dependent, this can be achieved from the user's specifications and by a systematic application of Bayes' formula.
2. Then, to cope with the combinatorial explosion, test cases are produced by implementing (and comparing) various *Gibbs samplers*, which are fruitfully employed for large systems encountered in physics. In particular, we strive to make a compromise between the convergence rate toward equilibrium, the percentage of generated duplicates and the path coverage, keeping in mind that the speed of convergence is exponential, a classical property deduced from the general theory of Markov chains.
3. The generation of rare events by mixing Gibbs samplers, large deviation techniques (LDT) and cross-entropy method is a work in progress.

The French car manufacturer *Groupe PSA* shows a great interest in these methods and has established a contractual collaboration involving ARMINES-Mines ParisTech (Guy Fayolle as associate researcher) and Can Tho University in Vietnam (Pr. Van Ly Tran).

LINKMEDIA Project-Team

7. New Results

7.1. Extracting and Representing Information

7.1.1. Text Mining in the Clinical Domain

Participants: Clément Dalloux, Vincent Claveau.

Clinical records cannot be shared, which is a real hurdle to develop and compare information extraction techniques. In the framework of the BigClin Project we have developed annotated corpora, that share the same linguistic properties than records, but can be freely distributed for research purposes. Several corpora and several types of annotation were proposed for French, Portuguese and English. They are made freely available for research purposes and are described in [27], [25]. These corpora will foster reproducible research on clinical text mining.

Thanks to these datasets, we have organized the DeFT text-mining competition in 2019. Several NLP techniques and tools have been developed within the project in order to identify relevant medical or linguistic information [30], [26]. They are all chiefly based on machine learning approaches, and for most of them, more specifically, on deep learning. For instance, we have developed a new Part-of-Speech tagger and lemmatizer for French, especially suited to handle medical texts; it is freely available as a web-service at https://allgo.inria.fr . The identification of negation and uncertainty is important to precisely understand the clinical texts. Thus, we have continued our work on neural techniques to find the negation/uncertainty cues and their scope (part of sentence concerned by the negation or uncertainty). It achieves state-of-the-art results on English, and is pioneer work for French and Portuguese for which it sets a new standard [4], [21]; it is available at https://allgo.inria.fr . Other achievements in text-mining include: numerical value extraction (finding concepts that are measured, such as lab results, numerical expressions, their units) in French, English and Portuguese, the identification of gender, age, outcome and admission reasons in French clinical texts, ...

7.1.2. Embedding in hyperbolic spaces

Participants: François Torregrossa, Vincent Claveau, Guillaume Gravier.

During this year, we have studied non-Euclidean spaces into which one can embed data (for instance, words). We have developed the HierarX tool which projects multiple datasources into hyperbolic manifolds: Lorentz or Poincaré. From similarities between word pairs or continuous word representations in high dimensional spaces, HierarX is able to embed knowledge in hyperbolic geometries with small dimensionality. Those shape information into continuous hierarchies. The source code is available on the Inria's GitLab.

7.1.3. Aggregation and embedding for group membership verification

Participants: Marzieh Gheisari Khorasgani, Teddy Furon, Laurent Amsaleg.

This paper proposes a group membership verification protocol preventing the curious but honest server from reconstructing the enrolled signatures and inferring the identity of querying clients [24]. The protocol quantizes the signatures into discrete embeddings, making reconstruction difficult. It also aggregates multiple embeddings into representative values, impeding identification. Theoretical and experimental results show the trade-off between the security and error rates.

7.1.4. Group Membership Verification with Privacy: Sparse or Dense?

Participants: Marzieh Gheisari Khorasgani, Teddy Furon, Laurent Amsaleg.

Group membership verification checks if a biometric trait corresponds to one member of a group without revealing the identity of that member. Recent contributions provide privacy for group membership protocols through the joint use of two mechanisms: quantizing templates into discrete embeddings, and aggregating several templates into one group representation. However, this scheme has one drawback: the data structure representing the group has a limited size and cannot recognize noisy query when many templates are aggregated. Moreover, the sparsity of the embeddings seemingly plays a crucial role on the performance verification. This contribution proposes a mathematical model for group membership verification allowing to reveal the impact of sparsity on both security, compactness, and verification performances [23]. This models bridges the gap towards a Bloom filter robust to noisy queries. It shows that a dense solution is more competitive unless the queries are almost noiseless.

7.1.5. Privacy Preserving Group Membership Verification and Identification

Participants: Marzieh Gheisari Khorasgani, Teddy Furon, Laurent Amsaleg.

When convoking privacy, group membership verification checks if a biometric trait corresponds to one member of a group without revealing the identity of that member. Similarly, group membership identification states which group the individual belongs to, without knowing his/her identity. A recent contribution provides privacy and security for group membership protocols through the joint use of two mechanisms: quantizing biometric templates into discrete embeddings, and aggregating several templates into one group representation. This paper significantly improves that contribution because it jointly learns how to embed and aggregate instead of imposing fixed and hard coded rules [10]. This is demonstrated by exposing the mathematical underpinnings of the learning stage before showing the improvements through an extensive series of experiments targeting face recognition. Overall, experiments show that learning yields an excellent trade-off between security/privacy and the verification/identification performances.

7.1.6. Intrinsic Dimensionality Estimation within Tight Localities

Participants: Laurent Amsaleg, Oussama Chelly [Microsoft Germany], Michael Houle [National Institute of Informatics, Japan], Ken-Ichi Kawarabayashi [National Institute of Informatics, Japan], Miloš Radovanović [Univ. Novi Sad, Serbia], Weeris Treeratanajaru [Chulalongkorn University, Thailand].

Accurate estimation of Intrinsic Dimensionality (ID) is of crucial importance in many data mining and machine learning tasks, including dimensionality reduction, outlier detection, similarity search and subspace clustering. However, since their convergence generally requires sample sizes (that is, neighborhood sizes) on the order of hundreds of points, existing ID estimation methods may have only limited usefulness for applications in which the data consists of many natural groups of small size. In this paper, we propose a local ID estimation strategy stable even for 'tight' localities consisting of as few as 20 sample points [31]. The estimator applies MLE techniques over all available pairwise distances among the members of the sample, based on a recent extreme-value-theoretic model of intrinsic dimensionality, the Local Intrinsic Dimension (LID). Our experimental results show that our proposed estimation technique can achieve notably smaller variance, while maintaining comparable levels of bias, at much smaller sample sizes than state-of-the-art estimators.

7.1.7. Selective Biogeography-Based Optimizer Considering Resource Allocation for Large-Scale Global Optimization

Participants: Meiji Cui [Tongji University, China], Li Li [Tongji University, China], Miaojing Shi.

Biogeography-based optimization (BBO), a recent proposed meta-heuristic algorithm, has been successfully applied to many optimization problems due to its simplicity and efficiency. However, BBO is sensitive to the curse of dimensionality; its performance degrades rapidly as the dimensionality of the search space increases. In [3], a selective migration operator is proposed to scale up the performance of BBO and we name it selective BBO (SBBO). The differential migration operator is selected heuristically to explore the global area as far as possible whist the normal distributed migration operator is chosen to exploit the local area. By the means of heuristic selection, an appropriate migration operator can be used to search the global optimum efficiently. Moreover, the strategy of cooperative co-evolution (CC) is adopted to solve large-scale global optimization problems (LSOPs). To deal with subgroup imbalance contribution to the whole solution in the context of

CC, a more efficient computing resource allocation is proposed. Extensive experiments are conducted on the CEC 2010 benchmark suite for large-scale global optimization, and the results show the effectiveness and efficiency of SBBO compared with BBO variants and other representative algorithms for LSOPs. Also, the results confirm that the proposed computing resource allocation is vital to the large-scale optimization within the limited computation budget.

7.1.8. Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm

Participants: Shugang Li [Shanghai University, China], Xuewei Song [Shanghai University, China], Hanyu Lu [Shanghai University, China], Linyi Zeng [Shanghai University, Industrial and Commercial Bank of China, China], Miaojing Shi, Fang Liu [Shanghai University, China].

Circle structure of online brand communities allows companies to conduct cross-marketing activities by the influence of friends in different circles and build strong and lasting relationships with customers. However, existing works on the friend recommendation in social network do not consider establishing friendships between users in different circles, which has the problems of network sparsity, neither do they study the adaptive generation of appropriate link prediction algorithms for different circle features. In order to fill the gaps in previous works, the intelligent attention allocation link prediction algorithm is proposed to adaptively build attention allocation index (AAI) according to the sparseness of the network and predict the possible friendships between users in different circles. The AAI reflects the amount of attention allocated to the user pair by their common friend in the triadic closure structure, which is decided by the friend count of the common friend. Specifically, for the purpose of overcoming the problem of network sparsity, the AAIs of both the direct common friends and indirect ones are developed. Next, the decision tree (DT) method is constructed to adaptively select the suitable AAIs for the circle structure based on the density of common friends and the dispersion level of common friends' attention. In addition, for the sake of further improving the accuracy of the selected AAI, its complementary AAIs are identified with support vector machine model according to their similarity in value, direction, and ranking. Finally, the mutually complementary indices are combined into a composite one to comprehensively portray the attention distribution of common friends of users in different circles and predict their possible friendships for cross-marketing activities. Experimental results on Twitter and Google+ show that the model has highly reliable prediction performance [5].

7.1.9. Revisiting the medial axis for planar shape decomposition

Participants: Nikos Papanelopoulos [NTUA, Greece], Yannis Avrithis, Stefanos Kollias [U. of Lincoln, UK].

We present a simple computational model for planar shape decomposition that naturally captures most of the rules and salience measures suggested by psychophysical studies, including the minima and short-cut rules, convexity, and symmetry. It is based on a medial axis representation in ways that have not been explored before and sheds more light into the connection between existing rules like minima and convexity. In particular, vertices of the exterior medial axis directly provide the position and extent of negative minima of curvature, while a traversal of the interior medial axis directly provides a small set of candidate endpoints for part-cuts. The final selection follows a prioritized processing of candidate part-cuts according to a local convexity rule that can incorporate arbitrary salience measures. Neither global optimization nor differentiation is involved. We provide qualitative and quantitative evaluation and comparisons on ground-truth data from psycho-physical experiments. With our single computational model, we outperform even an ensemble method on several other competing models [6].

7.1.10. Graph-based Particular Object Discovery

Participants: Oriane Siméoni, Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Yannis Avrithis, Ondra Chum [Univ. Prague].

Severe background clutter is challenging in many computer vision tasks, including large-scale image retrieval. Global descriptors, that are popular due to their memory and search efficiency, are especially prone to corruption by such a clutter. Eliminating the impact of the clutter on the image descriptor increases the chance of retrieving relevant images and prevents topic drift due to actually retrieving the clutter in the case of query expansion. In this work, we propose a novel salient region detection method. It captures, in an unsupervised manner, patterns that are both discriminative and common in the dataset. Saliency is based on a centrality measure of a nearest neighbor graph constructed from regional CNN representations of dataset images. The proposed method exploits recent CNN architectures trained for object retrieval to construct the image representation from the salient regions. We improve particular object retrieval on challenging datasets containing small objects [7].

7.1.11. Label Propagation for Deep Semi-supervised Learning

Participants: Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Yannis Avrithis, Ondra Chum [Univ. Prague].

Semi-supervised learning is becoming increasingly important because it can combine data carefully labeled by humans with abundant unlabeled data to train deep neural networks. Classic methods on semi-supervised learning that have focused on transductive learning have not been fully exploited in the inductive framework followed by modern deep learning. The same holds for the manifold assumption—that similar examples should get the same prediction. In this work, we employ a transductive label propagation method that is based on the manifold assumption to make predictions on the entire dataset and use these predictions to generate pseudo-labels for the unlabeled data and train a deep neural network. At the core of the transductive method lies a nearest neighbor graph of the dataset that we create based on the embeddings of the same network. Therefore our learning process iterates between these two steps. We improve performance on several datasets especially in the few labels regime and show that our work is complementary to current state of the art [12], [38].

7.1.12. Dense Classification and Implanting for Few-Shot Learning

Participants: Yann Lifchitz, Yannis Avrithis, Sylvaine Picard [SAFRAN Group], Andrei Bursuc [Valéo].

Few-shot learning for deep neural networks is a highly challenging and key problem in many computer vision tasks. In this context, we are targeting knowledge transfer from a set with abundant data to other sets with few available examples. We propose in [14], [40] two simple and effective solutions: (i) dense classification over feature maps, which for the first time studies local activations in the domain of few- shot learning, and (ii) implanting, that is, attaching new neurons to a previously trained network to learn new, task-specific features. Implanting enables training of multiple layers in the few-shot regime, departing from most related methods derived from metric learning that train only the final layer. Both contributions show consistent gains when used individually or jointly and we report state of the art performance on few-shot classification on miniImageNet.

7.1.13. Point in, Box out: Beyond Counting Persons in Crowds

Participants: Yuting Liu [Sichuan University, China], Miaojing Shi, Qijun Zhao [Sichuan University, China], Xiaofang Wang [RAINBOW Team, IRISA].

Modern crowd counting methods usually employ deep neural networks (DNN) to estimate crowd counts via density regression. Despite their significant improvements, the regression-based methods are incapable of providing the detection of individuals in crowds. The detection-based methods, on the other hand, have not been largely explored in recent trends of crowd counting due to the needs for ex- pensive bounding box annotations. In this work, we instead propose a new deep detection network with only point supervision required [15]. It can simultaneously detect the size and location of human heads and count them in crowds. We first mine useful person size information from point-level annotations and initialize the pseudo ground truth bounding boxes. An online updating scheme is introduced to refine the pseudo ground truth during training; while a locally- constrained regression loss is designed to provide additional constraints on the size of the predicted boxes in a local neighborhood. In the end, we propose a curriculum learning strategy to train the network from images of relatively accurate and easy pseudo ground truth first. Extensive experiments are conducted in both detection and counting tasks on several standard benchmarks, e.g. ShanghaiTech, UCF CC

50, WiderFace, and TRANCOS datasets, and the results show the superiority of our method over the state-of-the-art.

7.1.14. Revisiting Perspective Information for Efficient Crowd Counting

Participants: Miaojing Shi, Zhaohui Yang [Peking University, China], Chao Xu [Peking University, China], Qijun Chen [Tongji University, China].

Crowd counting is the task of estimating people numbers in crowd images. Modern crowd counting methods employ deep neural networks to estimate crowd counts via crowd density regressions. A major challenge of this task lies in the perspective distortion, which results in drastic person scale change in an image. Density regression on the small person area is in general very hard. In this work, we propose a perspective-aware convolutional neural network (PACNN) for efficient crowd counting, which integrates the perspective information into density regression to provide additional knowledge of the person scale change in an image [18]. Ground truth perspective maps are firstly generated for training; PACNN is then specifically designed to predict multi-scale perspective maps, and encode them as perspective-aware weighting layers in the network to adaptively combine the outputs of multi-scale density maps. The weights are learned at every pixel of the maps such that the final density combination is robust to the perspective distortion. We conduct extensive experiments on the ShanghaiTech, WorldExpo'10, UCF CC 50, and UCSD datasets, and demonstrate the effectiveness and efficiency of PACNN over the state-of-the-art.

7.1.15. Local Features and Visual Words Emerge in Activations

Participants: Oriane Siméoni, Yannis Avrithis, Ondra Chum [Univ. Prague].

We propose a novel method of deep spatial matching (DSM) for image retrieval [19], [41]. Initial ranking is based on image descriptors extracted from convolutional neural network activations by global pooling, as in recent state-of- the-art work. However, the same sparse 3D activation tensor is also approximated by a collection of local features. These local features are then robustly matched to approximate the optimal alignment of the tensors. This hap- pens without any network modification, additional layers or training. No local feature detection happens on the original image. No local feature descriptors and no visual vocabulary are needed throughout the whole process. We experimentally show that the proposed method achieves the state-of-the-art performance on standard benchmarks across different network architectures and different global pooling methods. The highest gain in performance is achieved when diffusion on the nearest-neighbor graph of global descriptors is initiated from spatially verified images.

7.1.16. Combining convolutional side-outputs for road image segmentation

Participants: Raquel Almeida, Simon Malinowski, Ewa Kijak, Silvio Guimaraes [PUC Minas].

Image segmentation consists in creating partitions within an image into meaningful areas and objects. It can be used in scene understanding and recognition, in fields like biology, medicine, robotics, satellite imaging, amongst others. In this work [17], we take advantage of the learned model in a deep architecture, by extracting side-outputs at different layers of the network for the task of image segmentation. We study the impact of the amount of side-outputs and evaluate strategies to combine them. A post-processing filtering based on mathematical morphology idempotent functions is also used in order to remove some undesirable noises. Experiments were performed on the publicly available KITTI Road Dataset for image segmentation. Our comparison shows that the use of multiples side outputs can increase the overall performance of the network, making it easier to train and more stable when compared with a single output in the end of the network. Also, for a small number of training epochs (500), we achieved a competitive performance when compared to the best algorithm in KITTI Evaluation Server.

7.1.17. BRIEF-based mid-level representations for time series classification

Participants: Raquel Almeida, Simon Malinowski, Silvio Guimaraes [PUC Minas].

Time series classification has been widely explored over the last years. Amongst the best approaches for that task, many are based on the Bag-of-Words framework, in which time series are transformed into a histogram of word occurrences. These words represent quantized features that are extracted beforehand. In this work [20], we aim to evaluate the use of accurate mid-level representation called BossaNova in order to enhance the Bag-of-Words representation and to propose a new binary time series descriptor, called BRIEF-based descriptor. More precisely, this kind of representation enables to reduce the loss induced by feature quantization. Experiments show that this representation in conjunction to BRIEF-based descriptor is statistically equivalent to traditional Bag-of-Words, in terms time series classification accuracy, being about 4 times faster. Furthermore, it is very competitive when compared to the state-of-the-art.

7.1.18. Toward a Framework for Seasonal Time Series Forecasting Using Clustering

Participants: Simon Malinowski, Thomas Guyet [LACODAM Team], Colin Leverger [LACODAM Team], Alexandre Termier [LACODAM Team].

Seasonal behaviours are widely encountered in various applications. For instance, requests on web servers are highly influenced by our daily activities. Seasonal forecasting consists in forecasting the whole next season for a given seasonal time series. It may help a service provider to provision correctly the potentially required resources, avoiding critical situations of over- or under provision. In this article, we propose a generic framework to make seasonal time series forecasting. The framework combines machine learning techniques (1) to identify the typical seasons and (2) to forecast the likelihood of having a season type in one season ahead. We study in[13] this framework by comparing the mean squared errors of forecasts for various settings and various datasets. The best setting is then compared to state-of-the-art time series forecasting methods. We show that it is competitive with them.

7.1.19. Smooth Adversarial Examples

Participants: Hanwei Zhang, Yannis Avrithis, Teddy Furon, Laurent Amsaleg.

This paper investigates the visual quality of the adversarial examples. Recent papers propose to smooth the perturbations to get rid of high frequency artefacts. In this work, smoothing has a different meaning as it perceptually shapes the perturbation according to the visual content of the image to be attacked [44]. The perturbation becomes locally smooth on the flat areas of the input image, but it may be noisy on its textured areas and sharp across its edges. This operation relies on Laplacian smoothing, well-known in graph signal processing, which we integrate in the attack pipeline. We benchmark several attacks with and without smoothing under a white-box scenario and evaluate their transferability. Despite the additional constraint of smoothness, our attack has the same probability of success at lower distortion.

7.1.20. Walking on the Edge: Fast, Low-Distortion Adversarial Examples

Participants: Hanwei Zhang, Yannis Avrithis, Teddy Furon, Laurent Amsaleg.

Adversarial examples of deep neural networks are receiving ever increasing attention because they help in understanding and reducing the sensitivity to their input. This is natural given the increasing applications of deep neural networks in our everyday lives. When white-box attacks are almost always successful, it is typically only the distortion of the perturbations that matters in their evaluation. In this work [45], we argue that speed is important as well, especially when considering that fast attacks are required by adversarial training. Given more time, iterative methods can always find better solutions. We investigate this speed-distortion trade-off in some depth and introduce a new attack called boundary projection (BP) that improves upon existing methods by a large margin. Our key idea is that the classification boundary is a manifold in the image space: we therefore quickly reach the boundary and then optimize distortion on this manifold.

7.1.21. Accessing watermarking information: Error exponents in the noisy case

Participant: Teddy Furon.

The study of the error exponents of zero-bit watermarking is addressed in the article by Comesana, Merhav, and Barni, under the assumption that the detector relies solely on second order joint empirical statistics of the received signal and the watermark. This restriction leads to the well-known dual hypercone detector, whose score function is the absolute value of the normalized correlation. They derive the false negative error exponent and the optimum embedding rule. However, they only focus on high SNR regime, i.e. the noiseless scenario. This work extends this theoretical study to the noisy scenario. It introduces a new definition of watermarking robustness based on the false negative error exponent, derives this quantity for the dual hypercone detector, and shows that its performances is almost equal to Costa's lower bound [22].

7.1.22. Detecting fake news and image forgeries

Participants: Cédric Maigrot, Vincent Claveau, Ewa Kijak.

Social networks make it possible to share information rapidly and massively. Yet, one of their major drawback comes from the absence of verification of the piece of information, especially with viral messages. Based on the work already presented in the previous years, C. Maigrot defended his thesis on the detection of image forgeries, classification of reinformation websites, and on the late fusion of models based on the text, image and source analysis [1]. This work was also given a large visibility thanks to numerous interviews in Press and TV (see the dedicated section about popularization).

7.1.23. Learning Interpretable Shapelets for Time Series Classification through Adversarial Regularization

Times series classification can be successfully tackled by jointly learning a shapelet-based representation of the series in the dataset and classifying the series according to this representation. However, although the learned shapelets are discriminative, they are not always similar to pieces of a real series in the dataset. This makes it difficult to interpret the decision, i.e. difficult to analyze if there are particular behaviors in a series that triggered the decision. In this work [29], we make use of a simple convolutional network to tackle the time series classification task and we introduce an adversarial regularization to constrain the model to learn more interpretable shapelets. Our classification results on all the usual time series benchmarks are comparable with the results obtained by similar state-of-the-art algorithms but our adversarially regularized method learns shapelets that are, by design, interpretable.

7.1.24. Using Knowledge Base Semantics in Context-Aware Entity Linking

Participants: Cheikh Brahim El Vaigh, Guillaume Gravier, Pascale Sébillot.

Done as part of the IPL iCODA, in collaboration with CEDAR Inria team.

Entity linking is a core task in textual document processing, which consists in identifying the entities of a knowledge base (KB) that are mentioned in a text. Approaches in the literature consider either independent linking of individual mentions or collective linking of all mentions. Regardless of this distinction, most approaches rely on the Wikipedia encyclopedic KB in order to improve the linking quality, by exploiting its entity descriptions (web pages) or its entity interconnections (hyperlink graph of web pages). We devised a novel collective linking technique which departs from most approaches in the literature by relying on a structured RDF KB [9]. This allows exploiting the semantics of the interrelationships that candidate entities may have at disambiguation time rather than relying on raw structural approximation based on Wikipedia's hyperlink, graph. The few approaches that also use an RDF KB simply rely on the existence of a relation between the candidate entities to which mentions may be linked. Instead, we weight such relations based on the RDF KB structure and propose an efficient decoding strategy for collective linking. Experiments on standard benchmarks show significant improvement over the state of the art.

7.1.25. Neural-based lexico-syntactic relation extraction in news archives

Participants: Guillaume Gravier, Cyrielle Mallart, Pascale Sébillot.

Done as part of the IPL iCODA, in collaboration with Ouest France

Relation extraction is the task of finding and classifying the relationship between two entities in a text. We pursued work on the detection of relations between entities, seen as a binary classification problem. In the context of large-scale news archives, we argue that detection is paramount before even considering classification, where most approaches consider the two tasks jointly with a null garbage class. This does hardly allow for the detection of relations for unseen categories, which are all considered as garbage. We designed a bi-LSTM sequence neural model acting on features extracted from the surface realization, the part-of-speech tags and the dependency parse tree and compared with a state-of-the-art relation detection LSTM-based approach. Experimental evaluations rely on a dataset derived from 200k Wikipedia articles in French containing 4M linked mentions of entities: 330k pairs of entities co-occur in the same sentence, of which 1 % are actual relations according to Wikidata. Results show the benefit of our binary detection approach over previous methods and over joint detection and classification.

7.1.26. Graph Convolutional Networks for Learning with Few Clean and Many Noisy Labels

Participants: Ahmet Iscen [Google Research], Giorgos Tolias [Univ. Prague], Yannis Avrithis, Ondra Chum [Univ. Prague], Cordelia Schmid [Google Research].

In this work we consider the problem of learning a classifier from noisy labels when a few clean labeled examples are given [39]. The structure of clean and noisy data is modeled by a graph per class and Graph Convolutional Networks (GCN) are used to predict class relevance of noisy examples. For each class, the GCN is treated as a binary classifier learning to discriminate clean from noisy examples using a weighted binary cross-entropy loss function, and then the GCN-inferred "clean" probability is exploited as a relevance measure. Each noisy example is weighted by its relevance when learning a classifier for the end task. We evaluate our method on an extended version of a few-shot learning problem, where the few clean examples of novel classes are supplemented with additional noisy data. Experimental results show that our GCN-based cleaning process significantly improves the classification accuracy over not cleaning the noisy data and standard few-shot classification where only few clean examples are used. The proposed GCN-based method outperforms the transductive approach (Douze et al., 2018) that is using the same additional data without labels.

7.1.27. Rethinking deep active learning: Using unlabeled data at model training

Participants: Oriane Siméoni, Mateusz Budnik, Yannis Avrithis, Guillaume Gravier.

Active learning typically focuses on training a model on few labeled examples alone, while unlabeled ones are only used for acquisition. In this work we depart from this setting by using both labeled and unlabeled data during model training across active learning cycles [42]. We do so by using unsupervised feature learning at the beginning of the active learning pipeline and semi-supervised learning at every active learning cycle, on all available data. The former has not been investigated before in active learning, while the study of latter in the context of deep learning is scarce and recent findings are not conclusive with respect to its benefit. Our idea is orthogonal to acquisition strategies by using more data, much like ensemble methods use more models. By systematically evaluating on a number of popular acquisition strategies and datasets, we find that the use of unlabeled data during model training brings a spectacular accuracy improvement in image classification, compared to the differences between acquisition strategies. We thus explore smaller label budgets, even one label per class.

7.1.28. Training Object Detectors from Few Weakly-Labeled and Many Unlabeled Images

Participants: Zhaohui Yang [Peking University], Miaojing Shi, Yannis Avrithis, Chao Xu [Peking University], Vittorio Ferrari [Google Research].

Weakly-supervised object detection attempts to limit the amount of supervision by dispensing the need for bounding boxes, but still assumes image-level labels on the entire training set are available. In this work, we study the problem of training an object detector from one or few clean images with image-level labels and a larger set of completely unlabeled images [43]. This is an extreme case of semi-supervised learning where the labeled data are not enough to bootstrap the learning of a classifier or detector. Our solution is to use a standard weakly-supervised pipeline to train a student model from image-level pseudo-labels generated on the unlabeled set by a teacher model, bootstrapped by region-level similarities to clean labeled images. By using

the recent pipeline of PCL and more unlabeled images, we achieve performance competitive or superior to many state of the art weakly-supervised detection solutions.

7.2. Accessing Information

7.2.1. Ontological modeling of human reading experience

Participants: Guillaume Gravier, Pascale Sébillot.

Done as part of the JPI CH READ-IT projects, in collaboration with Open University (UK) and Université Le Mans (FR)

Diaries, correspondence and authors' libraries provide important evidence into the evolution of ideas and society. Studying these phenomena is connected to understanding changes of perspective and values. Within the framework of the READ-IT project, we developed an ontological data approach modelling changes in the contents of diaries, correspondence and authors' libraries related to reading. By considering these three types of sources, we designed a conceptual data model to permit the study and increase the usability of sources containing evidence of reading experiences, highlighting common challenges and patterns related to changes to readers and to the medium of reading when confronting historical events [36], [8].

7.2.2. Integration of Exploration and Search: A Case Study of the M^3 Model

Participants: Snorri Gíslason [IT Univ. Copenhagen], Björn Þór Jónsson [IT Univ. Copenhagen], Laurent Amsaleg.

Effective support for multimedia analytics applications requires exploration and search to be integrated seamlessly into a single interaction model. Media metadata can be seen as defining a multidimensional media space, casting multimedia analytics tasks as exploration, manipulation and augmentation of that space. We present an initial case study of integrating exploration and search within this multidimensional media space [11]. We extend the M^3 model, initially proposed as a pure exploration tool, and show that it can be elegantly extended to allow searching within an exploration context and exploring within a search context. We then evaluate the suitability of relational database management systems, as representatives of today's data management technologies, for implementing the extended M^3 model. Based on our results, we finally propose some research directions for scalability of multimedia analytics.

7.2.3. Exquisitor: Breaking the Interaction Barrier for Exploration of 100 Million Images

Participants: Hanna Ragnarsdóttir [Reykjavik University], Þórhildur Þorleiksdóttir [Reykjavik University], Omar Shahbaz Khan [IT Univ. Copenhagen], Björn Þór Jónsson [IT Univ. Copenhagen], Gylfi Þór Gudmundsson [School of Computer Science, Reykjavik], Jan Zahálka [bohem.ai], Stevan Rudinac [University of Amsterdam], Laurent Amsaleg, Marcel Worring [University of Amsterdam].

We present Exquisitor, a media explorer capable of learning user preferences in real-time during interactions with the 99.2 million images of YFCC100M. Exquisitor owes its efficiency to innovations in data representation, compression, and indexing. Exquisitor can complete each interaction round, including learning preferences and presenting the most relevant results, in less than 30 ms using only a single CPU core and modest RAM. In short, Exquisitor can bring large-scale interactive learning to standard desktops and laptops, and even high-end mobile devices [16].

MAGRIT Team

7. New Results

7.1. Matching and localization

Participants: Marie-Odile Berger, Vincent Gaudilliere, Gilles Simon, Frédéric Sur, Matthieu Zins.

7.1.1. View synthesis for efficient and accurate pose computation

Estimating the pose of a camera from a scene model is a challenging problem when the camera is in a position not covered by the views used to build the model, because feature matching is difficult. Several viewpoint simulation techniques have been recently proposed in this context. They generally come with a high computational cost, are limited to specific scenes such as urban environments or object-centred scenes, or need an initial pose guess. A new method based on viewpoint simulation is presented in [15]. In this article, we show that view synthesis dramatically improves pose computation and that both the synthesis process and pose computation can be done in a very efficient way. Two major problems are especially addressed: the positioning of the virtual viewpoints with respect to the scene, and the synthesis of geometrically consistent patches. Experiments show that patch synthesis dramatically improves the accuracy of the pose in case of difficult registration, with a limited computational cost.

7.1.2. Localization from objects

We are interested in AR applications which take place in man-made GPS-denied environments, such as industrial or indoor scenes. In such environments, relocalization may fail due to repeated patterns and large changes in appearance which occur even for small changes in viewpoint. During this year, we have investigated a new method for relocalization which operates at the level of objects and takes advantage of the impressive progress realized in object detection. Recent works have opened the way towards object oriented reconstruction from elliptic approximation of objects detected in images. We have gone beyond that and have proposed a new method for pose computation based on ellipse/ellipsoid correspondences. In [18], we have proved that a closed form estimate of the translation can be uniquely inferred from the rotation matrix of the pose. When two or more correspondences are available, the rotation matrix is deduced through an optimization problem with three degrees of freedom. However, the pose cannot be uniquely computed from one correspondence. In [19], we consider the practical common case where an initial guess of the rotation matrix of the pose is known, for instance with an inertial sensor or from the estimation of orthogonal vanishing points [10]. The translation is recovered as in [18], [24]. We proved the effectiveness of the method on real scenes from a set of object detections generated by YOLO [33]. Globally, considering pose at the level of objects allows us to avoid common failures due to repeated structures. In addition, due to the small combinatorics induced by object correspondences, our method is well suited to fast rough localization even in large environments.

A patent was filed on this method in May 2019 [27]. An Inria technological transfer action (ATT) on the subject of object based localization will start in January 2020 with the aim to produce a demonstrator for industrial maintenance in complex environments.

7.2. Handling non-rigid deformations

Participants: Marie-Odile Berger, Jaime Garcia Guevara, Erwan Kerrien, Daryna Panicheva, Raffaella Trivisonne, Pierre-Frédéric Villard.

7.2.1. Compliance-based non rigid registration

Within J. Guevara's PhD thesis, we are investigating non rigid registration methods which exploit the matching of the vascular trees and are able to cope with large deformations of the organ. This year, we have developed a matching method which is entirely based on the mechanical properties of the organ. We thus avoid tedious parameter tuning which is required by many methods and instead use parameters whose values are known or can be measured. Our method makes use of an advanced biomechanical model which handles heterogeneities and anisotropy due to vasculature. The main originality of the method lies in the definition of a better and novel metric for generating improved graph-matching hypotheses, based on the notion of compliance, the inverse of stiffness. This method reduces the computation time by predicting first the most plausible matching hypotheses on a mechanical basis and reduces the sensitivity on the search space parameters. These contributions improve the registration quality and meet intra-operative timing constraints. Experiments have been conducted on ten realistic synthetic datasets and two real porcine datasets which where automatically segmented. This work was recently accepted in the journal *Annals of Biomedical Engineering* [9], [11].

7.2.2. Individual-specific heart valve modeling

Recent works on computer-based models of mitral valve behavior rely on manual extraction of the complex valve geometry, which is tedious and requires a high level of expertise. On the contrary, in the context of D. Panicheva's PhD thesis, we are investigating methods to segment the chordae with little human supervision which produce mechanically-coherent simulations of the mitral valve.

Valve chordae are generalized cylinders: Instead of being limited to a line, the central axis is a continuous curve; instead of a constant radius, the radius varies along the axis. Most of the time, chordae sections are flattened ellipses and classical model-based methods commonly used for vessel enhancement or vessel segmentation fail. We have exploited the fact that there are no other generalized cylinders than the chordae in the CT scan and we have proposed a topology-based method for chordae extraction. This approach is flexible and only requires the knowledge of an upper bound of the maximum radius of the chordae. The method has been tested on three CT scans. Overall, non-chordae structures are correctly identified and detected chordae ending points match up with actual chordae attachment points [21].

We then worked on evaluating the effectiveness of our approach. The valve behavior was simulated with a biomechanical framework based on the Finite Element Method. A structural model with no fluid-structure interaction was used. Physiological behavior was simulated by mechanical forces such as blood pressure, contact forces and tension forces applied from chordae tensions. The chordae segmentation was validated by comparing the simulation results to those obtained with manually segmented chordae [22].

7.2.3. Image-based biomechanical simulation of the diaphragm during mechanical ventilation

When intensive care patients are subjected to mechanical ventilation, the ventilator causes damage to the muscles that govern the normal breathing, leading to Ventilator Induced Diaphragmatic Dysfunction (VIDD). The INVIVE project aims to study the mechanics of respiration through numerical simulation in order to learn more about the onset of VIDD. We have worked during this year on how to compute solutions of the static linear elasticity equation using last year's work on the diaphragm geometry [26]. Since obtaining an analytical formulation of the boundary conditions in 3D is complex, we have worked on adapting our method to implicit geometries built from 2D data of the diaphragm. The idea is to have an analytical formulation of both the geometry and the boundary conditions to validate our radial basis framework. It is based on points belonging to a cross-section that has been chosen in the middle of the diaphragm. Points are gathered in groups inside rectangles based on a K-means classification. Rectangle dimensions are set so as to ensure cross-coverage. Curve patches are then computed for each rectangle using radial basis functions. A list of local curves is obtained from both the thoracic and abdomen zones and by combining them it is possible to evaluate the global implicit curve of the diaphragm.

7.2.4. 3D catheter navigation from monocular images

In interventional radiology, the 3D shape of the micro-tool (guidewire, micro-catheter or micro-coil) can be very difficult, if not impossible to infer from fluoroscopy images. We consider this question as a single view

3D curve reconstruction problem. Our aim is to assess whether, and under which conditions, a sophisticated physics-based model can be effective to compensate for the incomplete data in this ill-posed problem.

Raffaella Trivisonne started her PhD thesis in November 2015 (co-supervised by Stéphane Cotin, from MIMESIS team in Strasbourg) to address this research topic. An unscented Kalman filter is used as a fusion mechanism, in a non-rigid shape-from-motion approach: the observations are image data (opaque markers placed along the device), and the model is implemented through interactive physics-based simulation. Our contribution is to handle contacts, which introduce discontinuities in the first and second order derivatives of motion (resp. velocity and forces). Extensive validation on both synthetic and phantom-based data has been carried out this year [30], and various state vector parametrizations have been investigated, in particular in a view to achieve data assimilation of mechanical parameters to improve the predictability of simulation.

In this context, validation is made very complex by the need to acquire ground truth 3D curve shapes that are subjected to contacts and demonstrate highly transient dynamic deformations (e.g. stick and slip transitions after contact). Thomas Mangin was hired on a 1-year engineer contract (started in March 2019) to design and develop an experimental platform to acquire such ground truth data. The catheter is inserted in a translucent, silicon vascular phantom to generate contacts with no visual occlusion of the catheter shape. It is reconstructed from images acquired by a stereo rig made of two orthogonal high speed cameras. The motion is fully controlled by an original 3D-printed active device that induces accurate translation and rotation motions to the micro-tool. Monte-Carlo simulations are currently being carried out to certify the accuracy of the ground truth data produced by this system.

7.3. Image processing

Participants: Marie-Odile Berger, Fabien Pierre, Frédéric Sur.

7.3.1. Computational photomechanics

In computational photomechanics, mainly two methods are available for estimating displacement and strain fields on the surface of a material specimen subjected to a mechanical test, namely digital image correlation (DIC) and localized spectrum analysis (LSA). With both methods, a contrasted pattern marks the surface of the specimen: either a random speckle pattern for DIC or a regular pattern for LSA, this latter method being based on Fourier analysis. It is a challenging problem since strains are tiny quantities giving deformations often not visible to the naked eye. The recent outcomes of our collaboration with Institut Pascal (Université Clermont-Auvergne) focus on two areas.

We have investigated the optimization of the pattern marking the specimen [13], which is the topic of several recent papers. Checkerboard is the optimized pattern in terms of sensor noise propagation when the signal is correctly sampled, but its periodicity causes convergence issues with DIC. The consequence is that checkerboards are not used in DIC applications although they are optimal in terms of sensor noise propagation. We have shown that it is possible to use LSA to estimate displacement and strain fields from checkerboard images, although LSA was originally designed to process 2D grid images. A comparative study of checkerboards and grids shows that, under similar experimental conditions, the noise level in displacement and strain maps obtained with checkerboards is lower than that obtained with classic 2D grids. A patent on this topic was filed [28].

Another scientific contribution concerns the restoration of displacement and strain maps. DIC and LSA both provide displacement fields equal to the actual one convoluted by a kernel known a priori. The kernel indeed corresponds to the Savitzky-Golay filter in DIC, and to the analysis window of the windowed Fourier transform used in LSA. While convolution reduces noise level, it also gives a systematic measurement error. We have proposed a deconvolution method to retrieve the actual displacement and strain fields from the output of DIC or LSA [12]. The proposed algorithm can be considered as a variant of Van Cittert deconvolution, based on the small strain assumption. It is demonstrated that it allows enhancing fine details in displacement and strain maps, while improving spatial resolution.

7.3.2. Cartoon-texture decomposition

Decomposing an image as the sum of geometric and textural components is a popular problem of image analysis. In this problem, known as cartoon and texture decomposition, the cartoon component is piecewise smooth, made of the geometric shapes of the images, and the texture component is made of stationary or quasi-stationary oscillatory patterns filling the shapes. Microtextures being characterized by their power spectrum, we propose to extract cartoon and texture components from the information provided by the power spectrum of image patches. The contribution of texture to the spectrum of a patch is detected as statistically significant spectral components with respect to a null hypothesis modeling the power spectrum of a non-textured patch. The null-hypothesis model is built upon a coarse cartoon representation obtained by a basic yet fast filtering algorithm of the literature. The coarse decomposition is obtained in the spatial domain and is an input of the proposed spectral approach. We thus design a "dual domain" method. The statistical model is also built upon the power spectrum of patches with similar textures across the image. The proposed approach therefore falls within the family of non-local methods. Compared to variational methods or fast filers, the proposed non-local dual-domain approach [16] is shown to achieve a good compromise between computation time and accuracy. Matlab code is publicly available.

7.3.3. Variational methods for image processing

The work described in [20] aims to couple the powerful prediction of the convolutional neural network (CNN) to the accuracy at pixel scale of the variational methods. We have focused on a CNN which is able to compute a statistical distribution of the colors for each pixel of the image based on a learning stage on a large color image database. A variational method able to select a color candidate among a given set while performing regularization of the result is combined with a CNN, to design a fully automatic image colorization framework with an improved accuracy in comparison with CNN alone. To solve the proposed model, we have proposed in [17] a novel accelerated alternating optimization scheme to solve block biconvex nonsmooth problems whose objectives can be split into smooth (separable) regularizers and simple coupling terms. The proposed method performs a Bregman distance-based generalization of the well-known forward–backward splitting for each block, along with an inertial strategy which aims at getting empirical acceleration. We discuss the theoretical convergence of the proposed scheme and provide numerical experiments on image colorization.

MORPHEO Project-Team

7. New Results

7.1. Surface Motion Capture Animation Synthesis

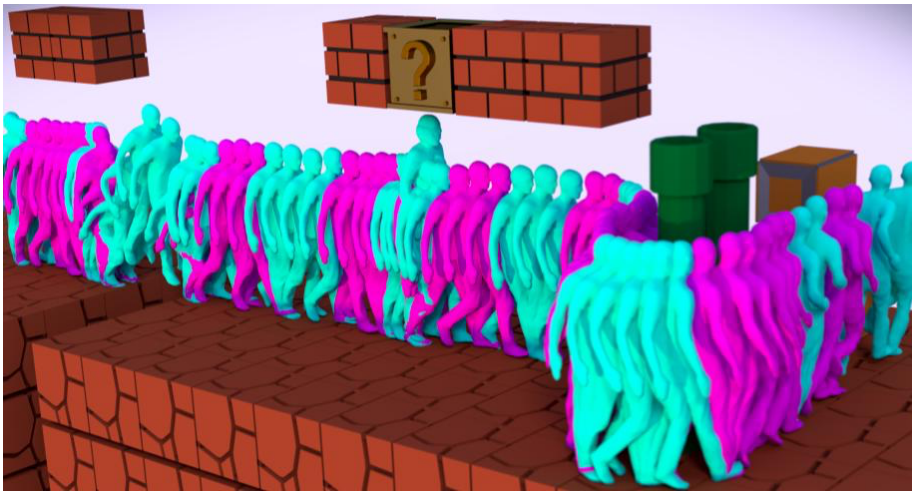


Figure 2. Animation Synthesis

We propose to generate novel animations from a set of elementary examples of video-based surface motion capture, under user-specified constraints. 4D surface capture animation is motivated by the increasing demand from media production for highly realistic 3D content. To this aim, data driven strategies that consider video-based information can produce animation with real shapes, kinematics and appearances. Our animations rely on the combination and the interpolation of textured 3D mesh data, which requires examining two aspects: (1) Shape geometry and (2) appearance. First, we propose an animation synthesis structure for the shape geometry, the Essential graph, that outperforms standard Motion graphs in optimality with respect to quantitative criteria, and we extend optimized interpolated transition algorithms to mesh data. Second, we propose a compact view-independent representation for the shape appearance. This representation encodes subject appearance changes due to viewpoint and illumination, and due to inaccuracies in geometric modelling independently. Besides providing compact representations, such decompositions allow for additional applications such as interpolation for animation (see figure 2).

This result was published in a prominent computer graphics journal, IEEE Transactions on Visualization and Computer Graphics [7].

7.2. CBCT of a Moving Sample from X-rays and Multiple Videos

We consider dense volumetric modeling of moving samples such as body parts. Most dense modeling methods consider samples observed with a moving X-ray device and cannot easily handle moving samples. We propose instead a novel method to observe shape motion from a fixed X-ray device and to build dense in-depth attenuation information. This yields a low-cost, low-dose 3D imaging solution, taking benefit of equipment widely available in clinical environments. Our first innovation is to combine a video-based surface motion

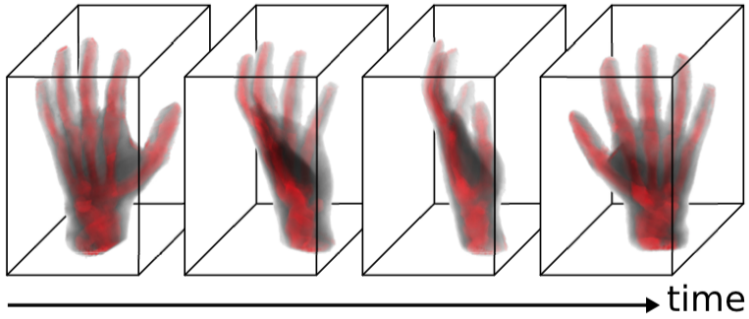


Figure 3. Dense volumetric attenuation reconstruction from a rigidly moving sample captured by a single planar X-ray imaging device and a surface motion capture system. Higher attenuation (here bone structure) is highlighted in red.

capture system with a single low-cost/low-dose fixed planar X-ray device, in order to retrieve the sample motion and attenuation information with minimal radiation exposure. Our second innovation is to rely on Bayesian inference to solve for a dense attenuation volume given planar radioscopic images of a moving sample. This approach enables multiple sources of noise to be considered and takes advantage of very limited prior information to solve an otherwise ill-posed problem. Results show that the proposed strategy is able to reconstruct dense volumetric attenuation models from a very limited number of radiographic views over time on synthetic and in-situ data, as illustrated in Figure 3 .

This result was published in a prominent medical journal, IEEE Transactions on Medical Imaging [9].

7.3. Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D sequences

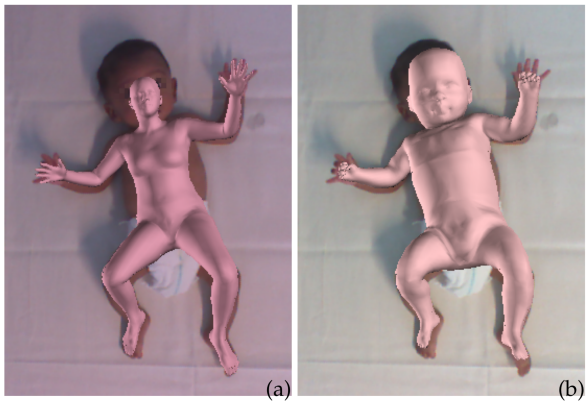


Figure 4. (a) Simply scaling a generic adult body model and fitting it to an infant does not work as body proportions significantly differ. (b) The proposed SMIL model properly captures the infants' shape and pose

Statistical models of the human body surface are generally learned from thousands of high-quality 3D scans in predefined poses to cover the wide variety of human body shapes and articulations. Acquisition of such data requires expensive equipment, calibration procedures, and is limited to cooperative subjects who can understand and follow instructions, such as adults. We presented a method for learning a statistical 3D Skinned Multi-Infant Linear body model (SMIL) from incomplete, low-quality RGB-D sequences of freely moving infants. Quantitative experiments show that SMIL faithfully represents the RGB-D data and properly factorizes the shape and pose of the infants. To demonstrate the applicability of SMIL, we fitted the model to RGB-D sequences of freely moving infants and show, with a case study, that our method captures enough motion detail for General Movements Assessment (GMA), a method used in clinical practice for early detection of neurodevelopmental disorders in infants. SMIL provides a new tool for analyzing infant shape and movement and is a step towards an automated system for GMA. This result was published in a prominent computer vision journal, IEEE Transactions on PAMI [8].

7.4. The Virtual Caliper: Rapid Creation of Metrically Accurate Avatars from 3D Measurements

Figure 5. Using the wand controllers of the HTC Vive, the Virtual Caliper produces a rigged 3D model with exactly the dimensions of the measured person.

Creating metrically accurate avatars is important for many applications such as virtual clothing try-on, ergonomics, medicine, immersive social media, telepresence, and gaming. Creating avatars that precisely represent a particular individual is challenging however, due to the need for expensive 3D scanners, privacy issues with photographs or videos, and difficulty in making accurate tailoring measurements. We overcome these challenges by creating "The Virtual Caliper", which uses VR game controllers to make simple measurements. First, we establish what body measurements users can reliably make on their own body. We find several distance measurements to be good candidates and then verify that these are linearly related to 3D body shape as represented by the SMPL body model. The Virtual Caliper enables novice users to accurately measure themselves and create an avatar with their own body shape. We evaluate the metric accuracy relative to ground truth 3D body scan data, compare the method quantitatively to other avatar creation tools, and perform extensive perceptual studies. We also provide a software application to the community that enables novices to rapidly create avatars in fewer than five minutes. Not only is our approach more rapid than existing methods, it exports a metrically accurate 3D avatar model that is rigged and skinned.

This result was published in a prominent computer graphics journal, IEEE Transactions on Visualization and Computer Graphics [10].

7.5. Adaptive Mesh Texture for Multi-View Appearance Modeling

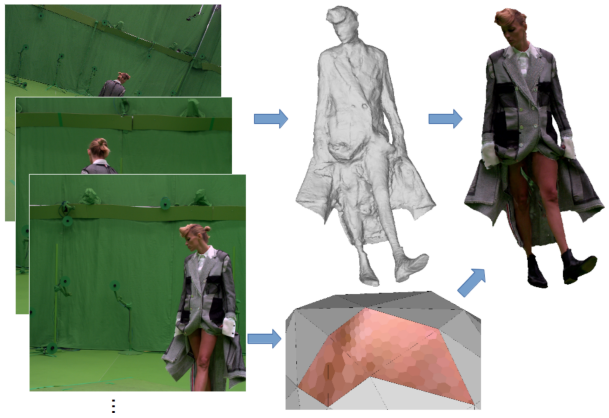


Figure 6. Texturing 3D models: given a set of input photographs (left), a geometric mesh is computed (top), along with an appearance function stored within the surface mesh structure (bottom).

Most applications in image based 3D modeling resort to texture maps, a 2D mapping of shape color information into image files. Despite their unquestionable merits, in particular the ability to apply standard image tools, including compression, image textures still suffer from limitations that result from the 2D mapping of information that originally belongs to a 3D structure. This is especially true with 2D texture atlases, a generic 2D mapping for 3D mesh models that introduces discontinuities in the texture space and plagues many 3D appearance algorithms. Moreover, the per-triangle texel density of 2D image textures cannot be individually adjusted to the corresponding pixel observation density without a global change in the atlas mapping function. To address these issues, we have proposed a new appearance representation for image-based 3D shape modeling, which stores appearance information directly on 3D meshes, rather than a texture atlas. We have shown this representation to allow for input-adaptive sampling and compression support. Our experiments demonstrated that it outperforms traditional image textures, in multi-view reconstruction contexts, with better visual quality and memory foot- print, which makes it a suitable tool when dealing with large amounts of data as with dynamic scene 3D models.

This result was published in the international conference on 3D Vision (3DV'19) [11].

7.6. Contact Preserving Shape Transfer for Motion Retargeting

Retargeting a motion from a source to a target character is an important problem in computer animation, as it allows to reuse existing rigged databases or transfer motion capture to virtual characters. Surface based pose transfer is a promising approach to avoid the trial-and-error process when controlling the joint angles. In this work we investigated whether shape transfer instead of pose transfer would better preserve the original contextual meaning of the source pose. To this end, we proposed an optimization-based method to deform the source shape+pose using three main energy functions: similarity to the target shape, body part volume preservation, and collision management (preserve existing contacts and prevent penetrations). The results show that this strategy is able to retarget complex poses, including several contacts, to very different morphologies. In particular, we introduced new contacts that are linked to the change in morphology, and which would be difficult to obtain with previous works based on pose transfer that aim at distance preservation between body parts.

This result was published in the ACM SIGGRAPH Conference on Motion Interaction and Games (MIG'19) [12].

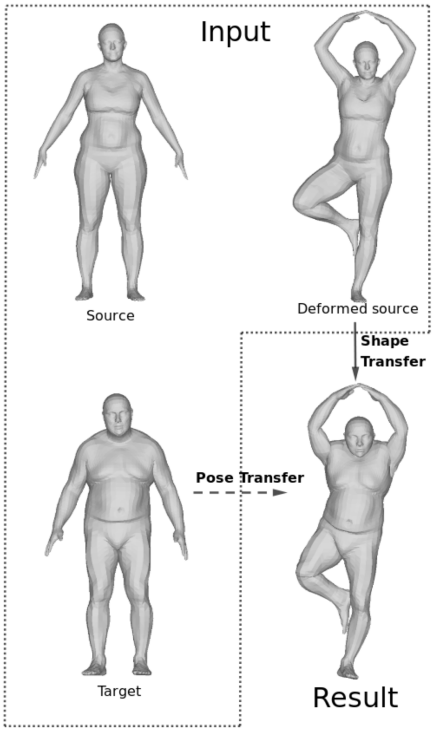


Figure 7. Motion Retargeting: Instead of transferring the pose from a source to a target shape, we propose to transfer the shape of the target to the deformed source character.

7.7. A Decoupled 3D Facial Shape Model by Adversarial Training

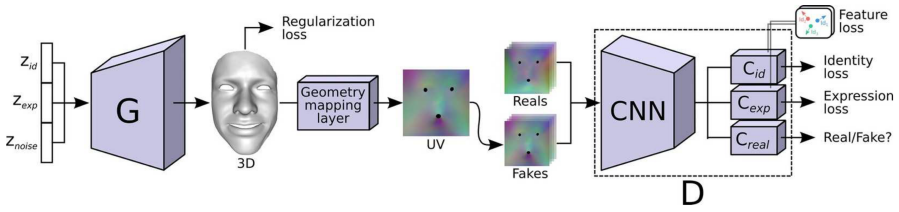


Figure 8. The face generator. Identity and expression codes zid , zexp are used to control the generator, and classification losses are added to decouple between the two. A feature loss is introduced to ensure consistency over features with fixed identities or expressions

Data-driven generative 3D face models are used to compactly encode facial shape data into meaningful parametric representations. A desirable property of these models is their ability to effectively decouple natural sources of variation, in particular identity and expression. While factorized representations have been proposed for that purpose, they are still limited in the variability they can capture and may present modeling artifacts when applied to tasks such as expression transfer. In this work, we explored a new direction with Generative Adversarial Networks and showed that they contribute to better face modeling performances, especially in decoupling natural factors, while also achieving more diverse samples. To train the model we introduced a novel architecture that combines a 3D generator with a 2D discriminator that leverages conventional CNNs, where the two components are bridged by a geometry mapping layer. We further presented a training scheme, based on auxiliary classifiers, to explicitly disentangle identity and expression attributes. Through quantitative and qualitative results on standard face datasets, we illustrated the benefits of our model and demonstrate that it outperforms competing state of the art methods in terms of decoupling and diversity.

This result was published in the international conference on computer vision (ICCV'19) [13]

7.8. Non-parametric 3D Human Shape Estimation from Single Images

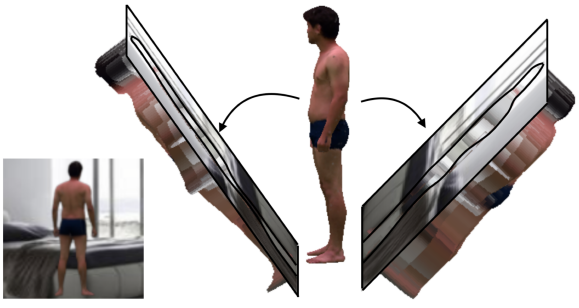


Figure 9. Given a single image, we estimate the "visible" and the "hidden" depth maps from the camera point of view. The two depth maps can be seen as the two halves of a virtual "mould".

In this work, we tackle the problem of 3D human shape estimation from single RGB images. While the recent progress in convolutional neural networks has allowed impressive results for 3D human pose estimation, estimating the full 3D shape of a person is still an open issue. Model-based approaches can output precise meshes of naked under-cloth human bodies but fail to estimate details and un-modelled elements such as hair or clothing. On the other hand, non-parametric volumetric approaches can potentially estimate complete shapes but, in practice, they are limited by the resolution of the output grid and cannot produce detailed estimates. In this work, we propose a non-parametric approach that employs a double depth map to represent the 3D shape of a person: a visible depth map and a "hidden" depth map are estimated and combined, to reconstruct the human 3D shape as done with a "mould". This representation through 2D depth maps allows a higher resolution output with a much lower dimension than voxel-based volumetric representations. Additionally, our fully derivable depth-based model allows us to efficiently incorporate a discriminator in an adversarial fashion to improve the accuracy and "humanness" of the 3D output. We train and quantitatively validate our approach on SURREAL and on 3D-HUMANS, a new photorealistic dataset made of semi-synthetic in-house images annotated with 3D ground truth surfaces.

This work was published in the international conference on computer vision (ICCV'19) [14]

7.9. Probabilistic Reconstruction Networks

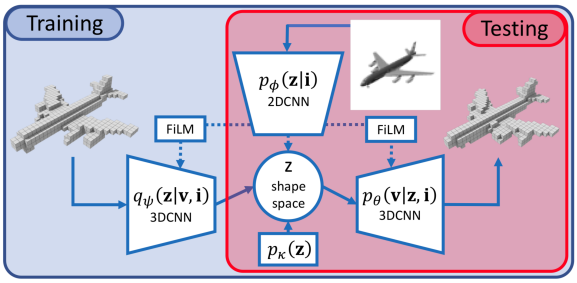


Figure 10. Probabilistic Reconstruction Networks for 3D shape inference from a single image. Arrows show the computational flow through the model, dotted arrows show optional image conditioning. The inference network q_Ψ is only used during training for variational inference

We study end-to-end learning strategies for 3D shape inference from images, in particular from a single image. Several approaches in this direction have been investigated that explore different shape representations and suitable learning architectures. We focus instead on the underlying probabilistic mechanisms involved and contribute a more principled probabilistic inference-based reconstruction framework, which we coin Probabilistic Reconstruction Networks. This framework expresses image conditioned 3D shape inference through a family of latent variable models, and naturally decouples the choice of shape representations from the inference itself. Moreover, it suggests different options for the image conditioning and allows training in two regimes, using either Monte Carlo or variational approximation of the marginal likelihood. Using our Probabilistic Reconstruction Networks we obtain single image 3D reconstruction results that set a new state of the art on the ShapeNet dataset in terms of the intersection over union and earth mover's distance evaluation metrics. Interestingly, we obtain these results using a basic voxel grid representation, improving over recent work based on finer point cloud or mesh based representations.

This work was published in the British machine vision conference (BMVC'19) [15] where it won the runner-up best paper award.

PERCEPTION Project-Team

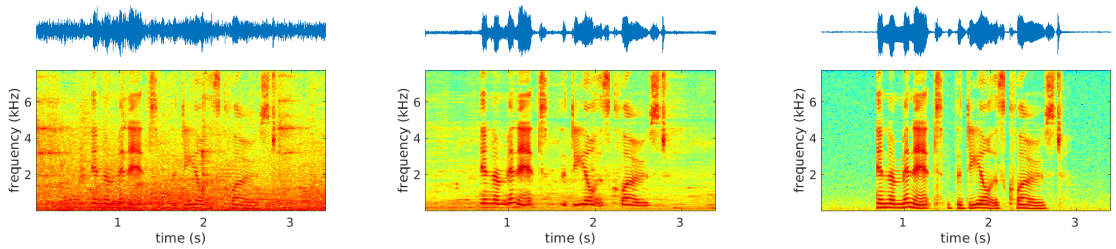
6. New Results

6.1. Multichannel Speech Separation and Enhancement Using the Convolutive Transfer Function

We addressed the problem of speech separation and enhancement from multichannel convolutive and noisy mixtures, *assuming known mixing filters*. We proposed to perform the speech separation and enhancement tasks in the short-time Fourier transform domain, using the convolutive transfer function (CTF) approximation [43], [44]. Compared to time-domain filters, CTF has much less taps, consequently it has less near-common zeros among channels and less computational complexity. The work proposes three speech-source recovery methods, namely: (i) the multichannel inverse filtering method, i.e. the multiple input/output inverse theorem (MINT), is exploited in the CTF domain, and for the multi-source case, (ii) a beamforming-like multichannel inverse filtering method applying single source MINT and using power minimization, which is suitable whenever the source CTFs are not all known, and (iii) a constrained Lasso method, where the sources are recovered by minimizing the ℓ_1-norm to impose their spectral sparsity, with the constraint that the ℓ_2-norm fitting cost, between the microphone signals and the mixing model involving the unknown source signals, is less than a tolerance. The noise can be reduced by setting a tolerance onto the noise power. Experiments under various acoustic conditions are carried out to evaluate the three proposed methods. The comparison between them as well as with the baseline methods is presented.

6.2. Speech Denoising and Enhancement with LTSMs

We have started to address the problems of multichannel speech denoising [45] and enhancement [51] in the short-time Fourier transform (STFT) domain and in the framework of sequence-to-sequence deep learning. In the case of denoising, the magnitude of noisy speech is mapped onto the noise power spectral density. In the case of speech enhancement, the noisy speech is mapped onto clean speech. A long short-time memory (LSTM) network takes as input a sequence of STFT coefficients associated with a frequency bin of multichannel noisy-speech signals. The network's output is a sequence of single-channel cleaned speech at the same frequency bin. We propose several clean-speech network targets, namely, the magnitude ratio mask, the complex ideal ratio mask, the STFT coefficients and spatial filtering [54]. A prominent feature of the proposed model is that the same LSTM architecture, with identical parameters, is trained across frequency bins. The proposed method is referred to as narrow-band deep filtering. This choice stays in contrast with traditional wide-band speech enhancement methods. The proposed deep filter is able to discriminate between speech and noise by exploiting their different temporal and spatial characteristics: speech is non-stationary and spatially coherent while noise is relatively stationary and weakly correlated across channels. This is similar in spirit with unsupervised techniques, such as spectral subtraction and beamforming. We describe extensive experiments with both mixed signals (noise is added to clean speech) and real signals (live recordings). We empirically evaluate the proposed architecture variants using speech enhancement and speech recognition metrics, and we compare our results with the results obtained with several state of the art methods. In the light of these experiments we conclude that narrow-band deep filtering has very good performance, and excellent generalization capabilities in terms of speaker variability and noise type, e.g. Figure 2 .
Website: https://team.inria.fr/perception/research/mse-lstm/.



Noisy speech input Filtered output Ground-truth speech signal

Figure 2. An example of narrow-band deep filtering for speech enhancement [54]. Waveforms and spectrograms of the noisy (unprocessed) input, the filtered output and the ground-truth clean-speech. Four microphones were used in this example. The signal-to-noise ratio in this example is 0 dB.

6.3. Multichannel Speech Enhancement with Variational Auto-Encoder

We addressed speaker-independent multichannel speech enhancement in unknown noisy environments. Our work is based on a well-established multichannel local Gaussian modeling framework. We propose to use a neural network for modeling the speech spectro-temporal content. The parameters of this supervised model are learned using the framework of variational autoencoders. The noisy recording environment is supposed to be unknown, so the noise spectro-temporal modeling remains unsupervised and is based on non-negative matrix factorization (NMF). We develop a Monte Carlo expectation-maximization algorithm and we experimentally show that the proposed approach outperforms its NMF-based counterpart, where speech is modeled using supervised NMF [49].
Website: https://team.inria.fr/perception/research/icassp-2019-mvae/

6.4. Audio-visual Speech Enhancement with Conditional Variational Auto-Encoder

Variational auto-encoders (VAEs) are deep generative latent variable models that can be used for learning the distribution of complex data. VAEs have been successfully used to learn a probabilistic prior over speech signals, which is then used to perform speech enhancement. One advantage of this generative approach is that it does not require pairs of clean and noisy speech signals at training. In this work, we propose audio-visual variants of VAEs for single-channel and speaker-independent speech enhancement. We developed a conditional VAE (CVAE) where the audio speech generative process is conditioned on visual information of the lip region, e.g. Figure 3 . At test time, the audio-visual speech generative model is combined with a noise model, based on nonnegative matrix factorization, and speech enhancement relies on a Monte Carlo expectation-maximization algorithm. Experiments were conducted with the recently published NTCD-TIMIT dataset. The results confirm that the proposed audio-visual CVAE effectively fuse audio and visual information, and it improves the speech enhancement performance compared with the audio-only VAE model, especially when the speech signal is highly corrupted by noise. We also showed that the proposed unsupervised audio-visual speech enhancement approach outperforms a state-of-the-art supervised deep learning method [55].
Website: https://team.inria.fr/perception/research/av-vae-se/

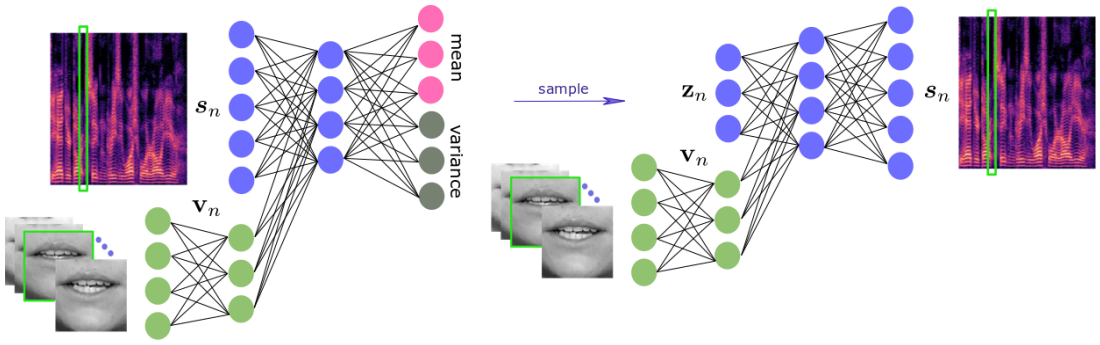


Figure 3. We proposed a conditional variational auto-encoder architecture for fusing audio and visual data for speech enhancement [55].

6.5. Variational Bayesian Inference of Audio-visual Speaker Tracking

We addressed the problem of tracking multiple speakers via the fusion of visual and auditory information [36]. We proposed to exploit the complementary nature of these two modalities in order to accurately estimate smooth trajectories of the tracked persons, to deal with the partial or total absence of one of the modalities over short periods of time, and to estimate the acoustic status – either speaking or silent – of each tracked person along time, e.g. Figure 1 . We proposed to cast the problem at hand into a generative audio-visual fusion (or association) model formulated as a latent-variable temporal graphical model. This may well be viewed as the problem of maximizing the posterior joint distribution of a set of continuous and discrete latent variables given the past and current observations, which is intractable. We proposed a variational inference model which amounts to approximate the joint distribution with a factorized distribution. The solution takes the form of closed-form expectation maximization procedures using Gaussian distributions [38]. We described in detail the inference algorithm, we evaluated its performance and we compared the results with several baseline methods. These experiments show that the proposed audio-visual tracker performs well in informal meetings involving a time-varying number of people. Real-time versions of the algorithm were implemented on our robotic platform [47].
Website: https://team.inria.fr/perception/research/var-av-track/.

6.6. Detection, Localization and Tracking of Multiple Audio Sources

We addressed the problem of online detection, localization and tracking of multiple moving speakers in reverberant environments [36]. The work has the following contributions. We used the direct-path relative transfer function (DP-RTF), an inter-channel feature that encodes acoustic information robust against reverberation, and we proposed an online algorithm well suited for estimating DP-RTFs associated with moving audio sources. Another crucial ingredient of the proposed method is its ability to properly assign DP-RTFs to audio-source directions. Towards this goal, we adopted a maximum-likelihood formulation and we proposed to use the exponentiated gradient (EG) to efficiently update source-direction estimates starting from their currently available values. The problem of multiple-speaker tracking is computationally intractable because the number of possible associations between observed source directions and physical speakers grows exponentially with time. We adopt a Bayesian framework and we proposed two variational approximations of the posterior filtering distributions associated with multiple speaker tracking, as well as two efficient variational expectation maximization (VEM) solvers [41], [37]. The proposed online localization and tracking methods were thoroughly evaluated using two datasets that contain recordings performed in real environments.
Websites:

https://team.inria.fr/perception/research/audiotrack-vonm/
https://team.inria.fr/perception/research/multi-speaker-tracking/.

6.7. The Kinovis Multiple-Speaker Tracking Datasets

The Kinovis multiple speaker tracking (Kinovis-MST) datasets contain live acoustic recordings of multiple moving speakers in a reverberant environment. The data were recorded in the Kinovis multiple-camera laboratory at Inria Grenoble Rhône-Alpes. The room size is $10.2 \times 9.9 \times 5.6$ meters with T60 = 0.53 seconds. The data were recorded with four microphones embedded into the head of a NAO robot. Because there is a fan located inside the robot head nearby the microphones, there is a fair amount of stationary and spatially correlated microphone noise. The signal-to-noise ratio of the microphone signals is of approximately 2.7 dB. The recordings contain between one and three moving participants that speak naturally, hence the number of active speech sources varies over time. The robot-to-speaker distance ranges between 1.5 and 3.5 meters. Ground-truth trajectories and speech activity information were obtained in the following way. Participants were wearing optical markers placed on their heads such that the Kinovis motion capture system provides accurate 3D trajectories for each participant. Moreover, an infrared marker is placed on the participants' foreheads. This enables the identification of each participant over time. Whenever time a participant is silent, he/she hides his/her infrared marker, thus allowing speaking/silent annotations of the recordings.
Website: https://team.inria.fr/perception/the-kinovis-mst-dataset/.

6.8. Deep Regression

Deep learning revolutionized data science, and recently its popularity has grown exponentially, as did the amount of papers employing deep networks. Vision tasks, such as human pose estimation, did not escape from this trend. There is a large number of deep models, where small changes in the network architecture, or in the data pre-processing, together with the stochastic nature of the optimization procedures, produce notably different results, making extremely difficult to sift methods that significantly outperform others. This situation motivates the current study, in which we perform a systematic evaluation and statistical analysis of vanilla deep regression, i.e. convolutional neural networks with a linear regression top layer. This is the first comprehensive analysis of deep regression techniques. We perform experiments on four vision problems, and report confidence intervals for the median performance as well as the statistical significance of the results, if any. Surprisingly, the variability due to different data pre-processing procedures generally eclipses the variability due to modifications in the network architecture. Our results reinforce the hypothesis according to which, in general, a general-purpose network (e.g. VGG-16 or ResNet-50) adequately tuned can yield results close to the state-of-the-art without having to resort to more complex and ad-hoc regression models, [40].
Website: https://team.inria.fr/perception/research/deep-regression/.

6.9. Deep Reinforcement Learning for Audio-Visual Robot Control

More recently, we investigated the use of reinforcement learning (RL) as an alternative to sensor-based robot control. The robotic task consists of turning the robot head (gaze control) towards speaking people. The method is more general in spirit than visual (or audio) servoing because it can handle an arbitrary number of speaking or non speaking persons and it can improve its behavior online, as the robot experiences new situations. An overview of the proposed method is shown in Fig. 4 . The reinforcement learning formulation enables a robot to learn where to look for people and to favor speaking people via a trial-and-error strategy.

Past, present and future HRI developments require datasets for training, validation, test as well as for benchmarking. HRI datasets are challenging because it is not easy to record realistic interactions between a robot and users. RL avoids systematic recourse to annotated datasets for training. In [39] we proposed the use of a simulated environment for pre-training the RL parameters, thus avoiding spending hours of tedious interaction.
Website: https://team.inria.fr/perception/research/deep-rl-for-gaze-control/.

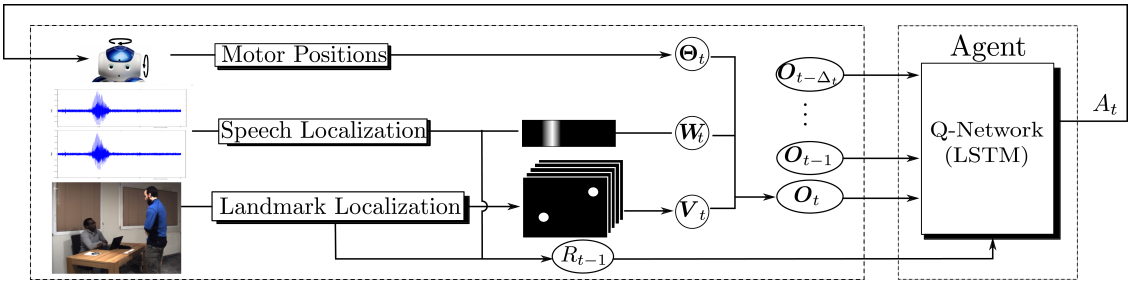


Figure 4. Overview of the proposed deep RL method for controlling the gaze of a robot. At each time index t, audio and visual data are represented as binary maps which, together with motor positions, form the set of observations $\mathbf{O}_t$. A motor action A_t (rotate the head left, right, up, down, or stay still) is selected based on past and present observations via maximization of current and future rewards. The rewards R are based on the number of visible persons as well as on the presence of speech sources in the camera field of view. We use a deep Q-network (DQN) model that can be learned both off-line and on-line. Please consult [39] for further details.

SIROCCO Project-Team

7. New Results

7.1. Visual Data Analysis

Scene depth, Scene flows, 3D modeling, Light-fields, 3D point clouds

7.1.1. Scene depth estimation from light fields

Participants: Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

While there exist scene depth estimation methods, these methods, mostly designed for stereo content or for pairs of rectified views, do not effectively apply to new imaging modalities such as light fields. We have focused on the problem of *scene depth estimation* for every viewpoint of a dense light field, exploiting information from only a sparse set of views [24]. This problem is particularly relevant for applications such as light field reconstruction from a subset of views, for view synthesis, for 3D modeling and for compression. Unlike most existing methods, the proposed algorithm computes disparity (or equivalently depth) for every viewpoint taking into account occlusions. In addition, it preserves the continuity of the depth space and does not require prior knowledge on the depth range.

We have then proposed a learning based depth estimation framework suitable for both densely and sparsely sampled light fields. The proposed framework consists of three processing steps: initial depth estimation, efficient fusion with occlusion handling and refinement. The estimation can be performed from a flexible subset of input views. The fusion of initial disparity estimates, relying on two warping errors measures, allows us to have an accurate estimation in occluded regions and along the contours. The use of trained neural networks has the advantage of a limited computational cost at estimation time. In contrast with methods relying on the computation of cost volumes, the proposed approach does not need any prior information on the disparity range. Experimental results show that the proposed method outperforms state-of-the-art light fields depth estimation methods for a large range of baselines [15].

The training of the proposed neural networks based architecture requires having ground truth disparity (or depth) maps. Although a few synthetic datasets exist for dense light fields with ground truth depth maps, no such dataset exists for sparse light fields with large baselines. This lack of training data with ground truth depth maps is a crucial issue for supervised learning of neural networks for depth estimation. We therefore created two datasets, namely SLFD and DLFD, containing respectively sparsely sampled and densely sampled synthetic light fields. To our knowledge, SLFD is the first available dataset providing sparse light field views and their corresponding ground truth depth and disparity maps. The created datasets have been made publicly available together with the code and the trained models.

7.1.2. Scene flow estimation from light fields

Participants: Pierre David, Christine Guillemot.

We have addressed the problem of scene flow estimation from sparsely sampled video light fields. Scene flows can be seen as 3D extensions of optical flows by also giving the variation in depth along time in addition to the optical flow. Scene flows are tools needed for temporal processing of light fields. Estimating dense scene flows in light fields poses obvious problems of complexity due to the very large number of rays or pixels. This is even more difficult when the light field is sparse, i.e., with large disparities, due to the problem of occlusions. The developments in this area are also made difficult due to the lack of test data, i.e., there is no publicly available synthetic video light fields with the corresponding ground truth scene flows. In order to be able to assess the performance of the proposed method, we have therefore created synthetic video light fields from the MPI Sintel dataset. This video light field data set has been produced with the Blender software by creating new production files placing multiple cameras in the scene, controlling the disparity between the set of views.

We have then developed a local 4D affine model to represent scene flows, taking into account light field epipolar geometry. The model parameters are estimated per cluster in the 4D ray space. We have first developed a sparse to dense estimation method that avoids the difficulty of computing matches in occluded areas [18], which we have further extended by developing a dense scene flow estimation method from light fields. The local 4D affine parameters are in this case derived by fitting the model on initial motion and disparity estimates obtained by using 2D dense optical flow estimation techniques.

We have shown that the model is very effective for estimating scene flows from 2D optical flows (see Fig.2). The model regularizes the optical flows and disparity maps, and interpolates disparity variation values in occluded regions. The proposed model allows us to benefit from deep learning-based 2D optical flow estimation methods while ensuring scene flow geometry consistency in the 4 dimensions of the light field.

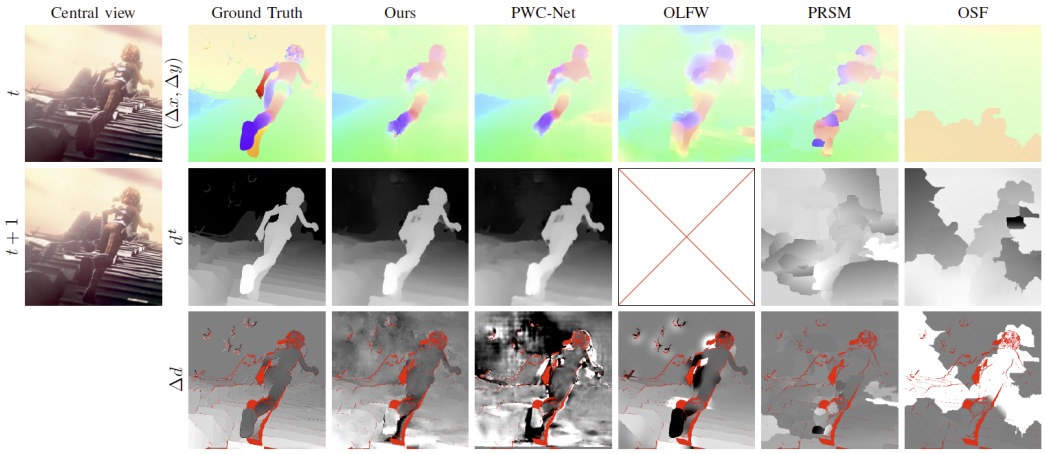


Figure 2. Visual comparison of our method with respect to reference methods (PWC-Net: deep learning method for optical flow estimation; oriented light field window (OLFW), Piece-wise Rigid Scene Model (PRSM), Object Scene Flow (OSF)). First row: optical flows; Second row: disparity maps; Third row: disparity variations. The red pixels are the occlusion mask where there is no ground truth disparity variation available.

7.1.3. Depth estimation at the decoder in the MPEG-I standard

Participants: Patrick Garus, Christine Guillemot, Thomas Maugey.

This study, in collaboration with Orange labs., addresses several downsides of the system under development in MPEG-I for coding and transmission of immersive media. We study a solution, which enables Depth-Image-Based Rendering for immersive video applications, while lifting the requirement of transmitting depth information. Instead, we estimate the depth information on the client-side from the transmitted views. We have observed that doing this leads to a significant rate saving (37.3% in average). Preserving perceptual quality in terms of MS-SSIM of synthesized views, it yields to 24.6% rate reduction for the same quality of reconstructed views after residue transmission under the MPEG-I common test conditions. Simultaneously, the required pixel rate, i.e. the number of pixels processed per second by the decoder, is reduced by 50% for any test sequence [22].

7.1.4. Spherical feature extraction for 360 light field reconstruction from omni-directional fish-eye camera captures

Participants: Christine Guillemot, Fatma Hawary, Thomas Maugey.

With the increasing interest in wide-angle or 360° scene captures, the extraction of descriptors well suited to the geometry of this content is a key problem for a variety of processing tasks. Algorithms designed for feature extraction in 2D images are hardly applicable to 360° images or videos as they do not well take into account their specific spherical geometry. To cope with this difficulty, it is quite common to perform an equirectangular projection of the spherical content, and to compute spherical features on projected and stitched content. However, this process introduces geometrical distortions with implications on the accuracy of applications such as angle estimation, depth calculation and 3D scene reconstruction. We adapt a spherical feature descriptor to the geometry of fish-eye cameras that avoids equirectangular projection. The captured image is directly mapped onto a spherical model of the 360° camera. In order to evaluate the interest of the proposed fish-eye adapted descriptor, we consider the angular coordinates of feature points on the sphere. We assess the stability of the corresponding angles when capturing the scene by a moving fish-eye camera. Experimental results show that the proposed fish-eye adapted descriptor allows a more stable angle estimation, hence a more robust feature detection, compared to spherical features on projected and stitched contents.

7.2. Signal processing and learning methods for visual data representation and compression

Sparse representation, data dimensionality reduction, compression, scalability, rate-distortion theory

7.2.1. Single sensor light field acquisition using coded masks

Participants: Christine Guillemot, Ehsan Miandji, Hoai Nam Nguyen.

We developed a simple variational approach for reconstructing color light fields in the compressed sensing framework with very low sampling ratio, using both coded masks and color filter arrays (CFA). A coded mask is placed in front of the camera sensor to optically modulate incoming rays, while a color filter array is assumed to be implemented at the sensor level to compress color information. Hence, the light field coded projections, operated by a combination of the coded mask and the CFA, measure incomplete color samples with a three times lower sampling ratio than reference methods that assume full color (channel-by-channel) acquisition. We then derived adaptive algorithms to directly reconstruct the light field from raw sensor measurements by minimizing a convex energy composed of two terms. The first one is the data fidelity term which takes into account the use of CFAs in the imaging model, and the second one is a regularization term which favors the sparse representation of light fields in a specific transform domain. Experimental results show that the proposed approach produces a better reconstruction both in terms of visual quality and quantitative performance when compared to reference reconstruction methods that implicitly assume prior color interpolation of coded projections.

We then pursued this study by developing a unifying image formation model that abstracts the architecture of most existing compressive-sensing light-field cameras, equipped with single lens and coded masks, as an equivalent multi-mask camera. It allows to compare different designs with a number of criteria: compression rate, light efficiency, measurement incoherence, as well as acquisition quality. Moreover, the underlying multi-mask camera can be flexibly adapted for various applications, such as single and multiple acquisitions, spatial super-resolution, parallax reconstruction, and color restoration. We also derived a generic variational algorithm solving all these concrete problems by considering appropriate sampling operators.

7.2.2. 3D point cloud processing and plenoptic point cloud compression

Participants: Christian Galea, Christine Guillemot, Maja Krivokuca.

Light fields, by capturing light rays emitted by a 3D scene along different orientations, give a very rich description of the scene enabling a variety of computer vision applications. The recorded 4D light field gives in particular information about the parallax and depth of the scene. The estimated depth can then be used to construct 3D models of the scene, e.g. in the form of a 3D point cloud, The constructed 3D point clouds, however, generally contain distortions and artefacts primarily caused by inaccuracies in the depth maps. We have developed a method for noise removal in 3D point clouds constructed from light fields [21]. While existing methods discard outliers, the proposed approach instead attempts to correct the positions of points,

and thus reduce noise without removing any points, by exploiting the consistency among views in a light-field. The proposed 3D point cloud construction and denoising method exploits uncertainty measures on depth values.

Beyond classical 3D point clouds, plenoptic point clouds can be seen as natural extensions of 3D point clouds to Surface Light Fields (SLF). While the concept of surface light field (SLF) has been introduced as a function that assigns a color to each ray originating on a surface, plenoptic point clouds represent in each voxel illumination and color seen from different camera viewpoints. In other words, instead of each point being associated with a single colour value, there can be multiple values to represent the colour at that point as perceived from different viewpoints. This concept aims at combining the best of light fields and computer graphics modeling, for photo-realistic rendering from arbitrary points of view. However, this representation leads to color maps per voxel, hence to large volumes of data. We have addressed the problem of efficient compression of this data based on the Region-Adaptive Hierarchical Transform (RAHT) method in which we have introduced clustering and speculat/diffuse components separation showing better adapted plenoptic point cloud color maps transforms.

7.2.3. Low-rank models and representations for light fields

Participants: Elian Dib, Christine Guillemot, Xiaoran Jiang.

We have addressed the problem of light field dimensionality reduction. We have introduced a local low-rank approximation method using a parameteric disparity model. The local support of the approximation is defined by super-rays. Superrays can be seen as a set of super-pixels that are coherent across all light field views. The light field low-rank assumption depends on how much the views are correlated, i.e. on how well they can be aligned by disparity compensation. We have therefore introduced a disparity estimation method using a low-rank prior. We have considered a parametric model describing the local variations of disparity within each super-ray, and alternatively search for the best parameters of the disparity model and of the low-rank approximation. We have assessed the proposed disparity parametric model, by considering an affine disparity model. We have shown that using the proposed disparity parametric model and estimation algorithm gives an alignment of superpixels across views that favours the low-rank approximation compared with using disparity estimated with classical computer vision methods. The low-rank matrix approximation is then computed on the disparity compensated super-rays using a singular value decomposition (SVD). A coding algorithm has been developed for the different components of the proposed disparity-compensated low-rank approximation [20].

We have also, in collaboration with Trinity College Dublin, introduced a new Light Field representation for efficient Light Field processing and rendering called Fourier Disparity Layers (FDL) [12]. The proposed FDL representation samples the Light Field in the depth (or equivalently the disparity) dimension by decomposing the scene as a discrete sum of layers. The layers can be constructed from various types of Light Field inputs including a set of sub-aperture images, a focal stack, or even a combination of both. From our derivations in the Fourier domain, the layers are simply obtained by a regularized least square regression performed independently at each spatial frequency, which is efficiently parallelized in a GPU implementation. Our model is also used to derive a gradient descent based calibration step that estimates the input view positions and an optimal set of disparity values required for the layer construction. Once the layers are known, they can be simply shifted and filtered to produce different viewpoints of the scene while controlling the focus and simulating a camera aperture of arbitrary shape and size. A direct implementation in the Fourier domain allows real time Light Field rendering. Finally, direct applications such as view interpolation or extrapolation and denoising have also been evaluated [12]. The use of this representation for view synthesis based compression has also been assessed in [19].

7.2.4. Graph-based transforms and prediction for light fields

Participants: Christine Guillemot, Thomas Maugey, Mira Rizkallah.

We have investigated Graph-based transforms for low dimensional embedding of light field data. Both non separable and separable transforms have been considered. The low-dimensional embedding can be learned with a few eigen vectors of the graph Laplacian. However, the dimension of the data (e.g. light fields) has obvious implications on the storage footprint of the Laplacian matrix and on the eigenvectors computation complexity, making graph-based non separable transforms impractical for such data. To cope with this difficulty, we have developed local super-rays based non separable and separable (spatial followed by angular) weighted and unweighted transforms to jointly capture light fields correlation spatially and across views [14]. Despite the local support of limited size defined by the super-rays, the Laplacian matrix of the non separable graph remains of high dimension and its diagonalization to compute the transform eigen vectors remains computationally expensive. To solve this problem, we have then performed the local spatio-angular transform in a separable manner.

Separable transforms on super-rays allow us to significantly decrease the eigenvector computation complexity. However, the basis functions of the spatial graph transforms to be applied on the super-ray pixels of each view are often not compatible. We have indeed shown that when the shape of corresponding super-pixels in the different views is not isometric, the basis functions of the spatial transforms are not coherent, resulting in decreased correlation between spatial transform coefficients, hence in a loss of performance of the angular transform, compared to the non-separable case. We have therefore developed a graph construction optimization procedure which seeks to find the eigen-vectors which align the best with those of a reference one while still approximately diagonalizing their respective Laplacians [14]. The proposed optimization method aims at preserving angular correlation even when the shapes of the super-pixels are not isometric. Experimental results show the benefit of the approach in terms of energy compaction. A coding scheme has also been developed to assess the rate-distortion perfomances of the proposed transforms

The use of local transforms with limited supports is a way to cope with the computational difficulty. Unfortunately, the locality of the support may not allow us to fully exploit long term signal dependencies present in both the spatial and angular dimensions in the case of light fields. We have therefore introduced sampling and prediction schemes, based on graph sampling theory, with local graph-based transforms enabling to efficiently compact the signal energy and exploit dependencies beyond the local graph support [31], [13]. The proposed approach has been shown to be very efficient in the context of spatio-angular transforms for quasi-lossless compression of light fields.

7.2.5. Intra-coding of 360-degree images on the sphere

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

Omni-directional images are characterized by their high resolution (usually 8K) and therefore require high compression efficiency. Existing methods project the spherical content onto one or multiple planes and process the mapped content with classical 2D video coding algorithms. However, this projection induces sub-optimality. Indeed, after projection, the statistical properties of the pixels are modified, the connectivity between neighboring pixels on the sphere might be lost, and finally, the sampling is not uniform. Therefore, we propose to process uniformly distributed pixels directly on the sphere to achieve high compression efficiency. In particular, a scanning order and a prediction scheme are proposed to exploit, directly on the sphere, the statistical dependencies between the pixels. A Graph Fourier Transform is also applied to exploit local dependencies while taking into account the 3D geometry. Experimental results demonstrate that the proposed method provides up to 5.6% bitrate reduction and on average around 2% bitrate reduction over state-of-the-art methods. This work has led to a publication in the PCS conference 2019 [26].

7.3. Algorithms for inverse problems in visual data processing

Inpainting, view synthesis, super-resolution

7.3.1. View synthesis in light fields and stereo set-ups

Participants: Simon Evain, Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

We have developed a learning-based framework for light field view synthesis from a subset of input views. Building upon a light-weight optical flow estimation network to obtain depth maps, our method employs two reconstruction modules in pixel and feature domains respectively. For the pixel-wise reconstruction, occlusions are explicitly handled by a disparity-dependent interpolation filter, whereas inpainting on disoccluded areas is learned by convolutional layers. Due to disparity inconsistencies, the pixel-based reconstruction may lead to blurriness in highly textured areas as well as on object contours. On the contrary, the feature-based reconstruction performs well on high frequencies, making the reconstruction in the two domains complementary. End-to-end learning is finally performed including a fusion module merging pixel and feature-based reconstructions. Experimental results show that our method achieves state-of-the-art performance on both synthetic and real-world datasets, moreover, it is even able to extend light fields baseline by extrapolating high quality views without additional training.

We have also designed a very lightweight neural network architecture, trained on stereo data pairs, which performs view synthesis from one single image [7]. With the growing success of multi-view formats, this problem is indeed increasingly relevant. The network returns a prediction built from disparity estimation, which fills in wrongly predicted regions using a occlusion handling technique. To do so, during training, the network learns to estimate the left-right consistency structural constraint on the pair of stereo input images, to be able to replicate it at test time from one single image. The method is built upon the idea of blending two predictions: a prediction based on disparity estimation, and a prediction based on direct minimization in occluded regions. The network is also able to identify these occluded areas at training and at test time by checking the pixelwise left-right consistency of the produced disparity maps. At test time, the approach can thus generate a left-side and a right-side view from one input image, as well as a depth map and a pixelwise confidence measure in the prediction. The work outperforms visually and metric-wise state-of-the-art approaches on the challenging KITTI dataset, all while reducing by a very significant order of magnitude (5 or 10 times) the required number of parameters (6.5 M).

7.3.2. Inverse problems in light field imaging with 4D anisotropic diffusion and neural networks

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

We have addressed inverse problems in ligt field imaging by following two methodological directions. We first introduced a 4D anisotropic diffusion framework based on PDEs [4]. The proposed regularization method operated in the 4D ray space and, unlike the methods operating on epipolar plane images, does not require prior estimation of disparity maps. The method performs a PDE-based diffusion with anisotropy steered by a tensor field based on local structures in the 4D ray space that we extract using a 4D tensor structure. To enhance coherent structures, the smoothing along directions, surfaces, or volumes in the 4D ray space is performed along the eigenvectors directions. Although anisotropic diffusion is well understood for 2D imaging, its interpretation and understanding in the 4D space is far from being straightforward. We have analysed the behaviour of the diffusion process on a light field toy example, i.e. a tesseract (a 4D cube). This simple light field example allows an in-depth analysis of how each eigenvector influences the diffusion process. The proposed ray space regularizer is a tool that has enabled us to tackle a variety of inverse problems (denoising, angular and spatial interpolation, regularization for enhancing disparity estimation as well as inpainting) in the ray space.

In collaboration with the university of Malta (Pr. Reuben Farrugia), we have explored the benefit of low-rank priors in light field super-resolution with deep neural networks. This led us to design a learning-based spatial light field super-resolution method that allows the restoration of the entire light field with consistency across all sub-aperture images [8]. The algorithm first uses optical flows to align the light field views and then reduces its angular dimension using low-rank approximation. We then consider the linearly independent columns of the resulting low-rank model as an embedding, which is restored using a deep convolutional neural network. The super-resolved embedding is then used to reconstruct the remaining sub-aperture images. The original disparities are restored using inverse warping where missing pixels are approximated using a novel light field inpainting algorithm. We pursued this study by designing an approach that, thanks to a low-rank approximation model, can leverage models learned for 2D image super-resolution [9]. This approach avoids the need for a

large amount of light field training data which is, unlike 2D images, not available. It also allows us to reduce the dimension, hence the number of parameters, of the network to be learned.

7.3.3. Neural networks for axial light field super-resolution

Participants: Christine Guillemot, Zhaolin Xiao.

Axial light field resolution refers to the ability to distinguish features at different depths by refocusing. The axial refocusing precision corresponds to the minimum distance in the axial direction between two distinguishable refocusing planes. High refocusing precision can be essential for some light field applications like microscopy. We first introduced a refocusing precision model based on a geometrical analysis of the flow of rays within the virtual camera. The model establishes the relationship between the feature distinguishability by refocusing and different camera settings. We have then developed a learning-based method to extrapolate novel views from axial volumes of sheared epipolar plane images (EPIs (see an example of extrapolated views in Fig.3)). As extended numerical aperture (NA) in classical imaging, the extrapolated light field gives refocused images with a shallower depth of field (DOF), leading to more accurate refocusing results. Most importantly, the proposed approach does not need accurate depth estimation. Experimental results with both synthetic and real light fields, including with microscopic data, demonstrate that our approach can effectively enhance the light field axial refocusing precision.

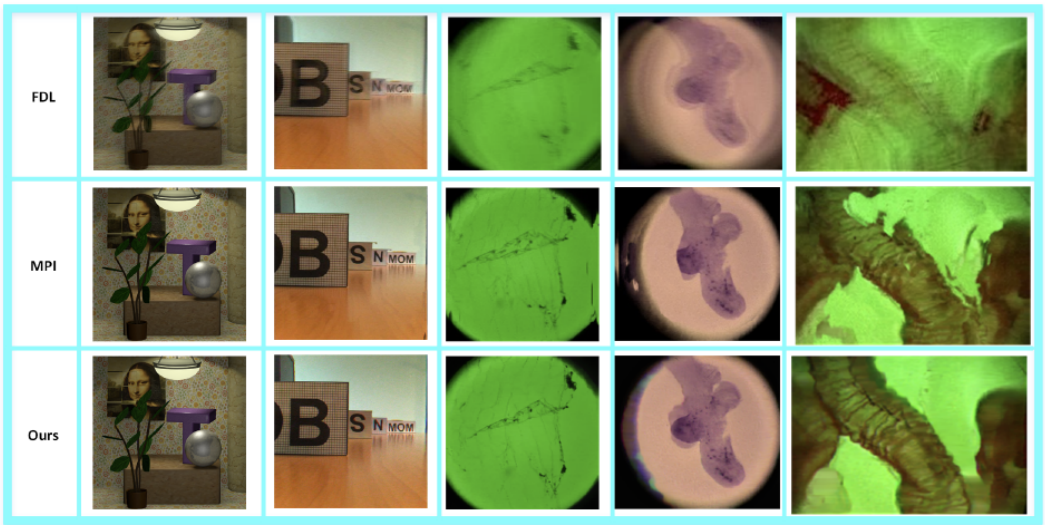


Figure 3. Extrapolation results with a 4X larger baseline, in comparison with reference methods using multiple plane images (MPI) and Fourier disparity layers (FDL).

7.3.4. Neural networks for inverse problems in 2D imaging

Participants: Christine Guillemot, Aline Roumy, Alexander Sagel.

The Deep Image Prior has been recently introduced to solve inverse problems in image processing with no need for training data other than the image itself. However, the original training algorithm of the Deep Image Prior constrains the reconstructed image to be on a manifold described by a convolutional neural network. For some problems, this neglects prior knowledge and can render certain regularizers ineffective. We have developed an alternative approach that relaxes this constraint and fully exploits all prior knowledge. We have evaluated our algorithm on the problem of reconstructing a high-resolution image from a downsampled version and observed a significant improvement over the original Deep Image Prior algorithm.

7.4. Distributed coding for interactive communication

Information theory, stochastic modeling, robust detection, maximum likelihood estimation, generalized likelihood ratio test, error and erasure resilient coding and decoding, multiple description coding, Slepian-Wolf coding, Wyner-Ziv coding, information theory, MAC channels

7.4.1. Interactive compression scheme for interactive media

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

We propose a new interactive compression scheme for omnidirectional images and 3D model. This requires two characteristics: efficient compression of data, to lower the storage cost, and random access ability to extract part of the compressed stream requested by the user (for reducing the transmission rate). For efficient compression, data needs to be predicted by a series of references that have been pre-defined and compressed. This contrasts with the spirit of random accessibility. We propose a solution for this problem based on incremental codes implemented by rate adaptive channel codes. This scheme encodes the image while adapting to any user request and leads to an efficient coding that is flexible in extracting data depending on the available information at the decoder. Therefore, only the information which is needed to be displayed at the user's side is transmitted during the user's request as if the request was already known at the encoder (see Fig. 4). The experimental results demonstrate that our coder obtains a better transmission rate than the state-of-the-art tile-based methods at a small cost in storage. Moreover, the transmission cost grows gradually with the size of the request and avoids a staircase effect, which shows the perfect suitability of our coder for interactive transmission. This work has led to a journal submission and several conference publications. In [25], we have proposed a new framework for evaluating the compression performance of interactive schemes. Indeed, interactive compression schemes can be characterized by tree criteria: the storage cost, the transmission rate and distortion. This contrasts with classical compression scheme, where only transmission rate and distortion are used. 3D-performance evaluation criteria are proposed. In [29], we have proposed to use the geometry to efficiently compress the 3D mesh texture. An interactive coding extension has been presented in [27].

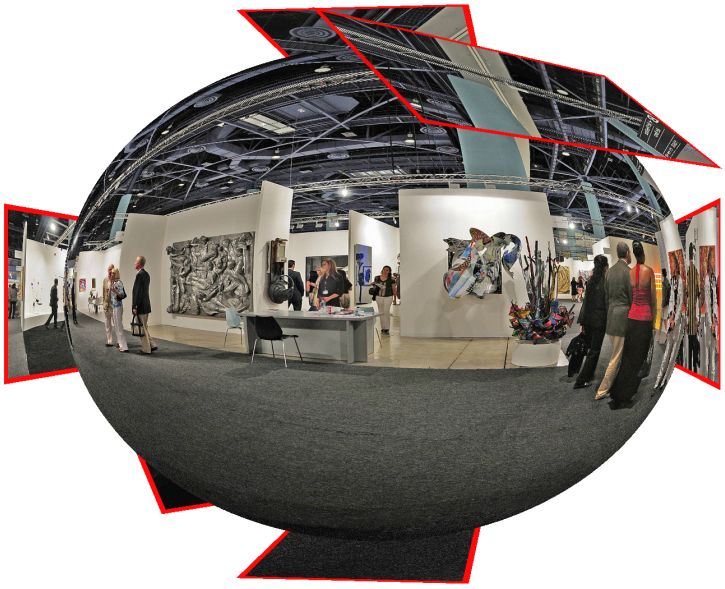


Figure 4. A spherical image and several viewports corresponding to different user's requests.

7.4.2. Reference source positioning for interactive compression

Participants: Thomas Maugey, Mai Quyen Pham, Aline Roumy.

Large databases containing many HD videos or records from sensors over long time intervals, have to be efficiently compressed, to reduce their size. The compression has also to allow efficient access to random parts of the databases upon request from the users. Efficient compression is usually achieved with prediction between data points. However, this creates dependencies between the compressed representations, which is contrary to the idea of random access. Prediction methods rely in particular on reference data points, used to predict other data points, and the placement of these references balances compression efficiency and random access. Existing solutions to position the references use ad hoc methods. We study this joint problem of compression efficiency and random access. We introduce the storage cost as a measure of the compression efficiency and the transmission cost for the random access ability. We show that the reference placement problem that trades off storage with transmission cost is an integer linear programming problem, that can be solved by standard optimizer. Moreover, we show that the classical periodic placement of the references is only optimal in a very restrictive case: namely, when the encoding costs of each data point are equal and when requests of successive data points are made.

Stars Project-Team

6. New Results

6.1. Introduction

This year Stars has proposed new results related to its three main research axes: (i) perception for activity recognition, (ii) action recognition and (iii) semantic activity recognition.

6.1.1. Perception for Activity Recognition

Participants: François Brémond, Juan Diego Gonzales Zuniga, Abhijit Das, Antitza Dantcheva, Ujjwal Ujjwal, Srijan Das, David Anghelone, Monique Thonnat.

The new results for perception for activity recognition are:

- Handling the Speed-Accuracy Trade-off in Deep Learning based Pedestrian Detection (see 6.2)
- Deep Learning applied on Embedded Systems for People Tracking (see 6.3)
- Partition and Reunion: A Two-Branch Neural Network for Vehicle Re- identification (see 6.4)
- Improving Face Sketch Recognition via Adversarial Sketch-Photo Transformation (see 6.5)
- Impact and Detection of Facial Beautification in Face Recognition: An Overview (see 6.6)
- Computer Vision and Deep Learning applied to Facial analysis in the invisible spectra (see 6.7)

6.1.2. Action Recognition

Participants: François Brémond, Juan Diego Gonzales Zuniga, Abhijit Das, Antitza Dantcheva, Ujjwal Ujjwal, Srijan Das, Monique Thonnat.

The new results for action recognition are:

- ImaGINator: Conditional Spatio-Temporal GAN for Video Generation (see 6.8)
- Characterizing the State of Apathy with Facial Expression and Motion Analysis (see 6.9)
- Dual-threshold Based Local Patch Construction Method for Manifold Approximation And Its Application to Facial Expression Analysis (see 6.10)
- A Weakly Supervised Learning Technique for Classifying Facial Expressions (see 6.11)
- Robust Remote Heart Rate Estimation from Face Utilizing Spatial- temporal Attention (see 6.12)
- Quantified Analysis for Epileptic Seizure Videos (see 6.13)
- Toyota Smarthome: Real-World Activities of Daily Living (see 6.15)
- Looking deeper into Time for Activities of Daily Living Recognition (see 6.15.1)
- Self-Attention Temporal Convolutional Network for Long-Term Daily Living Activity Detection (see 6.16)

6.1.3. Semantic Activity Recognition

Participants: François Brémond, Elisabetta de Maria, Antitza Dantcheva, Srijan Das, Abhijit Das, Daniel Gaffé, Thibaud L'Yvonnet, Sabine Moisan, Jean-Paul Rigault, Annie Ressouche, Ines Sarray, Yaohui Wang, S L Happy, Alexandra König, Philippe Robert, Monique Thonnat.

For this research axis, the contributions are:

- DeepSpa Project (see 6.17)
- Store Connect and Solitaria (see 6.18)
- Synchronous Approach to Activity Recognition (see 6.19)
- Probabilistic Activity Modeling (see 6.20)

6.2. Handling the Speed-Accuracy Trade-off in Deep Learning based Pedestrian Detection

Participants: François Brémond, Ujjwal Ujjwal.

Pedestrian detection is a specific instance of the more general problem of object detection. Pedestrian detection plays a fundamental role in many modern applications involving but not limited to *autonomous vehicles* and *surveillance systems*. These applications as many others are safety-critical. This implies that the cost of not correctly detecting a pedestrian is very high. At the same time applications such as the ones mentioned before, are expected to be real-time. This implies that a pedestrian be detected with minimum time delay. The subject of our recent work has been to design a pedestrian detector which is capable of detecting pedestrians with a high accuracy and high speed – two traits which are known to be difficult to achieve simultaneously.

Most of the pedestrian detectors in computer vision are derived from general-category object detectors. We reflect upon its implication in terms of speed and accuracy below.

6.2.1. Speed-Accuracy Trade-off

Speed and accuracy of object detectors are mutually trade-off factors. Emphasis on higher accuracy usually entails intensive computations which sacrifice the detection speed. On the other hand, emphasis on higher detection speed usually leads to simpler computations which sacrifice the detection accuracy.

We have recently been able to balance this trade-off by identifying that the means of computations on anchors are a major source of the speed-accuracy trade-off. Anchors are hypothetical bounding boxes and are reminiscent of sliding windows used in earlier works on object detection. There are two distinct means of processing anchors – *feature pooling* and *feature probing*. We have recently demonstrated that feature pooling is a costlier strategy than feature probing in terms of computational cost. However, in contrast, feature pooling is a more precise means to process anchors.

We leverage this difference in our approach by utilizing feature pooling throughout in our system. However, in order to gain in terms of run-time performance, we reduce the number of anchors to be processed. This reduction does allow us to process a small number of relevant anchors with high precision.

The block diagram of our proposed approach is shown in figure 4 .

We fuse the feature maps of multiple layers in order to improve the feature diversity. An increased feature diversity assists in learning from a range of hierarchical features generated by a convolutional neural network, often abbreviated as CNN. A depth-wise separable convolutional layer then further processes the fused feature map in order to reduce the number of feature dimensions. One of the prime novelties in our work is the use of pseudo-semantic segmentation. Pseudo-semantic segmentation allows one to obtain a rough estimate of the localization of pedestrians in the form of a heatmap. This step is important, as it provides us with a basis to select a small set of anchors instead of processing all the tiling anchors on the feature map. An anchor classification layer uses anchor-specific kernel sizes to classify a given anchor as positive or negative. A positive or negative anchor is characterized by the overlap between the anchor and the ground truth bounding box during training. This overlap is measured in terms of the well known intersection-over-union (IoU) metric in computer vision. The positive anchors are then pooled from, followed by classification and regression to obtain the final detection.

6.2.2. Results

Figure 5 summarizes the performance of the proposed approach vis-à-vis other approaches. The proposed approach provides significant improvements over other approaches in terms of both speed and accuracy. From figure 5 it is clear that we benefit from initial training on the citypersons data set. Moreover, we obtain the state-of-art performance on the citypersons data set, improving the existing best performing techniques by nearly 4 LAMR points.

6.3. Deep Learning applied on Embedded Systems for People Tracking

Participants: Juan Diego Gonzales Zuniga, Ujjwal Ujjwal, François Brémond, Serge Tissot [Kontron].

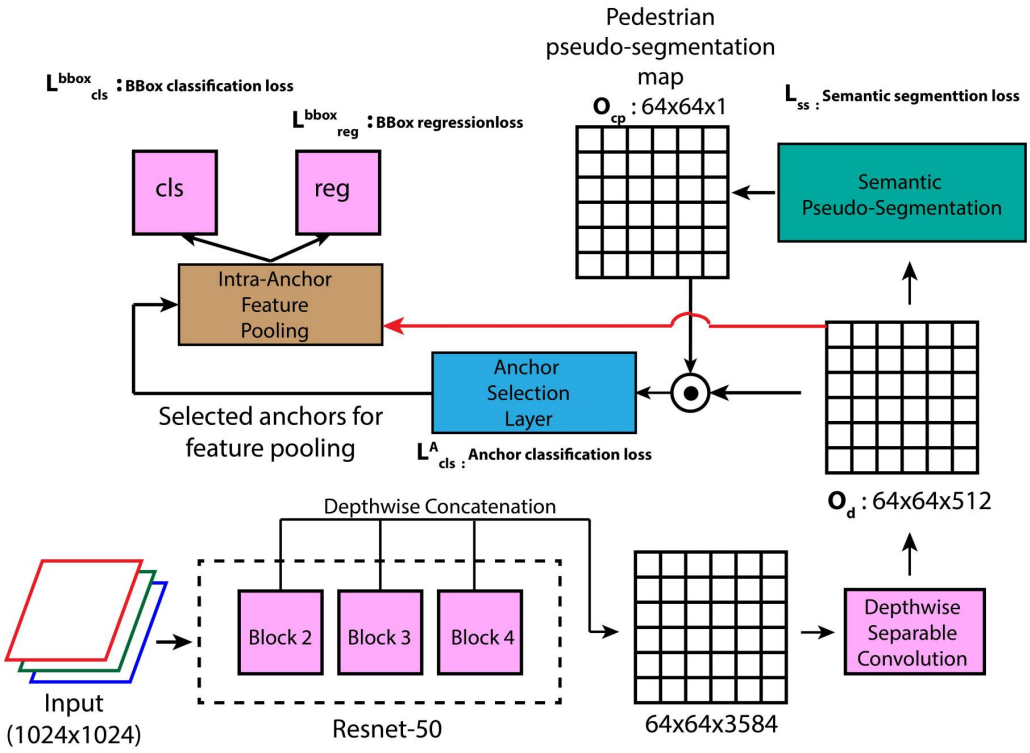


Figure 4. The block diagram of our proposed approach

Method	Stages	LAMR		Speed
		caltech-reasonable (test) (w/o CP pre-training) (CP pre-trained)	citypersons (val) (trained only on CP)	
Faster-RCNN	2	12.10	15.4	7
SSD	1	17.78 (16.36)	19.69	48
YOLOv2	1	21.62 (20.83)	NA	60
RPN-BF	2	9.6 (NA)	NA	7
MS-CNN	2	10.0 (NA)	NA	8
SDS-RCNN	2	7.6 (NA)	NA	5
ALF-Net	1	4.5 (NA)	12.0	20
Rep-Loss	2	5.0 (4.0)	13.2	-
Ours	1.5	**4.76 (3.99)**	**8.12**	**32**

Figure 5. Performance comparison of the proposed method with other methods for caltech-reasonable test set and citypersons validation set. The speed figures are in frames per second.

Our work objective is two-fold: a) Perform tracking of multiple people in videos, which is an instance of Multiple Object Tracking (MOT) problem, and b) optimize this tracking on embedded and open source hardware platforms such as OpenVINO and ROCm.

People tracking is a challenging and relevant problem since it needs multiple additional modules to perform the data association between nodes. In addition, state-of-the-art solutions require intensive memory allocation and power consumption which are not available on embedded hardware. Most architectures either require great amounts of memory or large computing time to achieve a state-of-the-art performance, these results are mostly achieved with dedicated hardware at data centers.

6.3.1. Online Joint Detection and Tracking

In people tracking, we are questioning the main paradigm that is tracking-by-detection which heavily relies on the performance of the underlying detection method. This requires access to a highly accurate and robust people detector. On the other hand, few frameworks attempt detect and track people jointly. Our intent is to perform people tracking *online* and *jointly with detection*.

We are trying to determinate a manner in which a single model can both perform detection and tracking simultaneously. Along these lines, we experimented with a variation of I3D on the Posetrack data set that takes an input of 8 frames in order to create heatmaps along multiple frames as seen in Figure 6 . Giving that the data of Posetrack or MOT cannot train a network as I3D, we are doing the pretraining with the synthetic JTA-Dataset.

This work is inspired by the less common methods of traking-by-tracks and tracking-by-tracklets. Both [40] and [41] generate multi-frame bounding box tuplet proposals and extract detection scores and features with a CNN and LSTM, respectively. Recent researches improve object detection by applying optical flow to propagate scores between frames.

Another method we implemented is by using the detections of previous frames as proposal for the data association, it only uses the IOU between two objects as a distance metric. This approach is simple and efficient assuming the objects do not move drastically. An improved method increases the performance by using a siamese network to conserve identity across frames and predictions for death and birth of tracks.

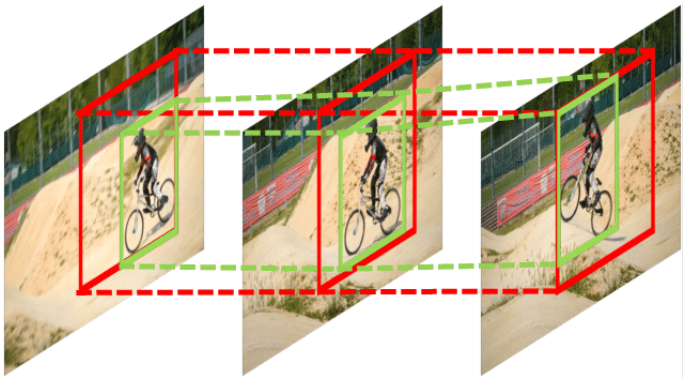


Figure 6. People tracking by tubelets

6.3.2. OpenVINO and ROCm

Regarding embedded hardware, we focus on enlarging both implementation and experimentation of two specific frameworks; OpenVINO and ROCm.

OpenVINO allows us to transfer deep learning models into Myriad and KeemBay chips, taking advantage of their capacity to compute multiple operations without the need of much power consumption. We have thoroughly tested their power consumption under different scenarios as well as implemented many qualitative algorithms with these two platforms, Figure 7 shows the Watt consumption and frame rate of the most popular backbone networks, making it viable to use on embedded applications with a reasonable 25FPS.

For ROCm, we have used the approach of [38] to optimize the compiler execution for a variety of CNN features and filters using a substitute GPU with similar computation capability as Nvidia but still remaining a low branch consumption around 15 Watts.

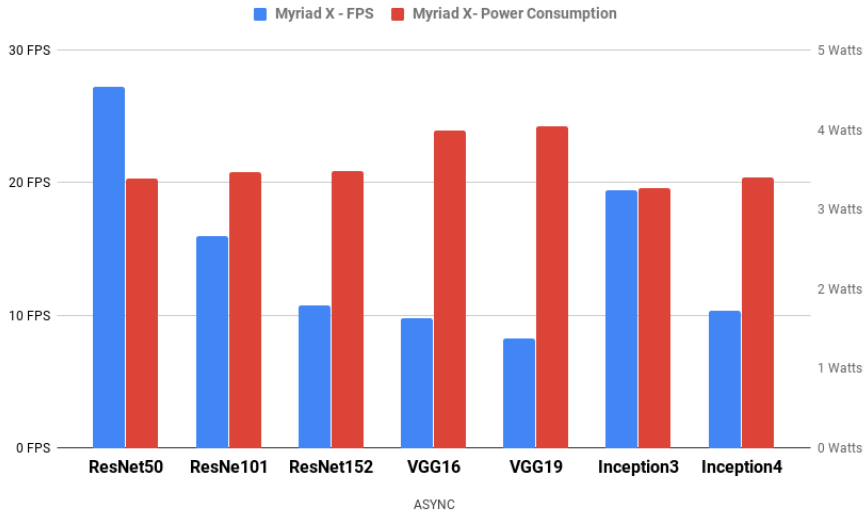


Figure 7. Power Consumption vs Frame rate

6.4. Partition and Reunion: A Two-Branch Neural Network for Vehicle Re-identification

Participants: Hao Chen, Benoit Lagadec, François Brémond.

The smart city vision raises the prospect that cities will become more intelligent in various fields, such as more sustainable environment and a better quality of life for residents. As a key component of smart cities, intelligent transportation system highlights the importance of vehicle re-identification (Re-ID). However, as compared to the rapid progress on person Re-ID, vehicle Re-ID advances at a relatively slow pace. Some previous state-of-the-art approaches strongly rely on extra annotation, like attributes (vehicle color and type) and key-points (wheels and lamps). Recent work on person Re-ID shows that extracting more local features can achieve a better performance without considering extra annotation. In this work, we propose an end-to-end trainable two-branch Partition and Reunion Network (PRN) for the challenging vehicle Re-ID task. Utilizing only identity labels, our proposed method outperforms existing state-of-the-art methods on four vehicle Re-ID benchmark datasets, including VeRi-776, VehicleID, VRIC and CityFlow-ReID by a large margin. The general architecture of our proposed method is represented in the Figure 8 .

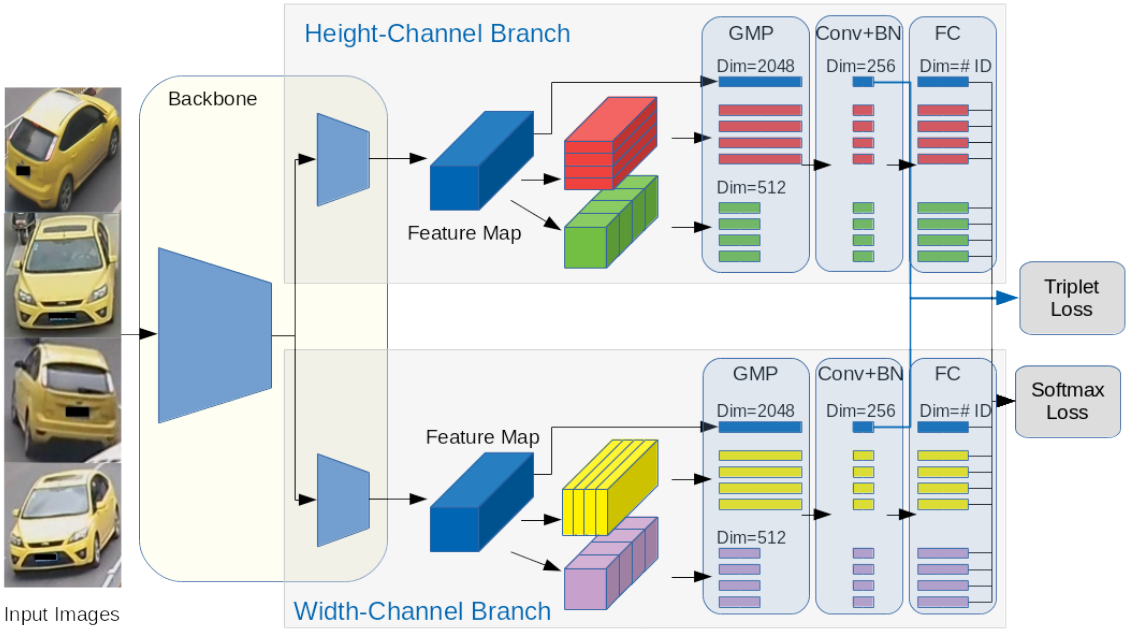


*Figure 8. General architecture of our proposed model. In this work, a ResNet-50 is used as our backbone network. Layers after conv4_1 in Resnet-50 are duplicated to split our network into 2 independent branches. GMP refers to Global Max Pooling. Conv refers to 1*1 convolutional layer, which aims to unify dimensions of global and local feature vectors. FC refers to fully connected layer. BN refers to Batch Normalization layer. In the test phase, all the feature vectors (Dim=256) after Batch Normalization layer are concatenated together as an appearance signature (Dim=256*18).*

6.4.1. Learning Discriminative and Generalizable Representations by Spatial-Channel Partition for Person Re-Identification

In Person Re-Identification (Re-ID) task, combining local and global features is a common strategy to overcome missing key parts and misalignment on models based only on global features. Using this combination, neural networks yield impressive performance in Re-ID task. Previous part-based models mainly focus on spatial partition strategies. Recently, operations on channel information, such as Group Normalization and Channel Attention, have brought significant progress to various visual tasks. However, channel partition has not drawn much attention in Person Re-ID. We conduct a study to exploit the potential of channel partition in Re-ID task [32]. Based on this study, we propose an end-to-end Spatial and Channel partition Representation network (SCR) in order to better exploit both spatial and channel information. Experiments conducted on three mainstream image-based evaluation protocols including Market-1501, DukeMTMC-ReID and CUHK03 and one video-based evaluation protocol MARS validate the performance of our model, which outperforms previous state-of-the-art in both single and cross domain Re-ID tasks. The general architecture of our proposed method is represented in the Figure 9 .

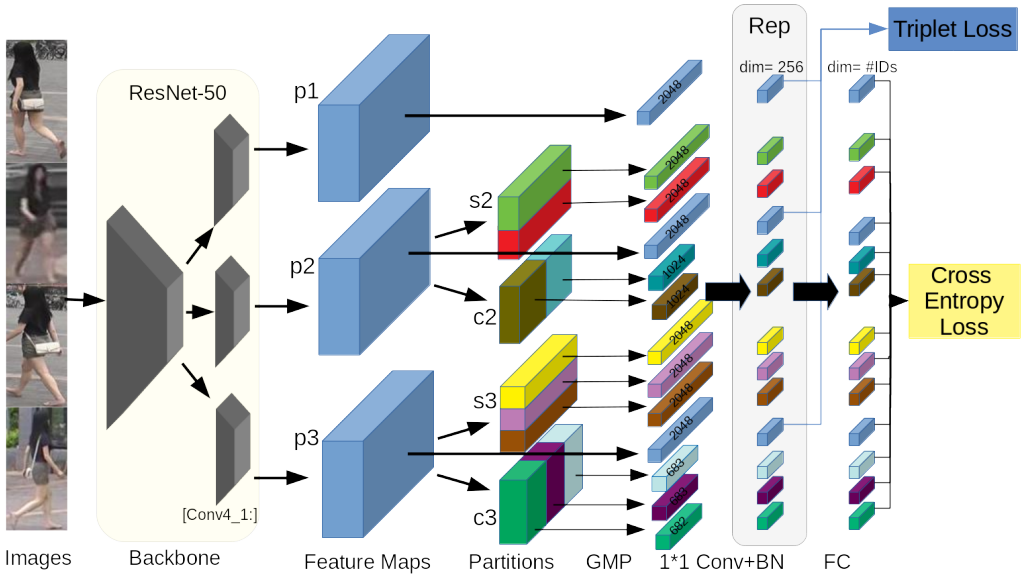


*Figure 9. Spatial and Channel Partition Representation network. For the backbone network, we duplicate layers after conv4_1 into 3 identical but independent branches that generate 3 feature maps "p1", "p2" and "p3". Then, multiple spatial-channel partitions are conducted on the feature maps. "s2" and "c2" refer to 2 spatial parts and 2 channel groups. "s3" and "c3" refer to 3 spatial parts and 3 channel groups. After global max pooling (GMP), dimensions of global (dim = 2048) and local (dim = 2048, 1024*2 and 683*2+682) features are unified by 1*1 convolution (1*1 Conv) and batch normalization (BN) to 256. Then, fully connected layers (FC) give identity predictions of input images. All the dimension unified feature vectors (dim = 256) are aggregated together as appearance representation (Rep) for testing.*

6.5. Improving Face Sketch Recognition via Adversarial Sketch-Photo Transformation

Participants: Antitza Dantcheva, Shikang Yu [Chinese Academy of Sciences], Hu Han [Chinese Academy of Sciences], Shiguang Shan [Chinese Academy of Sciences], Xilin Chen [Chinese Academy of Sciences].

participants

Face sketch-photo transformation has broad applications in forensics, law enforcement, and digital entertainment, particular for face recognition systems that are designed for photo-to-photo matching. While there are a number of methods for face photo-to-sketch transformation, studies on sketch-to-photo transformation remain limited. In this work, we proposed a novel conditional CycleGAN for face sketch-to-photo transformation. Specifically, we leveraged the advantages of CycleGAN and conditional GANs and designed a feature-level loss to assure the high quality of the generated face photos from sketches. The generated face photos were used, as a replacement of face sketches, and particularly for face identification against a gallery set of mugshot photos. Experimental results on the public-domain database CUFSF showed that the proposed approach was able to generate realistic photos from sketches, and the generated photos were instrumental in improving the sketch identification accuracy against a large gallery set. This work has been presented at the IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019) [30].

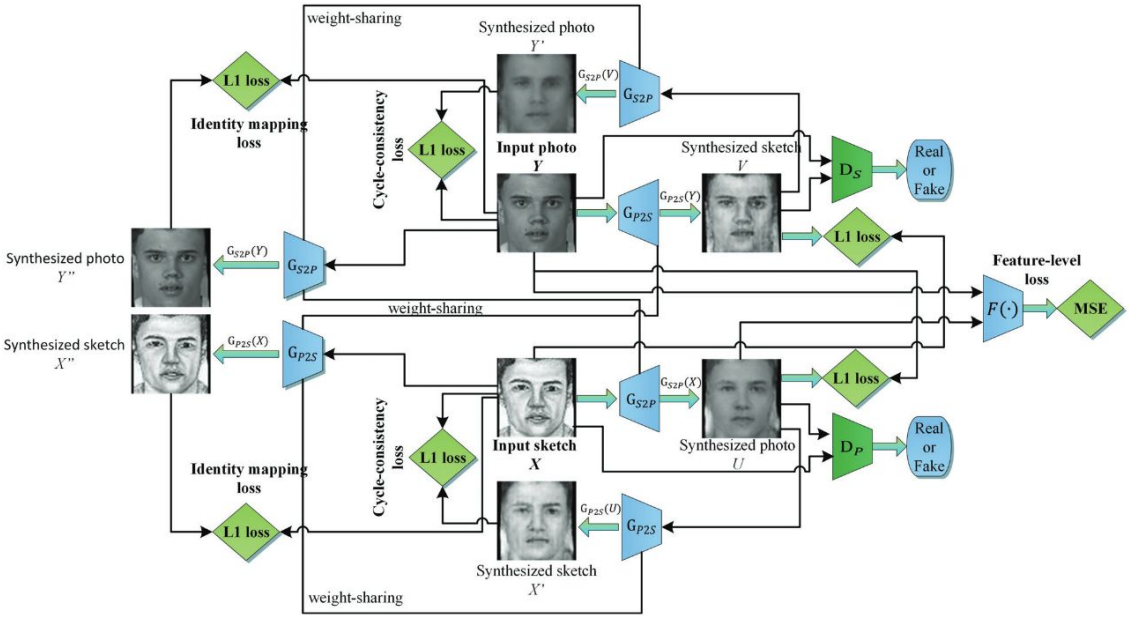


Figure 10. Overview of the proposed GAN for sketch-to-photo transformation using feature-level loss.

6.6. Impact and Detection of Facial Beautification in Face Recognition: An Overview

Participants: Antitza Dantcheva, Christian Rathgeb [Hochschule Darmstadt], Christoph Busch [Hochschule Darmstadt].

Facial beautification induced by plastic surgery, cosmetics or retouching has the ability to substantially alter the appearance of face images. Such types of beautification can negatively affect the accuracy of face recognition systems. In this work, a conceptual categorisation of beautification was presented, relevant scenarios with respect to face recognition were discussed, and related publications were revisited. Additionally, technical considerations and trade-offs of the surveyed methods were summarized along with open issues and challenges in the field. This survey is targeted to provide a comprehensive point of reference for biometric researchers

and practitioners working in the field of face recognition, who aim at tackling challenges caused by facial beautification. This work was published in IEEE Access [18].

6.7. Computer Vision and Deep Learning applied to Facial analysis in the invisible spectra

Participants: David Anghelone, Antitza Dantcheva.

The goal of our work is to analyze faces, as well as recognize events in the invisible spectra. In the last few years, face analysis has been a highly active area and has attracted a lot of interest from the scientific community. Limitations encountered in the visible spectrum such as illumination-restriction have the ability to be overcome in the infrared spectrum. We explored the state-of-the-Art of facial analysis in the invisible spectrum including low energy infrared waves, as well as ultraviolet waves. In this context we have captured images in each spectra and intend to process the data. We aim at designing a model, which extracts biometric features. The key challenges are the processing of contours, shape, etc. This subject is within the framework of the national project *SafeCity*: Security of Smart Cities.

6.8. ImaGINator: Conditional Spatio-Temporal GAN for Video Generation

Participants: Yaohui Wang, Antitza Dantcheva, Piotr Bilinski [University of Warsaw], François Brémond.

keywords: GANs, Video Generation

Generating human videos based on single images entails the challenging simultaneous generation of realistic and visual appealing appearance and motion. In this context, we propose a novel conditional GAN architecture, namely ImaGINator [35] (see Figure 11), which given a single image, a condition (label of a facial expression or action) and noise, decomposes appearance and motion in both latent and high level feature spaces, generating realistic videos. This is achieved by (i) a novel spatio-temporal fusion scheme, which generates dynamic motion, while retaining appearance throughout the full video sequence by transmitting appearance (originating from the single image) through all layers of the network. In addition, we propose (ii) a novel transposed (1+2)D convolution, factorizing the transposed 3D convolutional filters into separate transposed temporal and spatial components, which yields significantly gains in video quality and speed. We extensively evaluate our approach on the facial expression datasets MUG and UvA-NEMO, as well as on the action datasets NATOPS and Weizmann. We show that our approach achieves significantly better quantitative and qualitative results than the state-of-the-art (see Table 1).

Table 1. Evaluation of VGAN, MoCoGAN and proposed ImaGINator with respect to image quality (SSIM/PSNR) and video quality (FID).

	MUG	**NATOPS**
	SSIM/PSNR FID	SSIM/PSNR FID
VGAN	0.28/14.54 74.72	0.72/20.09 167.71
MoCoGAN	0.58/18.16 45.46	0.74/21.82 49.46
ImaGINator	**0.75/22.63 29.02**	**0.88/27.39 26.86**
	Weizmann	**UvA-NEMO**
	SSIM/PSNR FID	SSIM/PSNR FID
VGAN	0.29/15.78 127.31	0.21/13.43 30.01
MoCoGAN	0.42/17.58 116.08	0.45/16.58 29.81
ImaGINator	**0.73/19.67 99.80**	**0.66/20.04 16.16**

6.9. Characterizing the State of Apathy with Facial Expression and Motion Analysis

Participants: S L Happy, Antitza Dantcheva, Abhijit Das, François Brémond, Radia Zeghari [Cobtek], Philippe Robert [Cobtek].

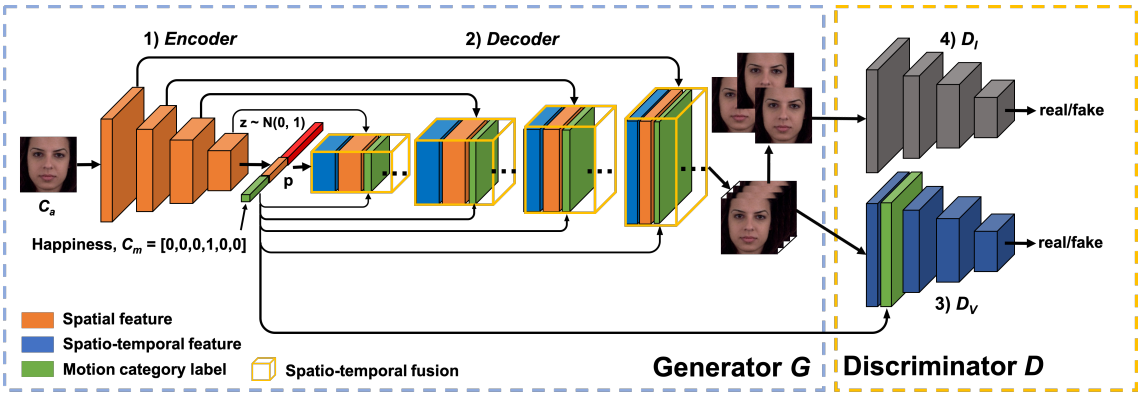


Figure 11. ***Overview of the proposed ImaGINator****. In the GeneratorG, the Encoder firstly encodes an input image c_a into a single vector p. Then, the Decoder produces a video based on a motion c_m and a random vector z. By using spatio-temporal fusion, low level spatial feature maps from the Encoder are directly concatenated into the Decoder. While D_I discriminates whether the generated images contain an authentic appearance, D_V additionally determines whether the generated videos contain an authentic motion.*

Reduced emotional response, lack of motivation, and limited social interaction comprise the major symptoms of apathy. Current methods for apathy diagnosis require the patient's presence in a clinic, and time consuming clinical interviews and questionnaires involving medical personnel, which are costly and logistically inconvenient for patients and clinical staff, hindering among other large scale diagnostics. In this work we introduced a novel machine learning framework to classify apathetic and non-apathetic patients based on analysis of facial dynamics, entailing both emotion and facial movement. Our approach catered to the challenging setting of current apathy assessment interviews, which include short video clips with wide face pose variations, very low-intensity expressions, and insignificant inter-class variations. We tested our algorithm on a dataset consisting of 90 video sequences acquired from 45 subjects and obtained an accuracy of 84% in apathy classification. Based on extensive experiments, we showed that the fusion of emotion and facial local motion produced the best feature set for apathy classification. In addition, we trained regression models to predict the clinical scores related to the mental state examination (MMSE) and the neuropsychiatric apathy inventory (NPI) using the motion and emotion features. Our results suggested that the performance can be further improved by appending the predicted clinical scores to the video-based feature representation. This work has been presented at the IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019) [25].

6.10. Dual-threshold Based Local Patch Construction Method for Manifold Approximation And Its Application to Facial Expression Analysis

Participants: S L Happy, Antitza Dantcheva, Aurobinda Routray [IIT Kharagpur].

In this paper, we propose a manifold based facial expression recognition framework which utilizes the intrinsic structure of the data distribution to accurately classify the expression categories. Specifically, we model the expressive faces as the points on linear subspaces embedded in a Grassmannian manifold, also called as expression manifold. We propose the dual-threshold based local patch (DTLP) extraction method for constructing the local subspaces, which in turn approximates the expression manifold. Further, we use the affinity of the face points from the subspaces for classifying them into different expression classes. Our method is evaluated on four publicly available databases with two well known feature extraction techniques. It is evident from the results that the proposed method efficiently models the expression manifold and improves

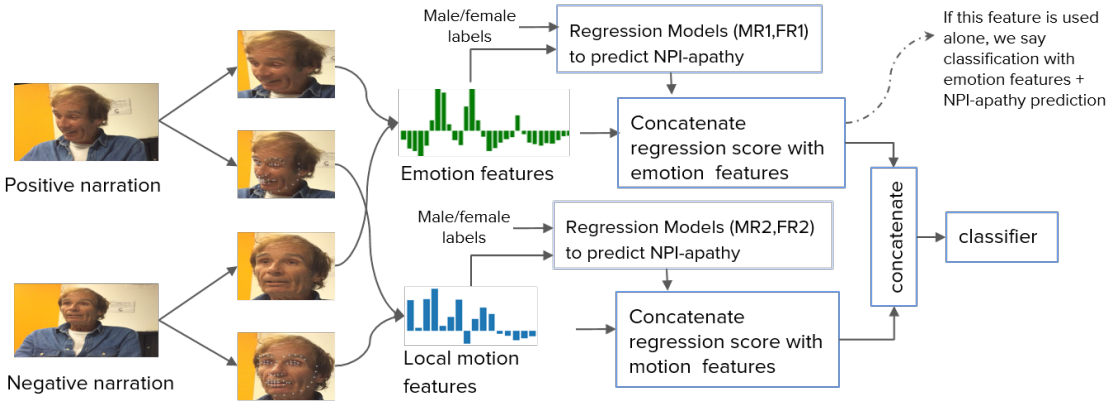


Figure 12. Overall framework for apathy detection from facial videos.

the recognition accuracy in spite of the simplicity of the facial representatives. This work has been presented at the European Signal Processing Conference (EUSIPCO'19) [26].

6.11. A Weakly Supervised Learning Technique for Classifying Facial Expressions

Participants: S L Happy, Antitza Dantcheva, François Brémond.

The universal hypothesis suggests that the six basic emotions: anger, disgust, fear, happiness, sadness, and surprise, are being expressed by similar facial expressions by all humans. While existing datasets support the universal hypothesis and comprise of images and videos with discrete disjoint labels of profound emotions, real-life data contains jointly occurring emotions and expressions of different intensities. Models, which are trained using categorical one-hot vectors often over-fit and fail to recognize low or moderate expression intensities. Motivated by the above, as well as by the lack of sufficient annotated data, we propose a weakly supervised learning technique for expression classification, which leveraged the information of unannotated data. Crucial in our approach was that we first trained a convolutional neural network (CNN) with label smoothing in a supervised manner and proceeded to tune the CNN-weights with both labelled and unlabelled data simultaneously. Experiments on four datasets demonstrated large performance gains in cross-database performance, as well as showed that the proposed method achieved to learn different expression intensities, even when trained with categorical samples. This work was published in Pattern Recognition Letters [15].

6.12. Robust Remote Heart Rate Estimation from Face Utilizing Spatial-temporal Attention

Participants: Antitza Dantcheva, Abhijit Das, Xuesong Niu [Chinese Academy of Sciences], Xingyuan Zhao [Chinese Academy of Sciences], Hu Han [Chinese Academy of Sciences], Shiguang Shan [Chinese Academy of Sciences], Xilin Chen [Chinese Academy of Sciences].

We proposed an end-to-end approach for robust remote heart rate (HR) measurement gleaned from facial videos. Specifically the approach was based on remote photoplethysmography (rPPG), which constitutes a pulse triggered perceivable chromatic variation, sensed in RGB-face videos. Incidentally rPPGs can be affected in less-constrained settings. To unpin the shortcoming, the proposed algorithm utilized a spatio-temporal attention mechanism, which placed emphasis on the salient features included in rPPG-signals. In addition,

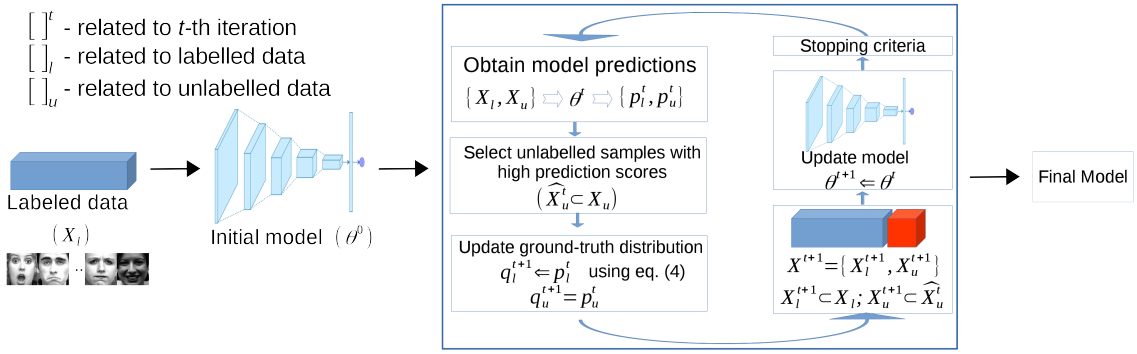


Figure 13. Workflow of the proposed method for weakly supervised learning of facial expressions.

we proposed an effective rPPG augmentation approach, generating multiple rPPG signals with varying HRs from a single face video. Experimental results on the public datasets VIPL-HR and MMSE-HR showed that the proposed method outperformed state-of-the-art algorithms in remote HR estimation. This work has been presented at the IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019) [28].

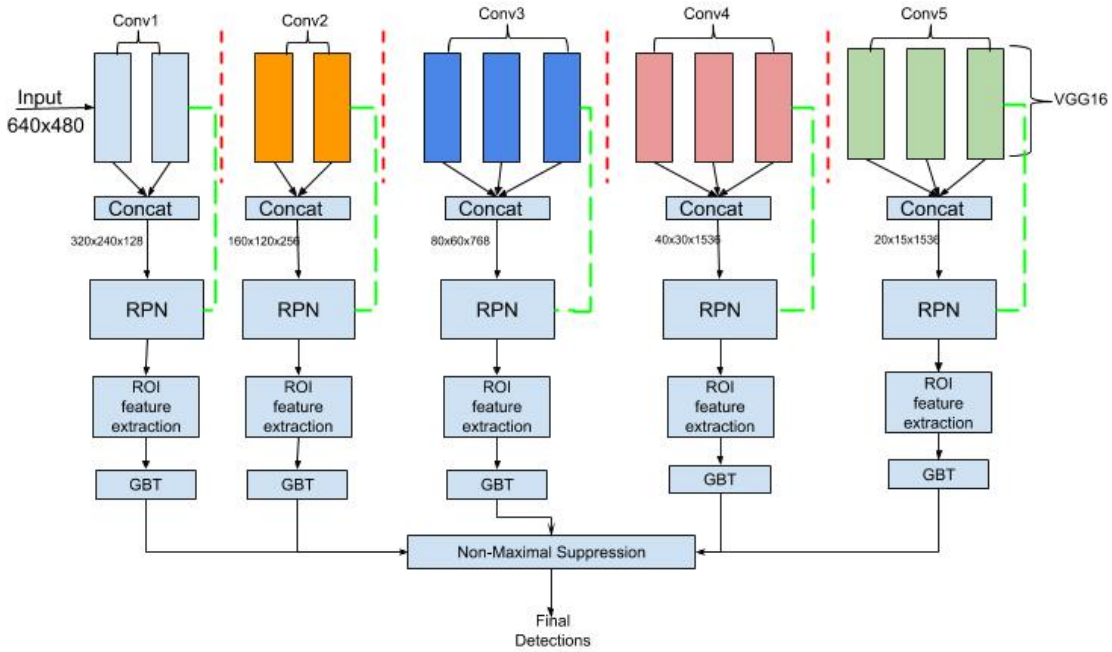


Figure 14. Overview of the proposed end-to-end trainable approach for rPPG based remote HR measurement via representation learning with spatial-temporal attention.

6.13. Quantified Analysis for Epileptic Seizure Videos

Participants: Jen-Cheng Hou, Monique Thonnat.

Epilepsy is a type of neurological disorder, affecting around 50 million people worldwide. Epilepsy's main symptoms are seizures, which are caused by abnormal neuronal activities in the brain. To determine appropriate treatments, neurologists assess manifestation of patients' behavior when seizures occur. Nevertheless, there are few objective criteria regarding the procedure, and diagnosis could be biased due to subjective evaluation. Hence it is important to quantify patients' ictal behaviors for better assessment of the disorder. In collaboration with Dr. Fabrice Bartolomei and Dr. Aileen McGonigal from Timone Hospital, Marseille, we have access to video recordings from epilepsy monitoring unit for analysis, with consent from ethics committee (IRB) and the patients involved.

6.13.1. Seizure Video Classification and Background Video Collection

In an epilepsy monitoring unit, EEG and video recording are usually collected. For patients who need brain surgery to remove lobes that produce seizures, stereo-EEG (SEEG) recordings are particularly measured. SEEG is an intrusive measurement and provides information of the seizure type. We have 86 seizure videos from 20 patients along with the corresponding SEEG conclusion (i.e. pre-frontal epilepsy, occipital epilepsy, etc.). In this study, the goal is to classify seizure videos to their seizure types. Classification was conducted by fine-tuning a pre-trained video classification model, I3D, with 10-fold cross-validation. Due to the relatively small volume of data we have and the challenging nature of our videos, the performance was not satisfactory enough. Inspired by recent semi-supervised works in leveraging large unlabeled dataset for better adaptation to certain tasks, we are collecting large volume of background videos in the epilepsy monitoring unit, in which patients' behavior are normal, such as eating, sleeping, and talking. The volume of the background video can be up to 1000 hours, which could be taken as unlabeled dataset for semi-supervised learning in our case.

6.13.2. Quantifying Rhythmic Rocking Movement with Head Tracking

In this study, six seizures from three patients with pre-frontal epilepsy were analyzed. The duration of rocking was 15-40 seconds, with marked regularity throughout each seizure. Our objective is to document time-evolving frequencies of antero-posterior rocking body movements occurring during seizures. We adopted MobileNet [39] as our backbone model for detecting head of the patient, and hence obtain the trajectories of head movement (see Figure 15). After smoothing the trajectories and find the valid peaks corresponding to the antero-posterior movement, we compute the time-evolving movement frequency for each seizure video. Whereas the rocking frequency varied substantially between patients and seizures (0.3-1Hz), coefficient of variation of frequency was low ($\leq 12\%$). The study report is under review for a medical journal.

6.14. Skeleton Image Representation for 3D Action Recognition

Participants: Carlos Caetano, François Brémond.

Due to the availability of large-scale skeleton datasets, 3D human action recognition has recently called the attention of computer vision community. Many works have focused on encoding skeleton data as skeleton image representations based on spatial structure of the skeleton joints, in which the temporal dynamics of the sequence is encoded as variations in columns and the spatial structure of each frame is represented as rows of a matrix. To further improve such representations, we introduce a novel skeleton image representation to be used as input of Convolutional Neural Networks (CNNs), named SkeleMotion. The proposed approach encodes the temporal dynamics by explicitly computing the magnitude and orientation values of the skeleton joints. Different temporal scales are employed to compute motion values to aggregate more temporal dynamics to the representation making it able to capture long-range joint interactions involved in actions as well as filtering noisy motion values. Experimental results demonstrate the effectiveness of the proposed representation on 3D action recognition outperforming the state-of-the-art on NTU RGB+D 120 dataset. This work has been published in AVSS 2019 [31].

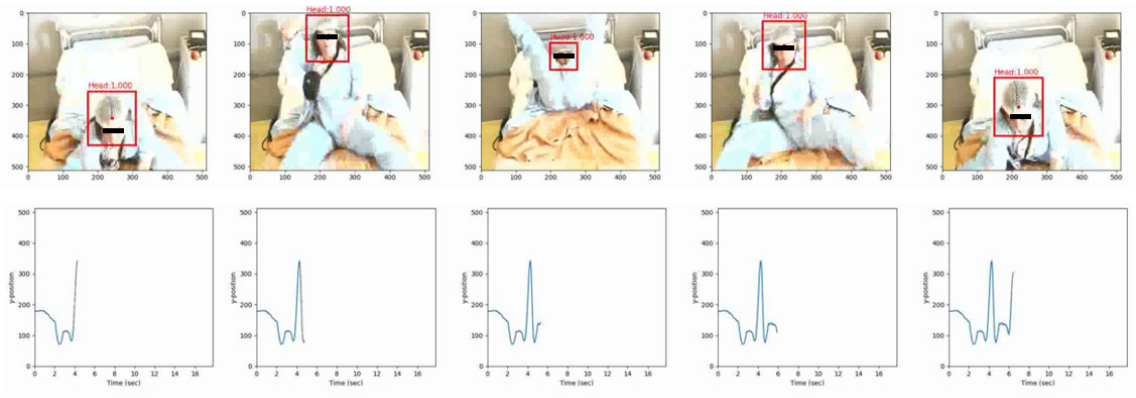


Figure 15. The first row demonstrates the image samples of the antero-posterior movement. The second row shows the position of the head through time in the vertical direction.

In another work, we have explore how to better represent motion information in a video. The temporal component of videos provides an important clue for activity recognition, as a number of activities can be reliably recognized based on the motion information. In view of that, this work proposes a novel temporal stream for two-stream convolutional networks based on images computed from the optical flow magnitude and orientation, named Magnitude-Orientation Stream (MOS), to learn the motion in a better and richer manner. Our method applies simple non-linear transformations on the vertical and horizontal components of the optical flow to generate input images for the temporal stream. Moreover, we also employ depth information to use as a weighting scheme on the magnitude information to compensate the distance of the subjects performing the activity to the camera. Experimental results, carried on two well-known datasets (UCF101 and NTU), demonstrate that using our proposed temporal stream as input to existing neural network architectures can improve their performance for activity recognition. Results demonstrate that our temporal stream provides complementary information able to improve the classical two-stream methods, indicating the suitability of our approach to be used as a temporal video representation. two-stream convolutional networks, spatiotemporal information, optical flow, depth information. This work has been published in the Journal of Visual Communication and Image Representation [14].

6.15. Toyota Smarthome: Real-World Activities of Daily Living

Participants: Srijan Das, Rui Dai, François Brémond.

The performance of deep neural networks is strongly influenced by the quantity and quality of annotated data. Most of the large activity recognition datasets consist of data sourced from the Web, which does not reflect challenges that exist in activities of daily living. In this work, we introduce a large real-world video dataset for activities of daily living: Toyota Smarthome. The dataset consists of 16K RGB+D clips of 31 activity classes, performed by seniors in a smarthome. Unlike previous datasets, videos were fully unscripted. As a result, the dataset poses several challenges: high intra-class variation, high class imbalance, simple and composite activities, and activities with similar motion and variable duration. Activities were annotated with both coarse and fine-grained labels. These characteristics differentiate Toyota Smarthome from other datasets for activity recognition as illustrated in 16 .

As recent activity recognition approaches fail to address the challenges posed by Toyota Smarthome, we present a novel activity recognition method with attention mechanism. We propose a pose driven spatio-temporal attention mechanism through 3D ConvNets. We show that our novel method outperforms state-of-the-art methods on benchmark datasets, as well as on the Toyota Smarthome dataset. We release the dataset

for research use at https://project.inria.fr/toyotasmarthome. This work is done in collaboration with Toyota Motors Europe and is published in ICCV 2019 [21].

Figure 16. Sample frames from Toyota Smarthome dataset: 1-7 label at the right top corner respectively correspond to camera view 1, 2, 3, 4, 5, 6 and 7 as marked in the plan of the apartment on the right. Image from camera view (1) Drink from can, (2) Drink from bottle, (3) Drink form glass and (4) Drink from cup are all fine grained activities with a coarse label drink. Image from camera view (5) Watch TV and (6) Insert tea bag show activities with large source-to-camera distance and occlusion. Images with camera view (7) Enter illustrate the RGB image and the provided 3D skeleton.

6.15.1. Looking deeper into Time for Activities of Daily Living Recognition

Participants: Srijan Das, Monique Thonnat, François Brémond.

In this work, we introduce a new approach for Activities of Daily Living (ADL) recognition. In order to discriminate between activities with similar appearance and motion, we focus on their temporal structure. Actions with subtle and similar motion are hard to disambiguate since long-range temporal information is hard to encode. So, we propose an end-to-end Temporal Model to incorporate long-range temporal information without losing subtle details. The temporal structure is represented globally by different temporal granularities and locally by temporal segments as illustrated in fig. 17 . We also propose a two-level pose driven attention mechanism to take into account the relative importance of the segments and granularities. We validate our approach on 2 public datasets: a 3D human activity dataset (NTU-RGB+D) and a human action recognition dataset with object interaction dataset (Northwestern-UCLA Multiview Action 3D). Our Temporal Model can also be incorporated with any existing 3D CNN (including attention based) as a backbone which reveals its robustness. This work has been accepted in WACV 2020 [20].

6.16. Self-Attention Temporal Convolutional Network for Long-Term Daily Living Activity Detection

Participants: Rui Dai, François Brémond.

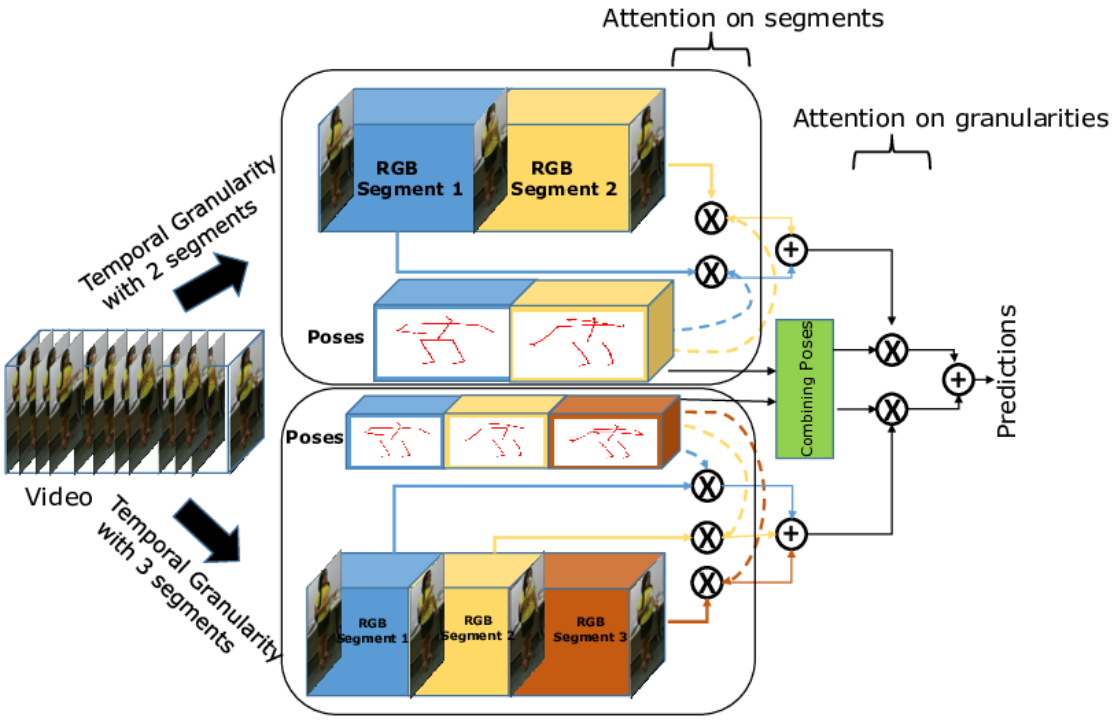


Figure 17. Framework of the proposed approach in a nutshell for two temporal granularities. The articulated poses soft-weight the temporal segments and the temporal granularities using a two-level attention mechanism.

This year, we proposed a Self-Attention - Temporal Convolutional Network (SA-TCN), which is able to capture both complex activity patterns and their dependencies within long-term untrimmed videos [34]. This attention block can also embed with other TCN-nased models. We evaluate our proposed model on DAily Home LIfe Activity Dataset (DAHLIA) and Breakfast datasets. Our proposed method achieves state-of-the-art performance on both datasets.

6.16.1. Work Flow

Given an untrimmed video, we represent each non-overlapping snippet by a visual encoding over 64 frames. This visual encoding is the input to the encoder-TCN, which is the combination of the following operations: 1D temporal convolution, batch normalization, ReLu, and max pooling. Next, we send the output of the encoder-TCN into the self-attention block to capture long-range dependencies. After that, the decoder-TCN applies the 1D convolution and up sampling to recover a feature map of the same dimension as visual encoding. Finally, the output will be sent to a fully connected layer with softmax activation to get the prediction. Fig 18 and 19 provide the structure of our model.

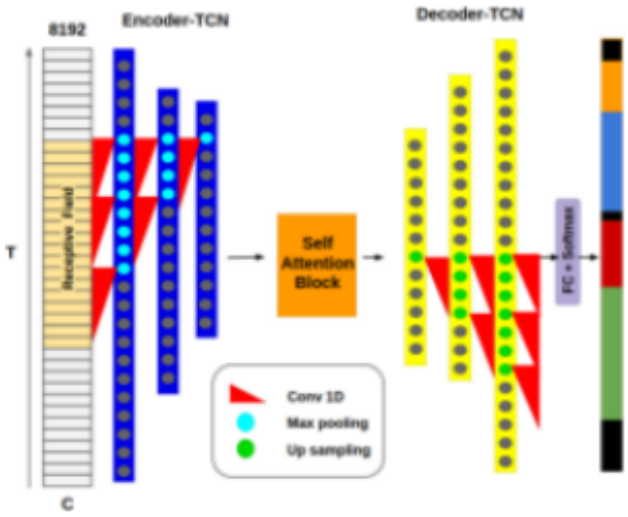


*Figure 18. **Overview**. The model contains mainly three parts: (1) visual encoding, (2) encoder-decoder structure, (3) attention block*

6.16.2. Result

We evaluated the proposed method on two daily-living activity datasets (DAHLIA, Breakfast) and achieved state-of-the-art performances. We compared with these following State-of the arts: DOHT, Negin *et al.*, GRU , ED-TCN, TCFPN.

6.17. DeepSpa Project

Participants: Alexandra König, Rachid Guerchouche, Minh Tran-Duc, Antitza Dantcheva, S L Happy, Abhijit Das.

The DeepSpa (Deep Speech Analysis, January 2019 - June 2020) project aims to deliver telecommunication-based neurocognitive assessment tools for early screening, early diagnostic and follow-up of cognitive disorders, mainly in elderly. The target is also clinical trials addressing Alzheimer's and other neurodegenerative diseases. By combining AI in speech recognition and video analysis for facial expression recognition, the proposed tools allow remote cognitive and psychological testing, thereby saving time and money.

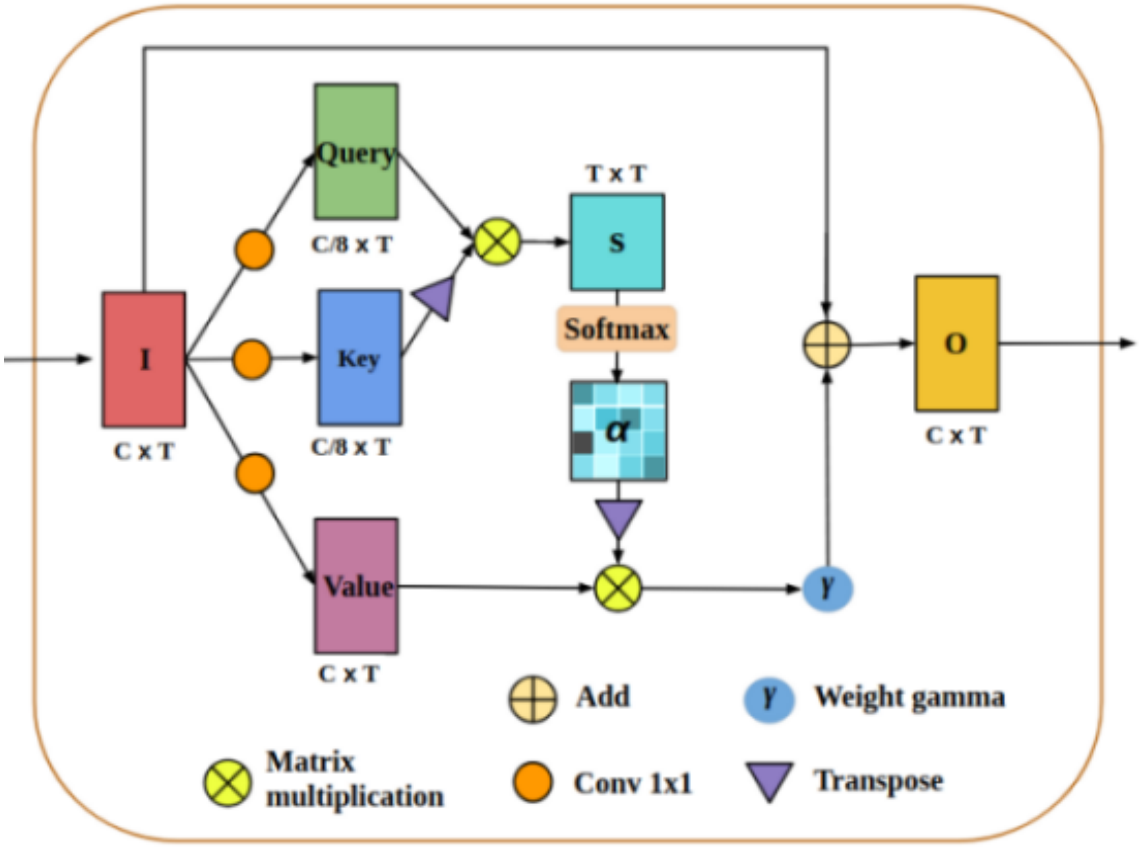


*Figure 19. **Attention block**. This figure presents the structure of attention block*

Table 2. Activity detection results on DAHLIA dataset with the average of view 1, 2 and 3. * marked methods have not been tested on DAHLIA in their original paper.

Model	FA1	F-score	IoU	mAP
DOHT	0.803	0.777	0.650	-
GRU*	0.759	0.484	0.428	0.654
ED-TCN*	0.851	0.695	0.625	0.826
Negin *et al.*	0.847	0.797	0.723	-
TCFPN*	0.910	**0.799**	0.738	**0.879**
SA-TCN	**0.921**	0.788	**0.740**	0.862

Table 3. Activity detection results on Breakfast dataset.

Model	FA1	F-Score	IoU	mAP
GRU	0.368	0.295	0.198	0.380
ED-TCN	0.461	0.462	0.348	0.478
TCFPN	**0.519**	0.453	0.362	0.466
SA-TCN	0.497	**0.494**	**0.385**	**0.480**

Table 4. Average precision of ED-TCN on DAHLIA.

Activities	**Background**	**House work**	**Working**	**Cooking**
AP	0.36	0.65	0.95	0.96
Activities	**Laying table**	**Eating**	**Clearing table**	**Wash dishes**
AP	0.90	0.97	0.80	0.97

Table 5. Combination of attention block with other TCN-based model: TCFPN. (Evaluated on DAHLIA dataset)

Model	**FA1**	**F-score**	**IoU**	**mAP**
TCFPN	0.910	**0.799**	0.738	0.879
SA-TCFPN	**0.917**	**0.799**	**0.748**	**0.894**

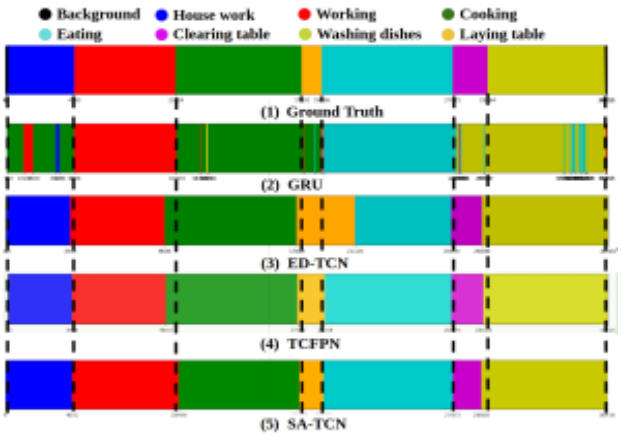


*Figure 20. **Detection visualization**. The detection visualization of video 'S01A2K1' in DAHLIA: (1) ground truth, (2) GRU, (3) ED-TCN, (4) TCFPN and (5) SA-TCN.*

The partners of the project are:

- Inria: technical partner and project coordinator
- University of Maastricht: clinical partner
- Jansen & Jansen: pharma partner and business champion
- Association Innovation Alzheimer: subgranted clinical partner
- Ki-element: subgranted technical partner.

6.17.1. Project structure

The DeepSpA project is structured in two use-cases:

- Use-case 1: remote assessment through phone for early screening of cognitive disorders (University of Maastricht, Jansen & Jansen and Ki-element): using AI based speech recognition; assessments through phone are made possible. A clinical trial is currently running in Maastricht (by end 2019, 70 subjects will be included, and 50 others will be included in 2020), the goal is to study the feasibility of such phone assessment in comparison to face-to-face assessment.
- Use-case 2: remote assessment through video-conference system (telemedicine tool) (Inria, Jansen & Jansen and Association Innovation Alzheimer): Inria developed a telemedicine tool which allows complete remote assessment. AI based speech and facial expression recognition empower the cognitive assessment by providing extra features useful for clinicians.

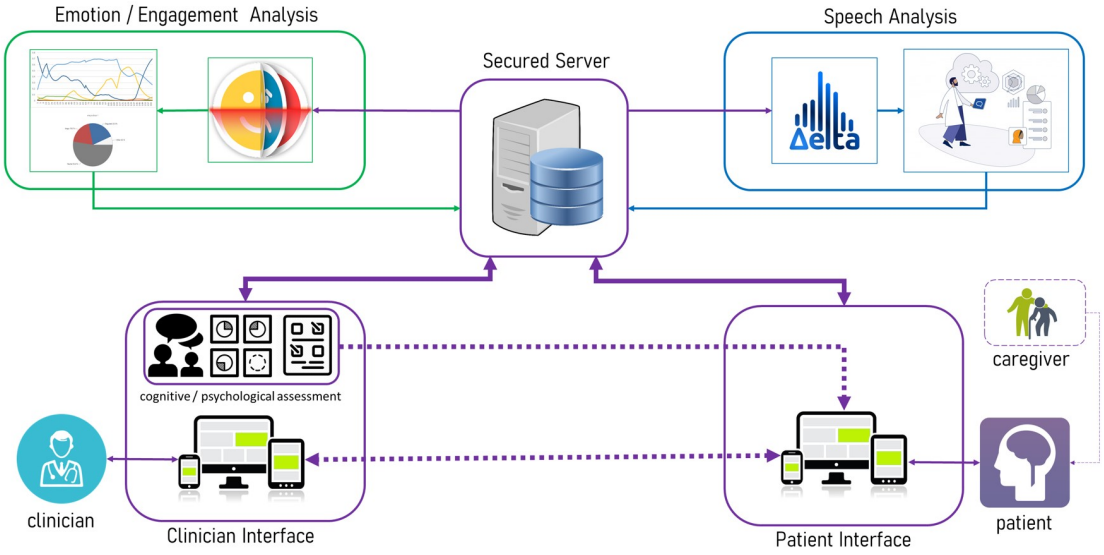


Figure 21. Global view of the telemedicine tool developed by Inria (STARS).

6.17.2. Telemedicine / Clinical Study with Digne-les-Bains

In order to evaluate the feasibility of remote assessment through the telemedicine tool, a collaboration with the city of Digne-les-Bains started in March 2019. The Hospital of Digne-les-Bains, la Maison de la Santé and the ADMR (association dealing with isolated people) are involved in a running clinical study, which aims at evaluating the feasibility of the remote assessment in two different setups:

- Clinical setup: a fixed place where the participant will undergo the telemedicine session: clinic, hospital, pharmacy, health centres ....
- Mobile Units: a mobile unit goes to the subjects home, the telemedicine session is done inside the mobile unit (e.g., van).

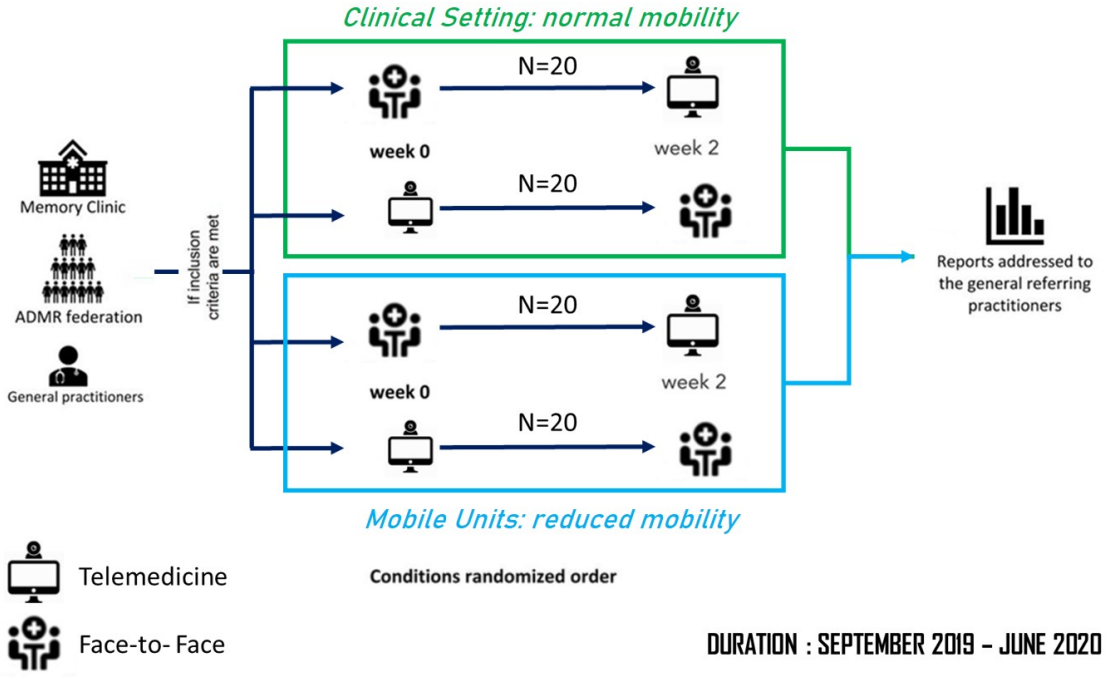


Figure 22. Global view of the clinical study with Digne-les-Bains.

We already started including subjects in the clinical setup case. By end 2019, we expect to include about 15 subjects, 25 extra subjects will be included during 2020. Mobile units setup will be tested during 2020.

First results and observations already showed that the telemedicine tool allows full assessments. Clinicians and patients showed strong interest and appreciation of such tool.

6.17.3. Facial expressions recognition and engagement evaluation in the telemedicine tool

The STARS team is doing research on facial expressions analysis, which could be integrated as part of the vision module of the telemedicine tool [25].

Notable software related to this research is the provided API on the cloud, which allows sending video files and retrieving emotions, gaze direction, facial movements and head direction (implemented by S L Happy).

6.18. Store Connect and Solitaria

Participants: Sébastien Gilabert, Minh Khue Phan Tran, François Brémond.

Store-Connect was a consortium aiming at detecting and positioning people in a supermarket. Several technologies were explore such as computing and merging trajectories obtained from the mobile phone of customers and from video cameras. In a second step, the goal was to detect all the 'stop' events of the customers while shopping in the store.

6.18.1. SupICP

We have developed with the SED team, SupICP, a platform for integrating all plugins developed by STARS team. Our main contribution is the Ontology Language Plugin. With this plugin, we can use contextual and knowledge information inside scenarios designed for video event recognition. Currently, we are improving this plugin for combining the Ontology Language with Deep Learning technology towards "Action recognition based on Deep Learning and Ontology Language".

We have also installed this software at the Institute Claude Pompidou, in order to conduct clinical trials, and to work with medical scientists.

6.18.2. Solitaria

The aim of this project is to combine data extracted from domestic sensors and from video cameras, and to implement this plugin into SupICP to monitor older people at home.

6.19. Synchronous Approach to Activity Recognition

Participants: Daniel Gaffé, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Ines Sarray.

Activity Recognition aims at recognizing and understanding sequences of actions and movements of mobile objects (human beings, animals or artifacts), that follow the predefined model of an activity. We propose to describe activities as a series of actions, triggered and driven by environmental events.

This year we mainly refined the ADeL description language, the semantics of some of its instructions and their compilation into equation systems. We also improved the recognition engine and the synchronizer to better handle the synchronous/asynchronous transformation.

Work remains to be done to complete a full framework to generate generic recognition systems and automatic tools to interface with static and dynamic analysis tools, such as model checkers or performance monitors.

6.19.1. Activity Description Language

The ADeL language was designed to describe various activities, it provides two different (and equivalent) formats: graphical and textual. This year we started to describe use case examples in the medical domain: serious games and exercises for patients having cognitive problems, such as Alzheimer or autistic persons. This kind of games are used to test patients and to evaluate their behavior and interactions. These use cases lead us to improve the language and part of its semantics. An example of the graphical format describing a simple exercise activity is given in figure 23 .

Work remains to be done to improve the usability of the language by our end-users.

6.19.2. Synchronizer

Using the synchronous paradigm makes time manipulation easy thanks to determinism and synchronous parallellism; moreover, tools exist to support formal verification. However, the sensor environment is asynchronous and it is thus necessary to transform asynchronous events given by sensors into synchronous logical instants. It is a difficult problem that does not have an exact and complete solution. We introduced a component called "synchronizer" between the environment sensors and the recognition engine. The synchronizer is responsible for filtering the sensor data, grouping them into logical instants, and sending these instants to the recognition engine.

We specified a generic algorithm, based on *awaited* events, i.e. the events which may trigger transitions to a next state. These events are provided by each automaton in each state. This algorithm is parametrized by heuristics to adapt to different situations. There are two main points of variation in the synchronizer where heuristics can be applied: when processing data coming from the sensors (to collect and combine raw data) and when building logical instants (to decide on the end of instants and to manage preemptions).

This year we finished to implement a first version of the synchronizer (for one single activity to recognize), we defined different heuristics, and we tested the synchronizer algorithm on some uses cases with these heuristics [11].

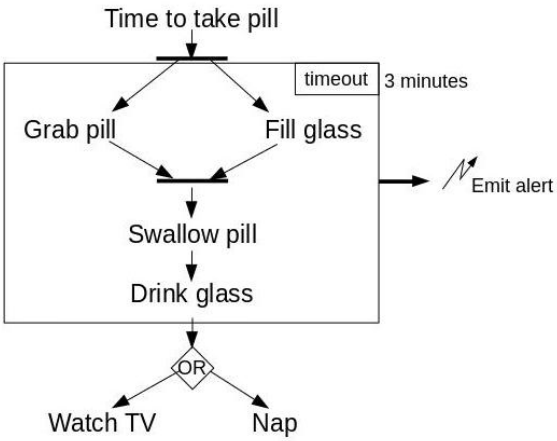


Figure 23. Example of a simple activity (patient should take a pill at a given time) including a parallel and a timeout instructions.

6.20. Probabilistic Activity Modeling

Participants: Elisabetta de Maria, Sabine Moisan, Jean-Paul Rigault, Thibaud L'Yvonnet.

Serious games constitute a domain in which real-time activity recognition is particularly relevant: the expected behavior is well identified and it is possible to rely on different sensors (biometric and external) while playing the game. We focus on games to help in diagnosis and treatment of patients.

We developed a formal approach to model such activities, taking into account possible variations in human behavior. All the scenarios of an activity are not equivalent: some are typical (thus frequent) while others seldom happen. We propose to quantify the likelihood of these variations by associating probabilities with the key actions of the activity description. We rely on a formal model based on probabilistic discrete-time Markov chains (DTMCs). We used the PRISM framework and its model checking facilities to express and check interesting temporal logic properties (PCTL).

As a use case, we considered a serious game to analyze the behavior of Alzheimer patients. We encoded this game as a DTMC in PRISM and we defined several meaningful PCTL properties that are then automatically tested thanks to the PRISM model checker. Two kinds of properties may be defined: those to verify the model and those oriented toward the medical domain. The latter may give indications to a practitioner regarding a patient's behavior. These properties include the use of PRISM "rewards" to quantify the performance of patients.

We expect that such a modeling approach could provide doctors with new indications for interpreting patients' performance and we identified three medically interesting outcomes for this approach. First, to evaluate a new patient before the first diagnosis of doctors, we can compare her game performance to a reference model representing a "healthy" behavior. Second, to monitor known patients, a customized model can be created according to their first results, and, over time, their health improvement or deterioration could be monitored. Finally, to pre-select a cohort of patients, we can use a reference model to determine, in a fast way, whether a new group of patients belongs to this specific category.

This year we first addressed the model definition and its suitability to check behavioral properties of interest [24]. Indeed, this is mandatory before envisioning any clinical study.

The next step will be to validate our approach as well as to test its scalability on three other serious games selected with the help of clinicians. We wrote a medical protocol to be submitted to CERNI proposing clinical experimentations with patients. This protocol will be a collaboration with the ICP institute, member of the CoBTEX laboratory. The new games will be modeled in PRISM and different configurations (for example for Mild, Moderate or Severe Alzheimer) will be set up with the participation of clinicians. Then, several groups of patients will play these games and their results will be recorded to calibrate our initial models.

THOTH Project-Team

7. New Results

7.1. Visual Recognition and Robotics

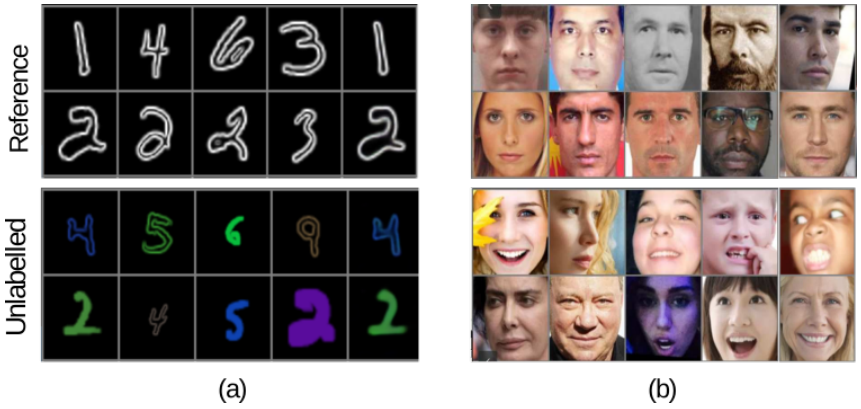


Figure 1. Illustration of different reference-based disentangling problems. (a) Disentangling style from digits. The reference distribution is composed by numbers with a fixed style (b) Disentangling factors of variations related with facial expressions. Reference images correspond to neutral faces. Note that pairing information between unlabelled and reference images is not available during training.

7.1.1. Learning Disentangled Representations with Reference-Based Variational Autoencoders

Participants: Adria Ruiz, Oriol Martinez, Xavier Binefa, Jakob Verbeek.

Learning disentangled representations from visual data, where different high-level generative factors are independently encoded, is of importance for many computer vision tasks. Supervised approaches, however, require a significant annotation effort in order to label the factors of interest in a training set. To alleviate the annotation cost, in [32] we introduce a learning setting which we refer to as "reference-based disentangling". Given a pool of unlabelled images, the goal is to learn a representation where a set of target factors are disentangled from others. The only supervision comes from an auxiliary "reference set" that contains images where the factors of interest are constant. See Fig. 1 for illustrative examples. In order to address this problem, we propose reference-based variational autoencoders, a novel deep generative model designed to exploit the weak supervisory signal provided by the reference set. During training, we use the variational inference framework where adversarial learning is used to minimize the objective function. By addressing tasks such as feature learning, conditional image generation or attribute transfer, we validate the ability of the proposed model to learn disentangled representations from minimal supervision.

7.1.2. Tensor Decomposition and Non-linear Manifold Modeling for 3D Head Pose Estimation

Participants: Dmytro Derkach, Adria Ruiz, Federico M. Sukno.

Head pose estimation is a challenging computer vision problem with important applications in different scenarios such as human-computer interaction or face recognition. In [5], we present a 3D head pose estimation algorithm based on non-linear manifold learning. A key feature of the proposed approach is that it allows modeling the underlying 3D manifold that results from the combination of rotation angles. To do so, we use tensor decomposition to generate separate subspaces for each variation factor and show that each of them has a clear structure that can be modeled with cosine functions from a unique shared parameter per angle (see Fig. 2). Such representation provides a deep understanding of data behavior. We show that the proposed framework can be applied to a wide variety of input features and can be used for different purposes. Firstly, we test our system on a publicly available database, which consists of 2D images and we show that the cosine functions can be used to synthesize rotated versions from an object from which we see only a 2D image at a specific angle. Further, we perform 3D head pose estimation experiments using other two types of features: automatic landmarks and histogram-based 3D descriptors. We evaluate our approach on two publicly available databases, and demonstrate that angle estimations can be performed by optimizing the combination of these cosine functions to achieve state-of-the-art performance.

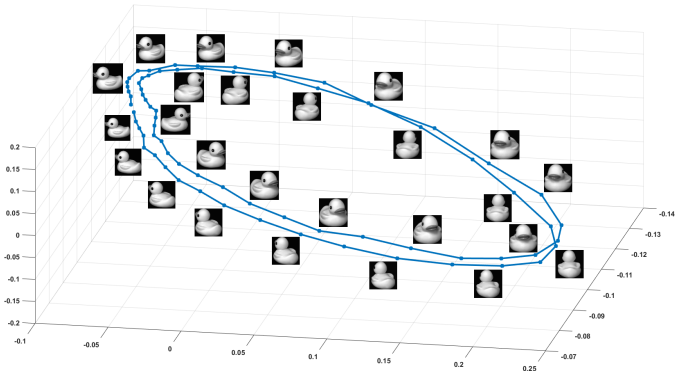


Figure 2. Visualization of the first three coefficients of the pose variation subspace for a dataset of single object rotated about the vertical axis.

7.1.3. Spreading vectors for similarity search

Participants: Alexandre Sablayrolles, Matthijs Douze, Cordelia Schmid, Hervé Jégou.

Discretizing multi-dimensional data distributions is a fundamental step of modern indexing methods. State-of-the-art techniques learn parameters of quantizers on training data for optimal performance, thus adapting quantizers to the data. In this work [29], we propose to reverse this paradigm and adapt the data to the quantizer: we train a neural net which last layer forms a fixed parameter-free quantizer, such as pre-defined points of a hyper-sphere. As a proxy objective, we design and train a neural network that favors uniformity in the spherical latent space, while preserving the neighborhood structure after the mapping. We propose a new regularizer derived from the Kozachenko–Leonenko differential entropy estimator to enforce uniformity and combine it with a locality-aware triplet loss. Experiments show that our end-to-end approach outperforms most learned quantization methods, and is competitive with the state of the art on widely adopted benchmarks. Furthermore, we show that training without the quantization step results in almost no difference in accuracy, but yields a generic catalyzer 3 that can be applied with any subsequent quantizer. The code is available online.

7.1.4. Diversity with Cooperation: Ensemble Methods for Few-Shot Classification

Participants: Nikita Dvornik, Cordelia Schmid, Julien Mairal.

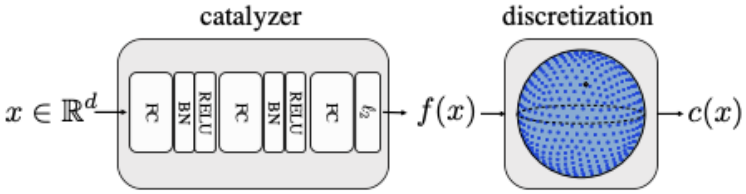


Figure 3. Our method learns a network that encodes the input space $\mathbb{R}^d$ into a code $c(x)$. It is learned end-to-end, yet the part of the network in charge of the discretization operation is fixed in advance, thereby avoiding optimization problems. The learnable function f, namely the "catalyzer", is optimized to increase the quality of the subsequent coding stage.

Few-shot classification consists of learning a predictive model that is able to effectively adapt to a new class, given only a few annotated samples. To solve this challenging problem, meta-learning has become a popular paradigm that advocates the ability to "learn to adapt". Recent works have shown, however, that simple learning strategies without meta-learning could be competitive. In our ICCV'19 paper [17], we go a step further and show that by addressing the fundamental high-variance issue of few-shot learning classifiers, it is possible to significantly outperform current meta-learning techniques. Our approach consists of designing an ensemble of deep networks to leverage the variance of the classifiers, and introducing new strategies to encourage the networks to cooperate, while encouraging prediction diversity, as illustrated in Figure 4 . Evaluation is conducted on the mini-ImageNet and CUB datasets, where we show that even a single network obtained by distillation yields state-of-the-art results.

7.1.5. Unsupervised Pre-Training of Image Features on Non-Curated Data

Participants: Mathilde Caron, Piotr Bojanowski [Facebook AI], Julien Mairal, Armand Joulin [Facebook AI].

Pre-training general-purpose visual features with convolutional neural networks without relying on annotations is a challenging and important task. Most recent efforts in unsupervised feature learning have focused on either small or highly curated datasets like ImageNet, whereas using non-curated raw datasets was found to decrease the feature quality when evaluated on a transfer task. Our goal is to bridge the performance gap between unsupervised methods trained on curated data, which are costly to obtain, and massive raw datasets that are easily available. To that effect, we propose a new unsupervised approach, DeeperCluster [13], described in Figure 5 which leverages self-supervision and clustering to capture complementary statistics from large-scale data. We validate our approach on 96 million images from YFCC100M, achieving state-of-the-art results among unsupervised methods on standard benchmarks, which confirms the potential of unsupervised learning when only non-curated raw data are available. We also show that pre-training a supervised VGG-16 with our method achieves 74.9% top-1 classification accuracy on the validation set of ImageNet, which is an improvement of $+0.8\%$ over the same network trained from scratch.

7.1.6. Learning to Augment Synthetic Images for Sim2Real Policy Transfer

Participants: Alexander Pashevich, Robin Strudel [Inria WILLOW], Igor Kalevatykh [Inria WILLOW], Ivan Laptev [Inria WILLOW], Cordelia Schmid.

Vision and learning have made significant progress that could improve robotics policies for complex tasks and environments. Learning deep neural networks for image understanding, however, requires large amounts of domain-specific visual data. While collecting such data from real robots is possible, such an approach

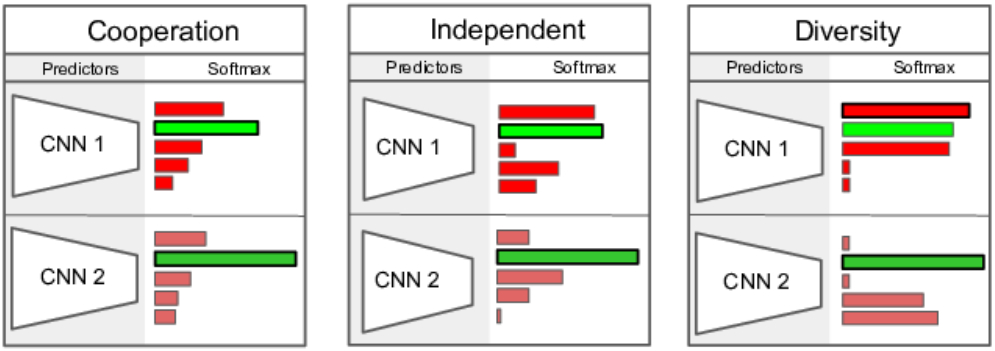


Figure 4. ***Illustration of the cooperation and diversity strategies on two networks.*** *All networks receive the same image as input and compute corresponding class probabilities with softmax. Cooperation encourages the non-ground truth probabilities (in red) to be similar, after normalization, whereas diversity encourages orthogonality.*

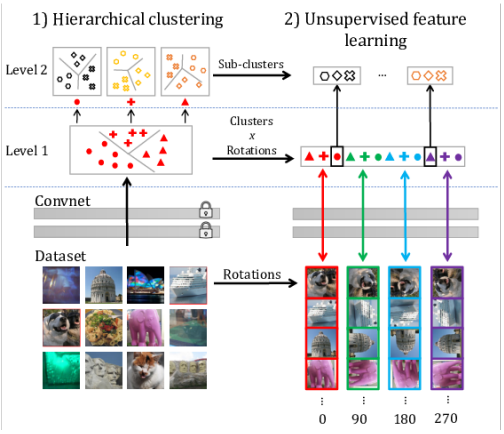


Figure 5. DeeperCluster alternates between a hierachical clustering of the features and learning the parameters of a convnet by predicting both the rotation angle and the cluster assignments in a single hierachical loss.

limits the scalability as learning policies typically requires thousands of trials. In this work [25] we attempt to learn manipulation policies in simulated environments. Simulators enable scalability and provide access to the underlying world state during training. Policies learned in simulators, however, do not transfer well to real scenes given the domain gap between real and synthetic data. We follow recent work on domain randomization and augment synthetic images with sequences of random transformations. Our main contribution is to optimize the augmentation strategy for sim2real transfer and to enable domain-independent policy learning, as illustrated in Figure 6 . We design an efficient search for depth image augmentations using object localization as a proxy task. Given the resulting sequence of random transformations, we use it to augment synthetic depth images during policy learning. Our augmentation strategy is policy-independent and enables policy learning with no real images. We demonstrate our approach to significantly improve accuracy on three manipulation tasks evaluated on a real robot.

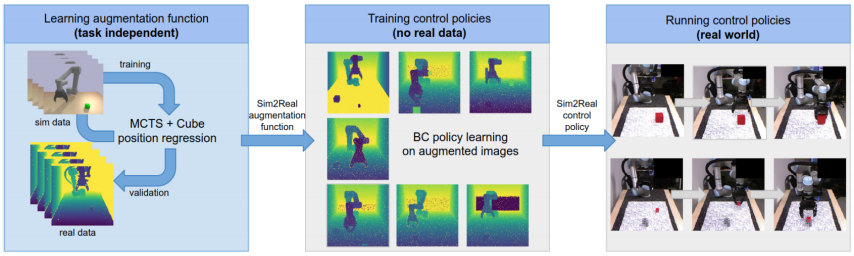


Figure 6. Overview of the method. Our contribution is the policy-independent learning of depth image augmentations (left). The resulting sequence of augmentations is applied to synthetic depth images while learning manipulation policies in a simulator (middle). The learned policies are directly applied to real robot scenes without finetuning on real images.

7.1.7. Learning to combine primitive skills: A step towards versatile robotic manipulation

Participants: Robin Strudel [Inria WILLOW], Alexander Pashevich, Igor Kalevatykh [Inria WILLOW], Ivan Laptev [Inria WILLOW], Josef Sivic [Inria WILLOW], Cordelia Schmid.

Manipulation tasks such as preparing a meal or assembling furniture remain highly challenging for robotics and vision. Traditional task and motion planning (TAMP) methods can solve complex tasks but require full state observability and are not adapted to dynamic scene changes. Recent learning methods can operate directly on visual inputs but typically require many demonstrations and/or task-specific reward engineering. In this work [40] we aim to overcome previous limitations and propose a reinforcement learning (RL) approach to task planning that learns to combine primitive skills illustrated in Figure 7 . First, compared to previous learning methods, our approach requires neither intermediate rewards nor complete task demonstrations during training. Second, we demonstrate the versatility of our vision-based task planning in challenging settings with temporary occlusions and dynamic scene changes. Third, we propose an efficient training of basic skills from few synthetic demonstrations by exploring recent CNN architectures and data augmentation. Notably, while all of our policies are learned on visual inputs in simulated environments, we demonstrate the successful transfer and high success rates when applying such policies to manipulation tasks on a real UR5 robotic arm.

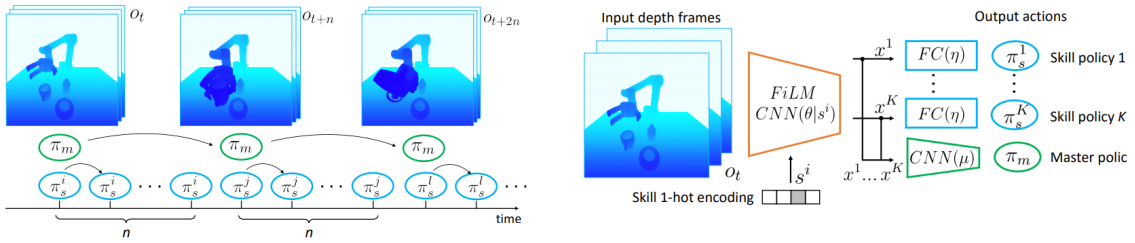


Figure 7. Illustration of our approach. (Left): Temporal hierarchy of master and skill policies. The master policy π_m is executed at a coarse interval of n time-steps to select among K skill policies $\pi_s^1 ... \pi_s^K$. Each skill policy generates control for a primitive action such as grasping or pouring. (Right): CNN architecture used for the skill and master policies.

7.1.8. Probabilistic Reconstruction Networks for 3D Shape Inference from a Single Image

Participants: Roman Klokov, Jakob Verbeek, Edmond Boyer [Inria Morpheo].

In our BMVC'19 paper [21], we study end-to-end learning strategies for 3D shape inference from images, in particular from a single image. Several approaches in this direction have been investigated that explore different shape representations and suitable learning architectures. We focus instead on the underlying probabilistic mechanisms involved and contribute a more principled probabilistic inference-based reconstruction framework, which we coin Probabilistic Reconstruction Networks. This framework expresses image conditioned 3D shape inference through a family of latent variable models, and naturally decouples the choice of shape representations from the inference itself. Moreover, it suggests different options for the image conditioning and allows training in two regimes, using either Monte Carlo or variational approximation of the marginal likelihood. Using our Probabilistic Reconstruction Networks we obtain single image 3D reconstruction results that set a new state of the art on the ShapeNet dataset in terms of the intersection over union and earth mover's distance evaluation metrics. Interestingly, we obtain these results using a basic voxel grid representation, improving over recent work based on finer point cloud or mesh based representations. In Figure 8 we show a schematic overview of our model.

7.1.9. Hierarchical Scene Coordinate Classification and Regression for Visual Localization

Participants: Xiaotian Li [Aalto Univ., Finland], Shuzhe Wang [Aalto Univ., Finland], Li Zhao [Aalto Univ., Finland], Jakob Verbeek, Juho Kannala [Aalto Univ., Finland].

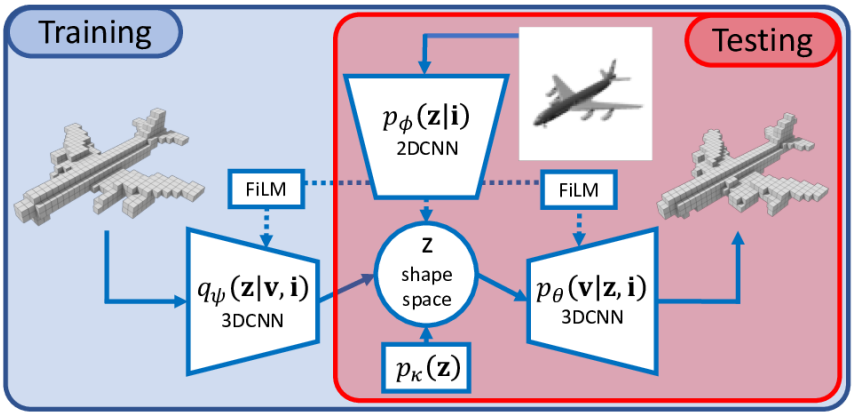


Figure 8. Probabilistic Reconstruction Networks for 3D shape inference from a single image. Arrows show the computational flow through the model, dotted arrows show optional image conditioning. Conditioning between 2D and 3D tensors is achieved by means of FiLM layers. The inference network q_ψ is only used during training for variational inference.

Visual localization is critical to many applications in computer vision and robotics. To address single-image RGB localization, state-of-the-art feature-based methods match local descriptors between a query image and a pre-built 3D model. Recently, deep neural networks have been exploited to regress the mapping between raw pixels and 3D coordinates in the scene, and thus the matching is implicitly performed by the forward pass through the network. However, in a large and ambiguous environment, learning such a regression task directly can be difficult for a single network. In our paper [37], we present a new hierarchical scene coordinate network to predict pixel scene coordinates in a coarse-to-fine manner from a single RGB image. The network consists of a series of output layers with each of them conditioned on the previous ones. The final output layer predicts the 3D coordinates and the others produce progressively finer discrete location labels. The proposed method outperforms the baseline regression-only network and allows us to train single compact models which scale robustly to large environments. It sets a new state-of-the-art for single-image RGB localization performance on the 7-Scenes, 12-Scenes, Cambridge Landmarks datasets, and three combined scenes. Moreover, for large-scale outdoor localization on the Aachen Day-Night dataset, our approach is much more accurate than existing scene coordinate regression approaches, and reduces significantly the performance gap w.r.t. explicit feature matching approaches. In Figure 9 we illustrate the scene coordinate predictions for the Aachen dataset experiments.

7.1.10. Moulding Humans: Non-parametric 3D Human Shape Estimation from Single Images

Participants: Valentin Gabeur, Jean-Sébastien Franco [Inria Morpheo], Xavier Martin, Cordelia Schmid, Gregory Rogez [NAVER LABS Europe].

While the recent progress in convolutional neural networks has allowed impressive results for 3D human pose estimation, estimating the full 3D shape of a person is still an open issue. Model-based approaches can output precise meshes of naked under-cloth human bodies but fail to estimate details and un-modelled elements such as hair or clothing. On the other hand, non-parametric volumetric approaches can potentially estimate complete shapes but, in practice, they are limited by the resolution of the output grid and cannot produce detailed estimates. In this paper [19], we propose a non-parametric approach that employs a double depth map 10 to represent the 3D shape of a person: a visible depth map and a "hidden" depth map are estimated and combined, to reconstruct the human 3D shape as done with a "mould". This representation through 2D

Figure 9. The scene coordinate predictions are visualized as 2D-2D matches between the query (left) and database (right) images. For each pair, the retrieved database image with the largest number of inliers is selected, and only the inlier matches are visualized. We show that our method is able to produce accurate correspondences for challenging queries.

depth maps allows a higher resolution output with a much lower dimension than voxel-based volumetric representations.

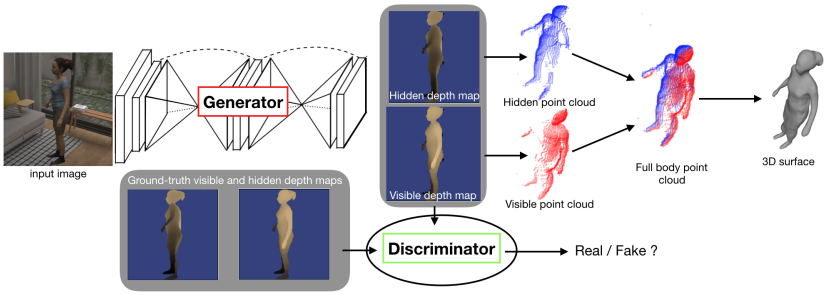


Figure 10. Given a single image, we estimate the "visible" and the "hidden" depth maps. The 3D point clouds of these 2 depth maps are combined to form a full-body 3D point cloud, as if lining up the 2 halves of a "mould". The 3D shape is then reconstructed using Poisson reconstruction. An adversarial training with a discriminator is employed to increase the humanness of the estimation.

7.1.11. Focused Attention for Action Recognition

Participants: Vladyslav Sydorov, Karteek Alahari.

In this paper [30], we introduce an attention model for video action recognition that allows processing video in higher resolution, by focusing on the relevant regions first. The network-specific saliency is utilized to guide the cropping, we illustrate the procedure in Figure 11 . We show performance improvement on the Charades dataset with this strategy.

7.2. Statistical Machine Learning

7.2.1. A Contextual Bandit Bake-off

Participants: Alberto Bietti, Alekh Agarwal [Microsoft Research], John Langford [Microsoft Research].

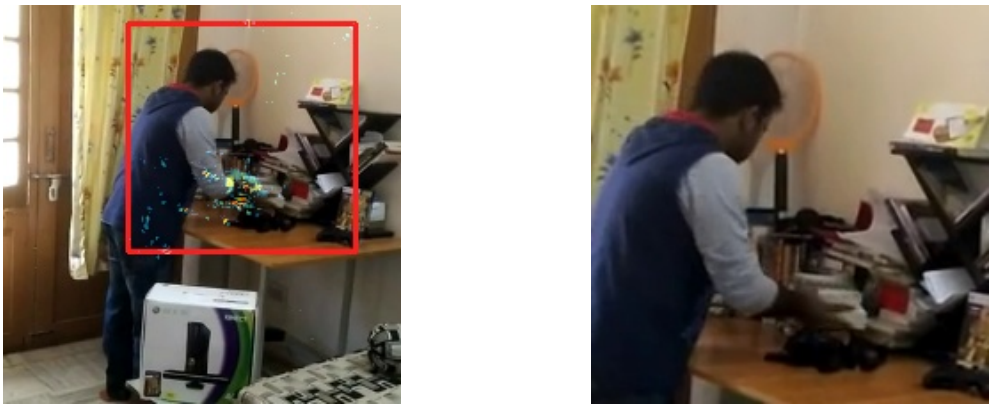


Figure 11. Example of attention on Charades action recognition dataset. (Left) Saliency scores (displayed as a heatmap) are localized around the object, a box maximizing the saliency measure within is selected. (Right) The network is provided with the relevant crop of the video, and can process it at a higher resolution.

Contextual bandit algorithms are essential for solving many real-world interactive machine learning problems. Despite multiple recent successes on statistically and computationally efficient methods, the practical behavior of these algorithms is still poorly understood. In , we leverage the availability of large numbers of supervised learning datasets to compare and empirically optimize contextual bandit algorithms, focusing on practical methods that learn by relying on optimization oracles from supervised learning. We find that a recent method using optimism under uncertainty works the best overall. A surprisingly close second is a simple greedy baseline that only explores implicitly through the diversity of contexts, followed by a variant of Online Cover which tends to be more conservative but robust to problem specification by design. Along the way, we also evaluate and improve several internal components of contextual bandit algorithm design. Overall, this is a thorough study and review of contextual bandit methodology.

7.2.2. A Generic Acceleration Framework for Stochastic Composite Optimization

Participants: Andrei Kulunchakov, Julien Mairal.

In [35], we introduce various mechanisms to obtain accelerated first-order stochasticoptimization algorithms when the objective function is convex or stronglyconvex. Specifically, we extend the Catalyst approach originally designed fordeterministic objectives to the stochastic setting. Given an optimizationmethod with mild convergence guarantees for strongly convex problems,the challenge is to accelerate convergence to a noise-dominated region, andthen achieve convergence with an optimal worst-case complexity depending on thenoise variance of the gradients.A side contribution of our work is also a generic analysis that canhandle inexact proximal operators, providing new insights about the robustness of stochastic algorithms when the proximal operator cannot be exactly computed. An illustration from this work is explained in Figure 12 .

7.2.3. Estimate Sequences for Variance-Reduced Stochastic Composite Optimization

Participants: Andrei Kulunchakov, Julien Mairal.

In [23], we propose a unified view of gradient-based algorithms for stochastic convex composite optimization. By extending the concept of estimate sequence introduced by Nesterov, we interpret a large class of stochastic optimization methods as procedures that iteratively minimize a surrogate of the objective. This point of view covers stochastic gradient descent (SGD), the variance-reduction approaches SAGA, SVRG, MISO, their proximal variants, and has several advantages: (i) we provide a simple generic proof of convergence for all of the aforementioned methods; (ii) we naturally obtain new algorithms with the same guarantees; (iii) we derive

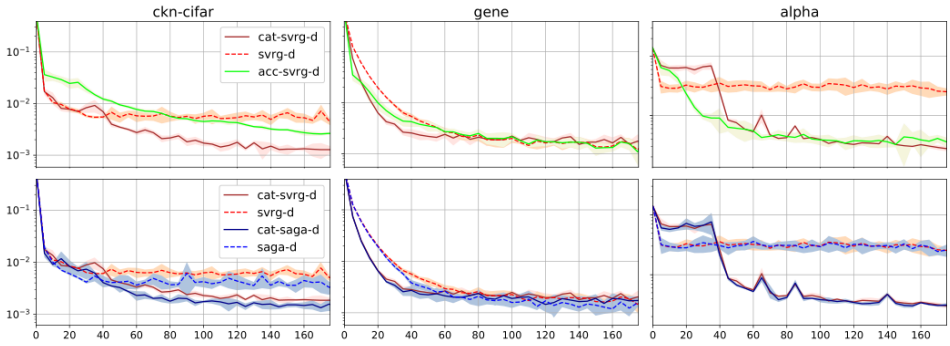


Figure 12. Accelerating SVRG-like (top) and SAGA (bottom) methods for ℓ_2-logistic regression with $\mu = 1/(100n)$ (bottom) for mild dropout, which imitates stochasticity in the gradients. All plots are on a logarithmic scale for the objective function value, and the x-axis denotes the number of epochs. The colored tubes around each curve denote a standard deviations across 5 runs. The curves show that acceleration may be useful even in the stochastic optimization regime.

generic strategies to make these algorithms robust to stochastic noise, which is useful when data is corrupted by small random perturbations. Finally, we show that this viewpoint is useful to obtain accelerated algorithms. A comparison with different approaches is shown in Figure 13 .

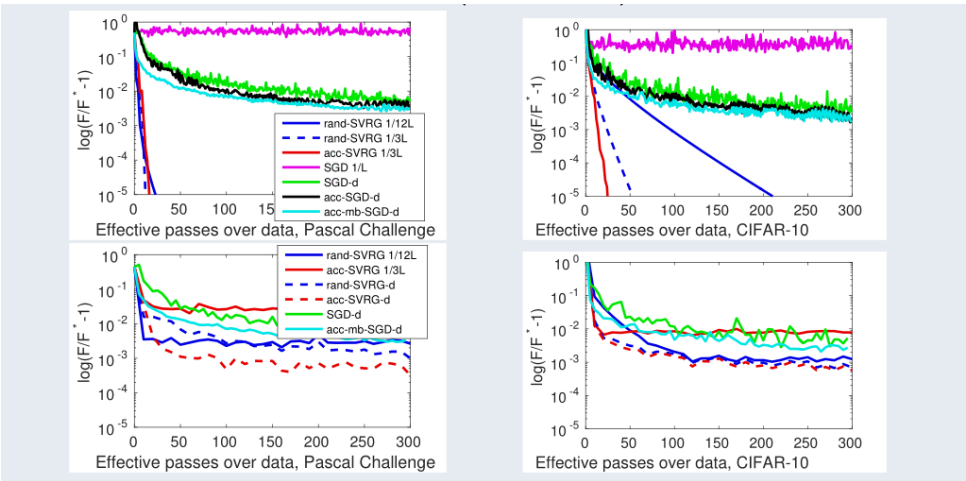


Figure 13. Comparison of different standard approaches with our developed method on two datasets for ℓ_2-logistic regression with mild dropout (bottom) and deterministic case (above). The case of exact gradient computations clearly shows benefits from acceleration, which consist in fast linear convergence. In the stochastic case, we demonstrate either superiority or high competitiveness of the developed method along with its unbiased convergence to the optimum. In both cases, we show that acceleration is able to generically comprise strengths of standard methods and even outperform them.

7.2.4. White-box vs Black-box: Bayes Optimal Strategies for Membership Inference

Participants: Alexandre Sablayrolles, Matthijs Douze, Yann Ollivier, Cordelia Schmid, Hervé Jégou.

Membership inference determines, given a sample and trained parameters of a machine learning model, whether the sample was part of the training set. In this paper [28], we derive the optimal strategy for membership inference with a few assumptions on the distribution of the parameters. We show that optimal attacks only depend on the loss function, and thus black-box attacks are as good as white-box attacks. As the optimal strategy is not tractable, we provide approximations of it leading to several inference methods 14 , and show that existing membership inference methods are coarser approximations of this optimal strategy. Our membership attacks outperform the state of the art in various settings, ranging from a simple logistic regression to more complex architectures and datasets, such as ResNet-101 and Imagenet.

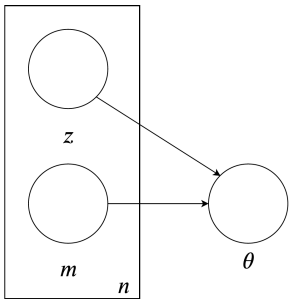


Figure 14. Plate notation of the membership inference problem: for each data point z_i, a binary membership variable m_i is sampled, and z_i belongs to the training set iff $m_i = 1$. Given the trained parameters θ and a sample z_i, we want to infer the value of m_i.

7.3. Theory and Methods for Deep Neural Networks

7.3.1. Group Invariance, Stability to Deformations, and Complexity of Deep Convolutional Representations

Participants: Alberto Bietti, Julien Mairal.

The success of deep convolutional architectures is often attributed in part to their ability to learn multiscale and invariant representations of natural signals. However, a precise study of these properties and how they affect learning guarantees is still missing. In the paper [3], we consider deep convolutional representations of signals; we study their invariance to translations and to more general groups of transformations, their stability to the action of diffeomorphisms, and their ability to preserve signal information. This analysis is carried by introducing a multilayer kernel based on convolutional kernel networks and by studying the geometry induced by the kernel mapping. We then characterize the corresponding reproducing kernel Hilbert space (RKHS), showing that it contains a large class of convolutional neural networks with homogeneous activation functions. This analysis allows us to separate data representation from learning, and to provide a canonical measure of model complexity, the RKHS norm, which controls both stability and generalization of any learned model. In addition to models in the constructed RKHS, our stability analysis also applies to convolutional networks with generic activations such as rectified linear units, and we discuss its relationship with recent generalization bounds based on spectral norms.

7.3.2. A Kernel Perspective for Regularizing Deep Neural Networks

Participants: Alberto Bietti, Grégoire Mialon, Dexiong Chen, Julien Mairal.

We propose a new point of view for regularizing deep neural networks by using the norm of a reproducing kernel Hilbert space (RKHS) [12]. Even though this norm cannot be computed, it admits upper and lower approximations leading to various practical strategies. Specifically, this perspective (i) provides a common umbrella for many existing regularization principles, including spectral norm and gradient penalties, or adversarial training, (ii) leads to new effective regularization penalties, and (iii) suggests hybrid strategies combining lower and upper bounds to get better approximations of the RKHS norm. We experimentally show this approach to be effective when learning on small datasets, or to obtain adversarially robust models.

7.3.3. On the Inductive Bias of Neural Tangent Kernels

Participants: Alberto Bietti, Julien Mairal.

State-of-the-art neural networks are heavily over-parameterized, making the optimization algorithm a crucial ingredient for learning predictive models with good generalization properties. A recent line of work has shown that in a certain over-parameterized regime, the learning dynamics of gradient descent are governed by a certain kernel obtained at initialization, called the neural tangent kernel. In [12], we study the inductive bias of learning in such a regime by analyzing this kernel and the corresponding function space (RKHS). In particular, we study smoothness, approximation, and stability properties of functions with finite norm, including stability to image deformations in the case of convolutional networks, and compare to other known kernels for similar architectures.

7.3.4. Large Memory Layers with Product Keys

Participants: Guillaume Lample, Alexandre Sablayrolles, Marc'Aurelio Ranzato, Ludovic Denoyer, Hervé Jégou.

This paper introduces a structured memory which can be easily integrated into a neural network. The memory is very large by design and significantly increases the capacity of the architecture, by up to a billion parameters with a negligible computational overhead. Its design and access pattern is based on product keys, which enable fast and exact nearest neighbor search. The ability to increase the number of parameters while keeping the same computational budget lets the overall system strike a better trade-off between prediction accuracy and computation efficiency both at training and test time. This memory layer, shown in Figure 15 , allows us to tackle very large scale language modeling tasks. In our experiments we consider a dataset with up to 30 billion words, and we plug our memory layer in a state-of-the-art transformer-based architecture. In particular, we found that a memory augmented model with only 12 layers outperforms a baseline transformer model with 24 layers, while being twice faster at inference time. We release our code for reproducibility purposes.

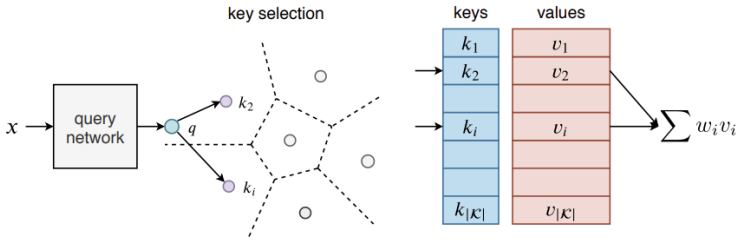


Figure 15. Overview of a key-value memory layer: The input x is processed through a query network that produces a query vector q, which is compared to all the keys. The output is the sparse weighted sum over the memories associated with the selected keys. For a large number of keys $|\mathcal{K}|$, the key selection procedure becomes too expensive in practice. Our product key method is exact and makes this search process very fast.

7.3.5. Understanding Priors in Bayesian Neural Networks at the Unit Level

Participants: Mariia Vladimirova, Jakob Verbeek, Pablo Mesejo [Univ. Granada, Spain], Julyan Arbel [Inria MISTIS].

In our ICML'19 paper [31], we investigate deep Bayesian neural networks with Gaussian weight priors and a class of ReLUlike nonlinearities. Bayesian neural networks with Gaussian priors are well known to induce an L2, "weight decay", regularization. Our results characterize a more intricate regularization effect at the level of the unit activations. Our main result establishes that the induced prior distribution on the units before and after activation becomes increasingly heavy-tailed with the depth of the layer. We show that first layer units are Gaussian, second layer units are sub-exponential, and units in deeper layers are characterized by sub-Weibull distributions. Our results provide new theoretical insight on deep Bayesian neural networks, which we corroborate with experimental simulation results.

7.3.6. Adaptative Inference Cost With Convolutional Neural Mixture Models

Participants: Adria Ruiz, Jakob Verbeek.

Despite the outstanding performance of convolutional neural networks (CNNs) for many vision tasks, the required computational cost during inference is problematic when resources are limited. In this paper [27], we propose Convolutional Neural Mixture Models (CNMMs), a probabilistic model embedding a large number of CNNs that can be jointly trained and evaluated in an efficient manner. Within the proposed framework, we present different mechanisms to prune subsets of CNNs from the mixture, allowing to easily adapt the computational cost required for inference (see Fig. 16). Image classification and semantic segmentation experiments show that our method achieve excellent accuracy-compute trade-offs. Moreover, unlike most of previous approaches, a single CNMM provides a large range of operating points along this trade-off, without any re-training.

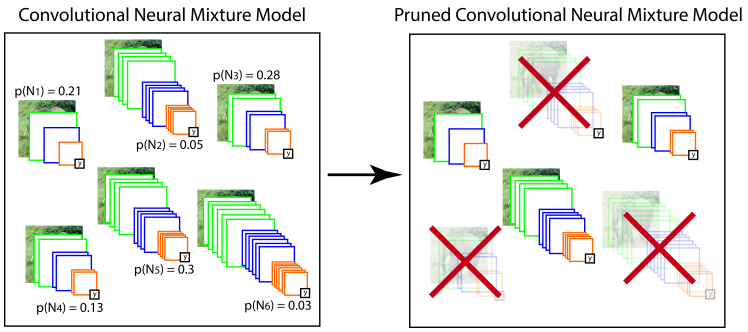


Figure 16. A Convolutional Neural Mixture Model embeds a large number of CNNs. Weight sharing enables efficient joint training of all networks and computation of the mixture output. The learned mixing weights can be used to remove networks from the mixture, and thus reduce the computational cost of inference.

7.4. Pluri-disciplinary Research

7.4.1. Biological Sequence Modeling with Convolutional Kernel Networks

Participants: Dexiong Chen, Laurent Jacob, Julien Mairal.

The growing number of annotated biological sequences available makes it possible to learn genotype-phenotype relationships from data with increasingly high accuracy. When large quan- tities of labeled samples are available for training a model, convolutional neural networks can be used to predict the phenotype of unannotated sequences with good accuracy. Unfortunately, their performance with medium- or small-scale datasets is mitigated, which requires inventing new data-efficient approaches. In this paper [4], [14], we introduce a hybrid approach between convolutional neural networks and kernel methods to model biological sequences. Our method, shown in Figure 17 , enjoys the ability of convolutional neural networks to learn data representations that are adapted to a specific task, while the kernel point of view yields algorithms that perform significantly better when the amount of training data is small. We illustrate these advantages for transcription factor binding prediction and protein homology detection, and we demonstrate that our model is also simple to interpret, which is crucial for discovering predictive motifs in sequences. The source code is freely available at https://gitlab.inria.fr/dchen/CKN-seq.

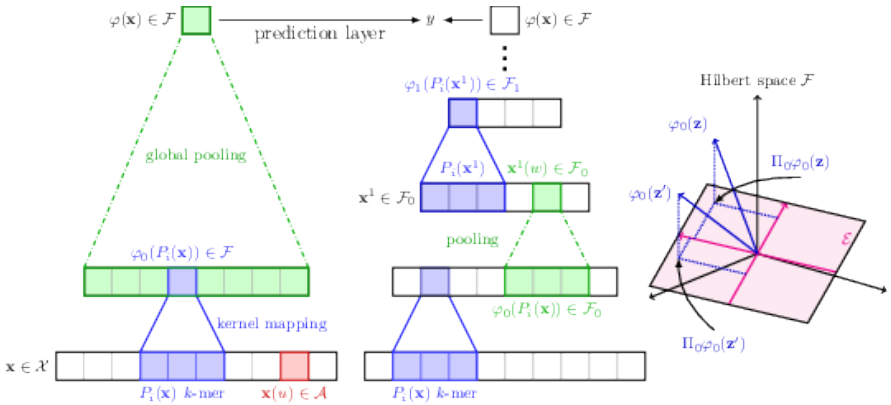


Figure 17. Construction of single-layer (left) and multilayer (middle) CKN-seq and the approximation of one layer (right). For a single-layer model, each k-mer $P_i(\mathbf{x})$ is mapped to $\varphi_0(P_i(\mathbf{x}))$ in $\mathcal{F}$ and projected to $\Pi\varphi_0(P_i(\mathbf{x}))$ parametrized by $\psi_0(P_i(\mathbf{x}))$. Then, the final finite-dimensional sequence is obtained by the global pooling, $\psi(\mathbf{x}) = \frac{1}{m}\sum_{i=0}^{m}\psi_0(P_i(\mathbf{x}))$. The multilayer construction is similar, but relies on intermediate maps, obtained by local pooling.

7.4.2. Recurrent Kernel Networks

Participants: Dexiong Chen, Laurent Jacob [CNRS, LBBE Laboratory], Julien Mairal.

Substring kernels are classical tools for representing biological sequences or text. However, when large amounts of annotated data are available, models that allow end-to-end training such as neural networks are often preferred. Links between recurrent neural networks (RNNs) and substring kernels have recently been drawn, by formally showing that RNNs with specific activation functions were points in a reproducing kernel Hilbert space (RKHS). In this paper [15], we revisit this link by generalizing convolutional kernel networks——originally related to a relaxation of the mismatch kernel——to model gaps in sequences. It results in a new type of recurrent neural network (Figure 18), which can be trained end-to-end with backpropagation, or without supervision by using kernel approximation techniques. We experimentally show that our approach is well suited to biological sequences, where it outperforms existing methods for protein classification tasks.

7.4.3. Depth-adaptive Transformer

Participants: Maha Elbayad, Jiatao Gu [Facebook AI], Edouard Grave [Facebook AI], Michael Auli [Facebook AI].

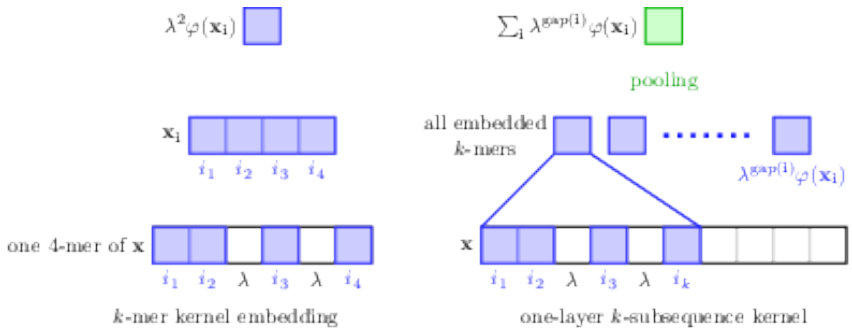


Figure 18. Representation of a sequence in a RKHS based on our kernel.

State of the art sequence-to-sequence models for large scale tasks perform a fixed number of computations for each input sequence regardless of whether it is easy or hard to process. In our ICLR'2020 paper [18], we train Transformer models which can make output predictions at different stages of the network and we investigate different ways to predict how much computation is required for a particular sequence. Unlike dynamic computation in Universal Transformers, which applies the same set of layers iteratively, we apply different layers at every step to adjust both the amount of computation as well as the model capacity. On IWSLT German-English translation our approach matches the accuracy of a well tuned baseline Transformer while using less than a quarter of the decoder layers. Figure 19 illustrates the different halting mechanisms investigated in this work. Namely, a sequence-level approach where we assume all the sequence's tokens are equally difficult and a token-level approach where tokens exit at varying depths.

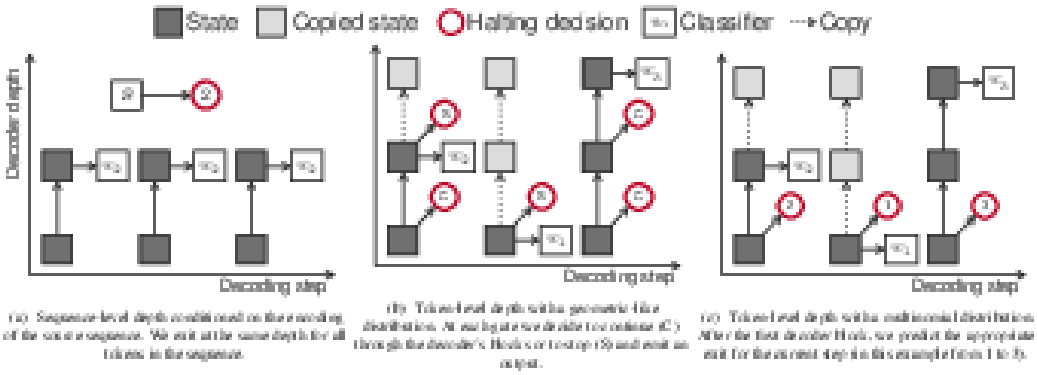


Figure 19. Illustration of the variant adaptive depth predictors: (a) the sequence-level and (b, c) at the token-level.

WILLOW Team

7. New Results

7.1. 3D object and scene modeling, analysis, and retrieval

7.1.1. Learning joint reconstruction of hands and manipulated objects

Participants: Yana Hasson, Gül Varol, Dimitrios Tzionas, Igor Kalevatykh, Michael Black, Ivan Laptev, Cordelia Schmid.

Estimating hand-object manipulations is essential for interpreting and imitating human actions. Previous work has made significant progress towards reconstruction of hand poses and object shapes in isolation. Yet, reconstructing hands and objects during manipulation is a more challenging task due to significant occlusions of both the hand and object. While presenting challenges, manipulations may also simplify the problem since the physics of contact restricts the space of valid hand-object configurations. For example, during manipulation, the hand and object should be in contact but not interpenetrate. In [14] we regularize the joint reconstruction of hands and objects with manipulation constraints. We present an end-to-end learnable model that exploits a novel contact loss that favors physically plausible hand-object constellations. Our approach improves grasp quality metrics over baselines, using RGB images as input. To train and evaluate the model, we also propose a new large-scale synthetic dataset, ObMan, with hand-object manipulations. We demonstrate the transferability of ObMan-trained models to real data. Figure 1 presents some example results.

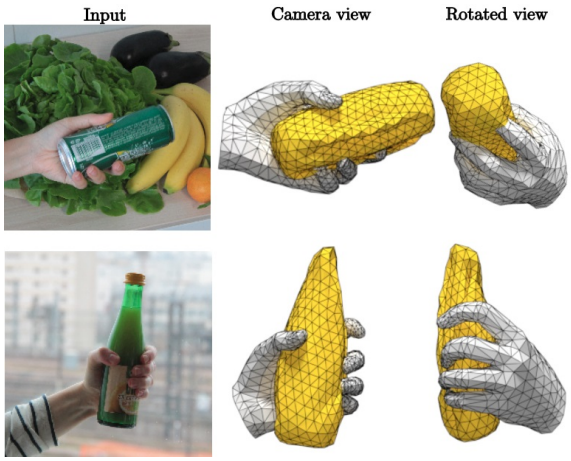


Figure 1. Our method jointly reconstructs hand and object meshes from a monocular RGB image. Note that the model generating the predictions for the above images, which we captured with an ordinary camera, was trained only on images from our synthetic dataset, ObMan.

7.1.2. D2-Net: A Trainable CNN for Joint Detection and Description of Local Features

Participants: Mihai Dusmanu, Ignacio Rocco, Tomas Pajdla, Marc Pollefeys, Josef Sivic, Akihiko Torii, Torsten Sattler.

In [13], we address the problem of finding reliable pixel-level correspondences under difficult imaging conditions. We propose an approach where a single convolutional neural network plays a dual role: It is simultaneously a dense feature descriptor and a feature detector, as illustrated in Figure 2 . By postponing the detection to a later stage, the obtained keypoints are more stable than their traditional counterparts based on early detection of low-level structures. We show that this model can be trained using pixel correspondences extracted from readily available large-scale SfM reconstructions, without any further annotations. The proposed method obtains state-of-the-art performance on both the difficult Aachen Day-Night localization dataset and the InLoc indoor localization benchmark, as well as competitive performance on other benchmarks for image matching and 3D reconstruction.

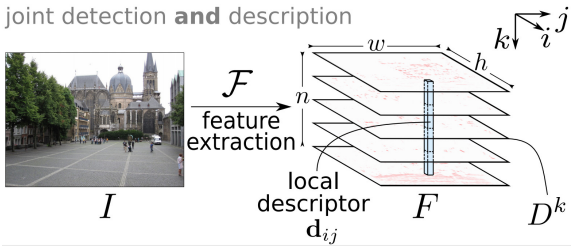


Figure 2. A feature extraction CNN $\mathcal{F}$ is used to extract feature maps that play a dual role: (i) local descriptors $\mathbf{d}_{ij}$ are simply obtained by traversing all the n feature maps D^k at a spatial position (i, j); (ii) detections are obtained by performing a non-local-maximum suppression on a feature map followed by a non-maximum suppression across each descriptor.

7.1.3. Is This The Right Place? Geometric-Semantic Pose Verification for Indoor Visual Localization

Participants: Hajime Taira, Ignacio Rocco, Jiri Sedlar, Masatoshi Okutomi, Josef Sivic, Tomas Pajdla, Torsten Sattler, Akihiko Torii.

Visual localization in large and complex indoor scenes, dominated by weakly textured rooms and repeating geometric patterns, is a challenging problem with high practical relevance for applications such as Augmented Reality and robotics. To handle the ambiguities arising in this scenario, a common strategy is, first, to generate multiple estimates for the camera pose from which a given query image was taken. The pose with the largest geometric consistency with the query image, e.g., in the form of an inlier count, is then selected in a second stage. While a significant amount of research has concentrated on the first stage, there is considerably less work on the second stage. In [21], we thus focus on pose verification. We show that combining different modalities, namely appearance, geometry, and semantics, considerably boosts pose verification and consequently pose accuracy, as illustrated in Figure 3 . We develop multiple hand-crafted as well as a trainable approach to join into the geometric-semantic verification and show significant improvements over state-of-the-art on a very challenging indoor dataset.

7.1.4. An Efficient Solution to the Homography-Based Relative Pose Problem With a Common Reference Direction

Participants: Yaqing Ding, Jian Yang, Jean Ponce, Hui Kong.

In [12], we propose a novel approach to two-view minimal-case relative pose problems based on homography with a common reference direction. We explore the rank-1 constraint on the difference between the Euclidean homography matrix and the corresponding rotation, and propose an efficient two-step solution for solving both the calibrated and partially calibrated (unknown focal length) problems. We derive new 3.5-point, 3.5-point, 4-point solvers for two cameras such that the two focal lengths are unknown but equal, one of them is unknown,

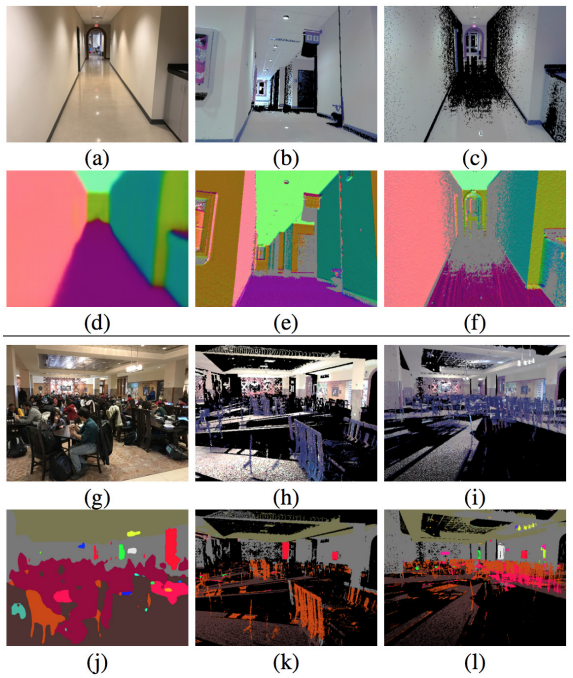


Figure 3. Given a set of camera pose estimates for a query image (a, g), we seek to identify the most accurate estimate. (b, h) Due to severe occlusion and weak textures, a state-of-the-art methodfails to identify the correct camera pose. To overcome those difficulties, we use several modalities along with visual appearance: (top) surface normals and (bottom) semantics. (c, i) Our approach verifies the estimated pose by comparing the semantics and surface normals extracted from the query (d, j) and database (f, l).

and both are unknown and possibly different, respectively. We present detailed analyses and comparisons with existing 6-and 7-point solvers, including results with smart phone images.

7.1.5. Coordinate-Free Carlsson-Weinshall Duality and Relative Multi-ViewGeometry

Participants: Matthew Trager, Martial Hebert, Jean Ponce.

In [23], we present a coordinate-free description of Carlsson-Weinshall duality between scene points and camera pinholes and use it to derive a new characterization of primal/dual multi-view geometry. In the case of three views, a particular set of reduced trilinearities provide a novel parameterization of camera geometry that, unlike existing ones, is subject only to very simple internal constraints. These trilinearities lead to new "quasi-linear" algorithms for primal and dual structure from motion. We include some preliminary experiments with real and synthetic data.

7.1.6. Build your own hybrid thermal/EO camera for autonomous vehicle

Participants: Yigong Zhang, Yicheng Gao, Shuo Gu, Yubin Guo, Minghao Liu, Zezhou Sun, Zhixing Hou, Hang Yang, Ying Wang, Jian Yang, Jean Ponce, Hui Kong.

In [24], we propose a novel paradigm to design a hybrid thermal/EO (Electro-Optical or visible-light) camera, whose thermal and RGB frames are pixel-wisely aligned and temporally synchronized. Compared with the existing schemes, we innovate in three ways in order to make it more compact in dimension, and thus more practical and extendable for real-world applications. The first is a redesign of the structure layout of the thermal and EO cameras. The second is on obtaining a pixel-wise spatial registration of the thermal and RGB frames by a coarse mechanical adjustment and a fine alignment through a constant homography warping. The third innovation is on extending one single hybrid camera to a hybrid camera array, through which we can obtain wide-view spatially aligned thermal, RGB and disparity images simultaneously. The experimental results show that the average error of spatial-alignment of two image modalities can be less than one pixel. Some results of our method are illustrated in Figure 4 .

7.2. Category-level object and scene recognition

7.2.1. Detecting unseen visual relations using analogies

Participants: Julia Peyre, Ivan Laptev, Cordelia Schmid, Josef Sivic.

In [19], we seek to detect visual relations in images of the form of triplets t = (subject, predicate, object), such as "person riding dog", where training examples of the individual entities are available but their combinations are unseen at training. This is an important set-up due to the combinatorial nature of visual relations: collecting sufficient training data for all possible triplets would be very hard. The contributions of this work are three-fold. First, we learn a representation of visual relations that combines (i) individual embeddings for subject, object and predicate together with (ii) a visual phrase embedding that represents the relation triplet. Second, we learn how to transfer visual phrase em- beddings from existing training triplets to unseen test triplets using analogies between relations that involve similar objects. Third, we demonstrate the benefits of our approach on three challenging datasets : on HICO-DET, our model achieves significant improvement over a strong baseline for both frequent and unseen triplets, and we observe similar improvement for the retrieval of unseen triplets with out-of- vocabulary predicates on the COCO-a dataset as well as the challenging unusual triplets in the UnRel dataset. Figure 5 presents an illustration of the approach.

7.2.2. SFNet: Learning Object-aware Semantic Correspondence

Participants: Junghyup Lee, Dohyung Kim, Jean Ponce, Bumsub Ham.

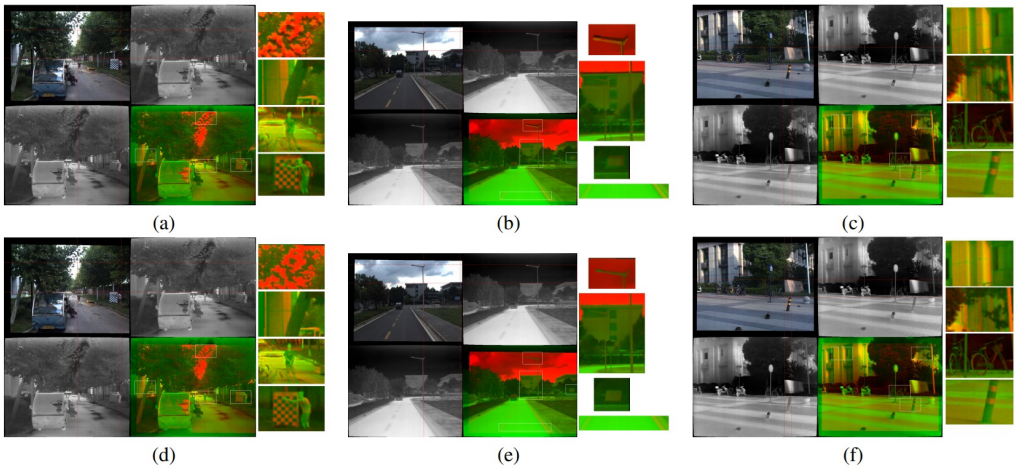


Figure 4. Results of alignment between the thermal and RGB frames of three sets of hybrid cameras before and after homography warping, respectively. (a), (b) and (c) are the alignment results before the homography warping, respectively. In each sub-figure, the layout of images is arranged as follows. Top-left: the aligned RGB image. Top-middle and bottom-left: the same aligned thermal image. Bottom-middle: the fusion image. (d), (e) and (f) are the alignment results after the homography warping, respectively. Likewise, the layout of images in each sub-figure is the same as those of (a), (b) and (c). To show the effect of homography rectification, we have overlaid red dotted lines horizontally and vertically onto the each sub-figure. In addition, the right column of each sub-figure zooms in four selected image regions to help us to view the warping result.

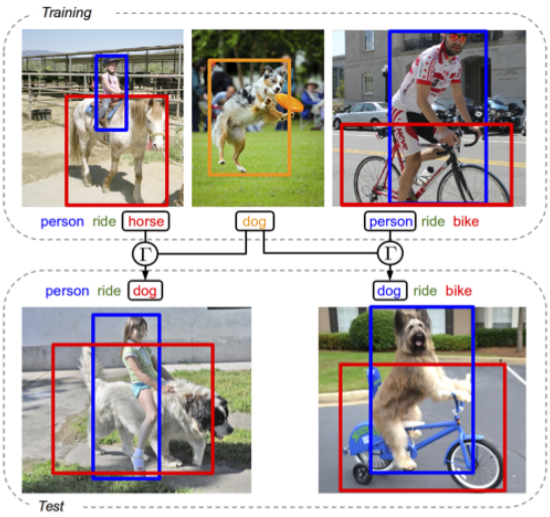


Figure 5. Illustration of transfer by analogy. We transfer visual representations of relations seen in the training set such as "person ride horse" to represent new unseen relations in the test set such as "person ride dog".

In [15], we address the problem of semantic correspondence, that is, establishing a dense flow field between images depicting different instances of the same object or scene category. We propose to use images annotated with binary foreground masks and subjected to synthetic geometric deformations to train a convolutional neural network (CNN) for this task. Using these masks as part of the supervisory signal offers a good compromise between semantic flow methods, where the amount of training data is limited by the cost of manually selecting point correspondences, and semantic alignment ones, where the regression of a single global geometric transformation between images may be sensitive to image-specific details such as background clutter. We propose a new CNN architecture, dubbed SFNet, which implements this idea. It leverages a new and differentiable version of the argmax function for end-to-end training, with a loss that combines mask and flow consistency with smoothness terms. Experimental results demonstrate the effectiveness of our approach, which significantly outperforms the state of the art on standard benchmarks. Figure 6 presents an illustration of the approach.

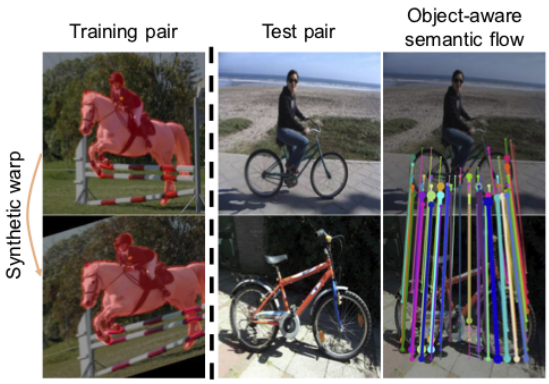


Figure 6. We use pairs of warped foreground masks obtained from a single image (left) as a supervisory signal to train our model.This allows us to establish object-aware semantic correspondences across images depicting different instances of the same object or scene category (right). No masks are required at test time.

7.2.3. Hyperpixel Flow: Semantic Correspondence with Multi-layer Neural Features

Participants: Juhong Min, Jongmin Kim, Jean Ponce, Minsu Cho.

In [17], we establish visual correspondences under large intra-class variations requires analyzing images at different levels , from features linked to semantics and context to local patterns, while being invariant to instance-specific details. To tackle these challenges, we represent images by "hyper-pixels" that leverage a small number of relevant features selected among early to late layers of a convolutional neural network. Taking advantage of the condensed features of hyperpixels, we develop an effective real-time matching algorithm based on Hough geometric voting. The proposed method, hyperpixel flow, sets a new state of the art on three standard benchmarks as well as a new dataset, SPair-71k, which contains a significantly larger number of image pairs than existing datasets, with more accurate and richer annotations for in-depth analysis. Figure 7 presents an illustration of the approach.

7.2.4. Exploring Weight Symmetry in Deep Neural Networks

Participants: Xu Shell Hu, Sergey Zagoruyko, Nikos Komodakis.

In [7], we propose to impose symmetry in neural network parameters to improve parameter usage and make use of dedicated convolution and matrix multiplication routines. Due to significant reduction in the number of parameters as a result of the symmetry constraints, one would expect a dramatic drop in accuracy. Surprisingly,

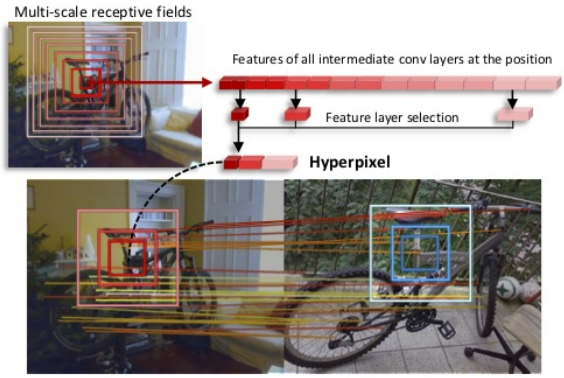


Figure 7. Hyperpixel flow. Top: The hyperpixel is a multi-layer pixel representation created with selected levels of features optimized for semantic correspondence. It provides multi-scale features, resolving local ambiguities. Bottom: The proposed method, hyperpixel flow, establishes dense correspondences in real time using hyperpixels.

we show that this is not the case, and, depending on network size, symmetry can have little or no negative effect on network accuracy, especially in deep overparameterized networks. We propose several ways to impose local symmetry in recurrent and convolutional neural networks, and show that our symmetry parameterizations satisfy universal approximation property for single hidden layer networks. We extensively evaluate these parameterizations on CIFAR, ImageNet and language modeling datasets, showing significant benefits from the use of symmetry. For instance, our ResNet-101 with channel-wise symmetry has almost 25% less parameters and only 0.2% accuracy loss on ImageNet.

7.2.5. Bilinear image translation for temporal analysis of photo collections

Participants: Théophile Dalens, Mathieu Aubry, Josef Sivic.

In [5], we propose an approach for analyzing unpaired visual data annotated with time stamps by generating how images would have looked like if they were from different times. To isolate and transfer time dependent appearance variations, we introduce a new trainable bilinear factor separation module. We analyze its relation to classical factored representations and concatenation-based auto-encoders. We demonstrate this new module has clear advantages compared to standard concatenation when used in a bottleneck encoder-decoder convolutional neural network architecture. We also show that it can be inserted in a recent adversarial image translation architecture, enabling the image transformation to multiple different target time periods using a single network. We apply our model to a challenging collection of more than 13,000 cars manufactured between 1920 and 2000 and a dataset of high school yearbook portraits from 1930 to 2009, as illustrated in Figure 8 . This allows us, for a given new input image, to generate a "history-lapse video" revealing changes over time by simply varying the target year. We show that by analyzing the generated history-lapse videos we can identify object deformations across time, extracting interesting changes in visual style over decades.

7.3. Image restoration, manipulation and enhancement

7.3.1. Deformable Kernel Networks for Joint Image Filtering

Participants: Beomjun Kim, Jean Ponce, Bumsub Ham.

Figure 8. Our method takes as input an image of an object (in green), such as a car, and generates what it would have looked like in another time-period (in blue). Each row shows temporal translation for a different input car image (in green). The translation model is trained on an unpaired dataset of cars with time stamps. We show that analyzing changes between the generated images reveal structural deformations in car shape and appearance over time.

Joint image filters are used to transfer structural details from a guidance picture used as a prior to a target image, in tasks such as enhancing spatial resolution and suppressing noise. Previous methods based on convolutional neural networks (CNNs) combine nonlinear activations of spatially-invariant kernels to estimate structural details and regress the filtering result. In this paper, we instead learn explicitly sparse and spatially-variant kernels. In [28], we propose a CNN architecture and its efficient implementation, called the deformable kernel network (DKN), that outputs sets of neighbors and the corresponding weights adaptively for each pixel. The filtering result is then computed as a weighted average. We also propose a fast version of DKN that runs about four times faster for an image of size 640×480. We demonstrate the effectiveness and flexibility of our models on the tasks of depth map upsampling, saliency map upsampling, cross-modality image restoration, texture removal, and semantic segmentation. In particular, we show that the weighted averaging process with sparsely sampled 3×3 kernels outperforms the state of the art by a significant margin.

7.3.2. Revisiting Non Local Sparse Models for Image Restoration

Participants: Bruno Lecouat, Jean Ponce, Julien Mairal.

In [29], we propose a differentiable algorithm for image restoration inspired by the success of sparse models and self-similarity priors for natural images. Our approach builds upon the concept of joint sparsity between groups of similar image patches, and we show how this simple idea can be implemented in a differentiable architecture, allowing end-to-end training. The algorithm has the advantage of being interpretable, performing sparse decompositions of image patches, while being more parameter efficient than recent deep learning methods. We evaluate our algorithm on grayscale and color denoising, where we achieve competitive results, and on demoisaicking, where we outperform the most recent state-of-the-art deep learning model with 47 times less parameters and a much shallower architecture. Figure 9 shows results of the proposed approach.

7.4. Human activity capture and classification

7.4.1. Video Face Clustering with Unknown Number of Clusters

Participants: Makarand Tapaswi, Marc T. Law, Sanja Fidler.

Figure 9. Demosaicking result obtained by our method. Top right: Ground truth. Middle: Image demosaicked with our sparse coding baseline without non-local prior. Bottom: demosaicking with sparse coding and non-local prior. The reconstruction does not exhibit any artefact on this image which is notoriously difficult for demosaicking.

Understanding videos such as TV series and movies requires analyzing who the characters are and what they are doing. We address the challenging problem of clustering face tracks based on their identity. Different from previous work in this area, we choose to operate in a realistic and difficult setting where: (i) the number of characters is not known a priori; and (ii) face tracks belonging to minor or background characters are not discarded.

To this end, we propose Ball Cluster Learning (BCL), a supervised approach to carve the embedding space into balls of equal size, one for each cluster (see Figure 10). The learned ball radius is easily translated to a stopping criterion for iterative merging algorithms. This gives BCL the ability to estimate the number of clusters as well as their assignment, achieving promising results on commonly used datasets. We also present a thorough discussion of how existing metric learning literature can be adapted for this task. This work has been published in [22].

7.4.2. Cross-task weakly supervised learning from instructional videos

Participants: Dimitri Zhukov, Jean-Baptiste Alayrac, Ramazan Gokberk Cinbis, David Fouhey, Ivan Laptev, Josef Sivic.

In [25], we investigate learning visual models for the steps of ordinary tasks using weak supervision via instructional narrations and an ordered list of steps instead of strong supervision via temporal annotations. At the heart of our approach is the observation that weakly supervised learning may be easier if a model shares components while learning different steps: “pour egg” should be trained jointly with other tasks involving “pour” and “egg”. We formalize this in a component model for recognizing steps and a weakly supervised learning framework that can learn this model under temporal constraints from narration and the list of steps. Past data does not permit systematic studying of sharing and so we also gather a new dataset, CrossTask, aimed at assessing cross-task sharing. Our experiments demonstrate that sharing across tasks im- proves performance, especially when done at the component level and that our component model can parse previously unseen tasks by virtue of its compositionality. Figure 11 illustrates the idea of sharing step components between different tasks.

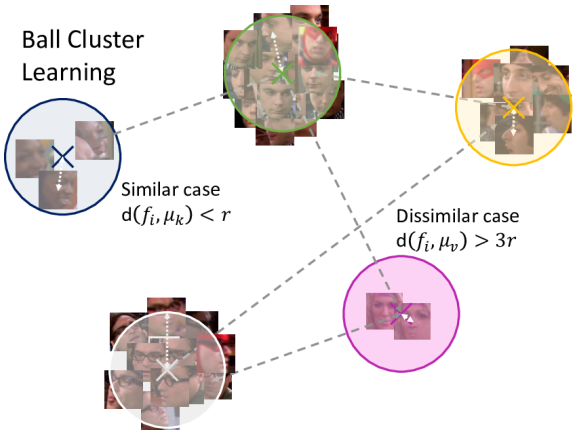


Figure 10. Ball Cluster Learning carves the feature space into balls of equal radius. The number of samples in the cluster does not affect the ball radius or minimum separation to other balls.

Figure 11. Our method begins with a collection of tasks, each con- sisting of an ordered list of steps and a set of instructional videos from YouTube. It automatically discovers both where the steps oc- cur and what they look like. To do this, it uses the order, narration and commonalities in appearance across tasks (e.g., the appear- ance of pour in both making pancakes and making meringue).

7.4.3. Leveraging the Present to Anticipate the Future in Videos

Participants: Antoine Miech, Ivan Laptev, Josef Sivic, Heng Wang, Lorenzo Torresani, Du Tran.

Anticipating actions before they are executed is crucial for a wide range of practical applications including autonomous driving and the moderation of live video streaming. While most prior work in this area requires partial observation of executed actions, in the paper we focus on anticipating actions seconds before they start (see Figure 12) Our proposed approach is the fusion of a purely anticipatory model with a complementary model constrained to reason about the present. In particular, the latter predicts present action and scene attributes, and reasons about how they evolve over time. By doing so, we aim at modeling action anticipation at a more conceptual level than directly predicting future actions. Our model outperforms previously reported methods on the EPIC-KITCHENS and Breakfast datasets. This paper was presented at the CVPR 2019 precognition workshop [34] and ranked second at the EPIC-KITCHENS action anticipation challenge.

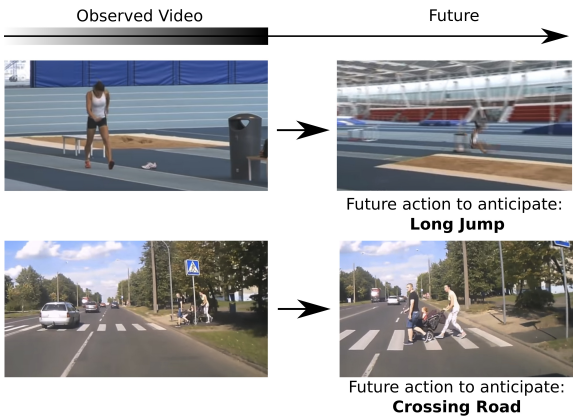


Figure 12. Examples of action anticipation in which the goal is to anticipate future actions in videos seconds before they are performed.

7.4.4. HowTo100M: Learning a Text-Video Embedding by Watching Hundred Million Narrated Video Clips

Participants: Antoine Miech, Dimitri Zhukov, Jean-Baptiste Alayrac, Makarand Tapaswi, Ivan Laptev, Josef Sivic.

Learning text-video embeddings usually requires a dataset of video clips with manually provided captions. However, such datasets are expensive and time consuming to create and therefore difficult to obtain on a large scale. In this work, we propose instead to learn such embeddings from video data with readily available natural language annotations in the form of automatically transcribed narrations (see Figure 13). The contributions of this work are three-fold. First, we introduce HowTo100M: a large-scale dataset of 136 million video clips sourced from 1.22M narrated instructional web videos depicting humans performing and describing over 23k different visual tasks. Our data collection procedure is fast, scalable and does not require any additional manual annotation. Second, we demonstrate that a text-video embedding trained on this data leads to state-of-the-art results for text-to-video retrieval and action localization on instructional video datasets such as YouCook2 or CrossTask. Finally, we show that this embedding transfers well to other domains: fine-tuning on generic Youtube videos (MSR-VTT dataset) and movies (LSMDC dataset) outperforms models trained on these datasets alone. This work was presented at ICCV 2019 [16].

Figure 13. We learn a joint text-video embedding by watching millions of narrated video clips of people performing diverse visual tasks. The learned embedding transfers well to other instructional and non-instructional text-video datasets.

7.4.5. Are Large-Scale 3D Models Really Necessary for Accurate Visual Localization?

Participants: Akihiko Torii, Hajime Taira, Josef Sivic, Marc Pollefeys, Masatoshi Okutomi, Tomas Pajdla, Torsten Sattler.

Accurate visual localization is a key technology for autonomous navigation. 3D structure-based methods, as illustrated in Figure 14 , employ 3D models of the scene to estimate the full 6 degree-of-freedom (DOF) pose of a camera very accurately. However, constructing (and extending) large-scale 3D models is still a significant challenge. In contrast, 2D image retrieval-based methods only require a database of geo-tagged images, which is trivial to construct and to maintain. They are often considered inaccurate since they only approximate the positions of the cameras. Yet, the exact camera pose can theoretically be recovered when enough relevant database images are retrieved. In [8], we demonstrate experimentally that large-scale 3D models are not strictly necessary for accurate visual localization. We create reference poses for a large and challenging urban dataset. Using these poses, we show that combining image-based methods with local reconstructions results in a higher pose accuracy compared to state-of-the-art structure-based methods, albeight at higher run-time costs. We show that some of these run-time costs can be alleviated by exploiting known database image poses. Our results suggest that we might want to reconsider the need for large-scale 3D models in favor of more local models, but also that further research is necessary to accelerate the local reconstruction process.

7.4.6. End-to-End Learning of Visual Representations from Uncurated Instructional Videos

Participants: Antoine Miech, Jean-Baptiste Alayrac, Lucas Smaira, Ivan Laptev, Josef Sivic, Andrew Zisserman.

Annotating videos is cumbersome, expensive and not scalable. Yet, many strong video models still rely on manually annotated data. With the recent introduction of the HowTo100M dataset, narrated videos now offer the possibility of learning video representations without manual supervision. In this work we propose a new learning approach, MIL-NCE, capable of addressing misalignments inherent to narrated videos (see Figure 15). With this approach we are able to learn strong video representations from scratch, without the need for any manual annotation. We evaluate our representations on a wide range of four downstream tasks over eight datasets: action recognition (HMDB-51, UCF-101, Kinetics-700), text-to-video retrieval (YouCook2, MSR-VTT), action localization (YouTube-8M Segments, CrossTask) and action segmentation (COIN). Our method outperforms all published self-supervised approaches for these tasks as well as several fully supervised baselines. This preprint [32] is currently under review.

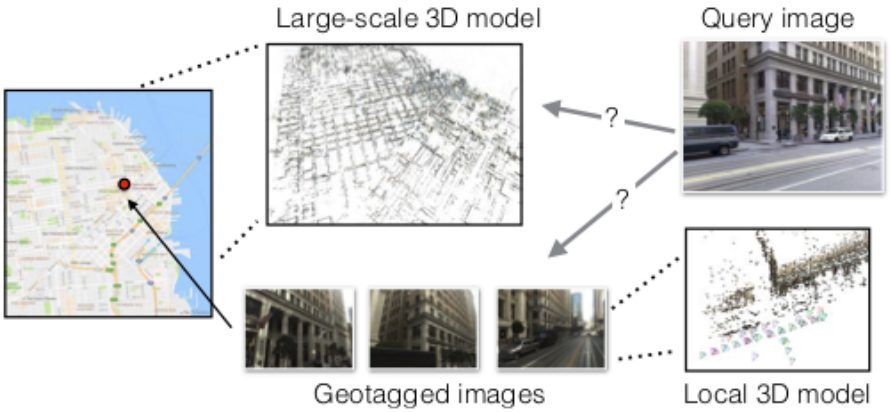


Figure 14. The state-of-the-art for large-scale visual localization. 2D image-based methods (bottom) use image retrieval and return the pose of the most relevant database image. 3D structure-based methods (top) use 2D-3D matches against a 3D model for camera pose estimation. Both approaches have been developed largely independently of each other and never compared properly before.

Figure 15. We describe an efficient approach to learn visual representations from highly misaligned and noisy narrations automatically extracted from instructional videos. Our video representations are learnt from scratch without relying on any manually annotated visual dataset yet outperform all self-supervised and many fully-supervised methods on several video recognition benchmarks.

7.4.7. Synthetic Humans for Action Recognition from Unseen Viewpoints

Participants: Gul Varol, Ivan Laptev, Cordelia Schmid, Andrew Zisserman.

In [35], the goal is to improve the performance of human action recognition for viewpoints unseen during training by using synthetic training data. Although synthetic data has been shown to be beneficial for tasks such as human pose estimation, its use for RGB human action recognition is relatively unexplored. We make use of the recent advances in monocular 3D human body reconstruction from real action sequences to automatically render synthetic training videos for the action labels. We make the following contributions: (i) we investigate the extent of variations and augmentations that are beneficial to improving performance at new viewpoints. We consider changes in body shape and clothing for individuals, as well as more action relevant augmentations such as non-uniform frame sampling, and interpolating between the motion of individuals performing the same action; (ii) We introduce a new dataset, SURREACT, that allows supervised training of spatio-temporal CNNs for action classification; (iii) We substantially improve the state-of-the-art action recognition performance on the NTU RGB+D and UESTC standard human action multi-view benchmarks; Finally, (iv) we extend the augmentation approach to in-the-wild videos from a subset of the Kinetics dataset to investigate the case when only one-shot training data is available, and demonstrate improvements in this case as well. Figure 16 presents an illustration of the approach.

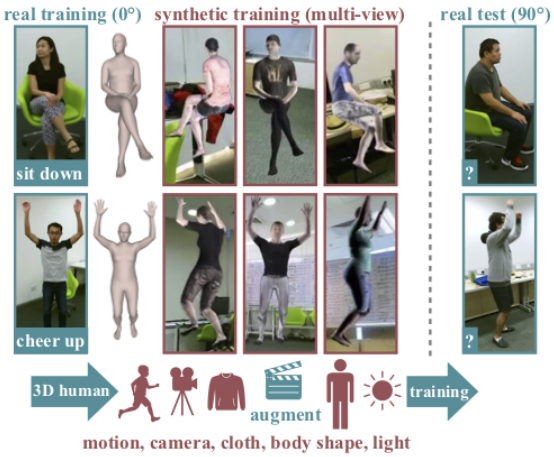


Figure 16. We estimate 3D shape from real videos and automatically render synthetic videos with action labels. We explore various augmentations for motions, viewpoints, and appearance. Training temporal CNNs with this data significantly improves the action recognition from unseen viewpoints.

7.5. Learning embodied representations and robotics

7.5.1. Roboticists and Reporters

Participants: Celine Pieters, Emmanuelle Danblon, Jean-Paul Laumond.

This paper reports on an experiment organized at the Cité des Sciences et de l'Industrie (CSI) of Paris in order to assess the importance of language in the representation and the integration of robots into the human culture. The experiment gathered specialized reporters and experts in robotics around a practical exercise of rhetoric. The objective of this work is to show that rhetoric is not a matter of communication, but a technique that allows to better understand the way roboticists understand their own discipline.

7.5.2. Robots

Participants: Jean-Paul Laumond, Denis Vidal.

What is a robot? How does it work? How is research progressing, what are the challenges and the economic and social questions posed by robotics in the twenty-first century? Today, as robots are becoming increasingly present in our professional, public and private lives, it is vital to understand their technological capabilities. We must more fully comprehend how they can help us and master their uses. Robots continue to fascinate us but our idea of them, stemming from literature and cinema, is often a purely imaginary one. This illustrated book accompanies the Robots exhibition at the Cité des sciences et de l'industrie. Figure 17 presents the front page of the exhibition.

Figure 17. Front page of the permanent exhibition at Cité des Sciences et de l'Industrie about Robotics.

7.5.3. Learning to Augment Synthetic Images for Sim2Real Policy Transfer

Participants: Alexander Pashevich, Robin Strudel, Igor Kalevatykh, Ivan Laptev, Cordelia Schmid.

Vision and learning have made significant progress that could improve robotics policies for complex tasks and environments. Learning deep neural networks for image understanding, however, requires large amounts of domain-specific visual data. While collecting such data from real robots is possible, such an approach limits the scalability as learning policies typically requires thousands of trials. In this paper [18], we attempt to learn manipulation policies in simulated environments. Simulators enable scalability and provide access to the underlying world state during training. Policies learned in simulators, however, do not transfer well to real scenes given the domain gap between real and synthetic data. We follow recent work on domain randomization and augment synthetic images with sequences of random transformations. Our main contribution is to optimize the augmentation strategy for sim2real transfer and to enable domain-independent policy learning. We design an efficient search for depth image augmentations using object localization as a proxy task. Given the resulting sequence of random transformations, we use it to augment synthetic depth images during policy learning. Our augmentation strategy is policy-independent and enables policy learning with no real images. We demonstrate our approach to significantly improve accuracy on three manipulation tasks evaluated on a real robot. Figure 18 presents an illustration of the approach.

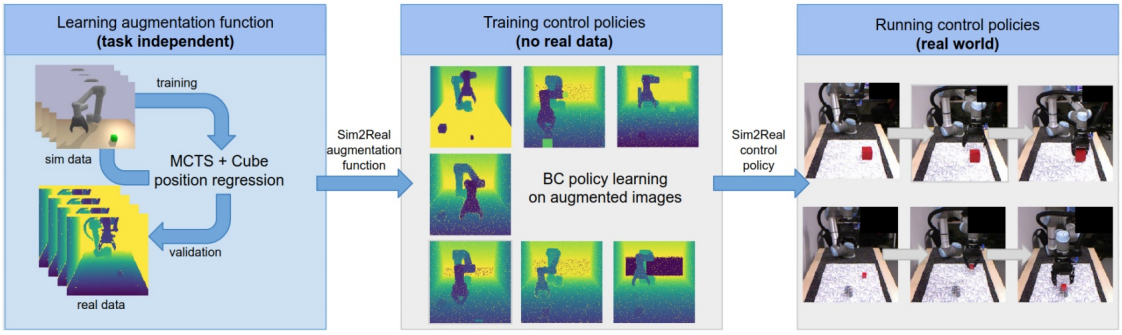


Figure 18. Overview of the method. Our contribution is the policy-independent learning of depth image augmentations (left). The resulting sequence of augmentations is applied to synthetic depth images while learning manipulation policies in a simulator (middle). The learned policies are directly applied to real robot scenes without finetuning on real images (right).

7.5.4. Learning to combine primitive skills: A step towards versatile robotic manipulation

Participants: Robin Strudel, Alexander Pashevich, Igor Kalevatykh, Ivan Laptev, Josef Sivic, Cordelia Schmid.

Manipulation tasks such as preparing a meal or assembling furniture remain highly challenging for robotics and vision. Traditional task and motion planning (TAMP) methods can solve complex tasks but require full state observability and are not adapted to dynamic scene changes. Recent learning methods can operate directly on visual inputs but typically require many demonstrations and/or task-specific reward engineering. In this paper [20], we aim to overcome previous limitations and propose a reinforcement learning (RL) approach to task planning that learns to combine primitive skills. First, compared to previous learning methods, our approach requires neither intermediate rewards nor complete task demonstrations during training. Second, we demonstrate the versatility of our vision-based task planning in challenging settings with temporary occlusions and dynamic scene changes. Third, we propose an efficient training of basic skills from few synthetic demonstrations by exploring recent CNN architectures and data augmentation. Notably, while all of our policies are learned on visual inputs in simulated environments, we demonstrate the successful transfer and high success rates when applying such policies to manipulation tasks on a real UR5 robotic arm. Figure 19 presents an illustration of the approach.

7.5.5. Monte-Carlo Tree Search for Efficient Visually Guided Rearrangement Planning

Participants: Sergey Zagoruyko, Yann Labbé, Igor Kalevatykh, Ivan Laptev, Justin Carpentier, Mathieu Aubry, Josef Sivic.

We address the problem of visually guided rearrangement planning with many movable objects, i.e., finding a sequence of actions to move a set of objects from an initial arrangement to a desired one, while relying on visual inputs coming from RGB camera. To do so, we introduce a complete pipeline relying on two key contributions. First, we introduce an efficient and scalable rearrangement planning method, based on a Monte-Carlo Tree Search exploration strategy. We demonstrate that because of its good trade-off between exploration and exploitation our method (i) scales well with the number of objects while (ii) finding solutions which require a smaller number of moves compared to the other state-of-the-art approaches. Note that on the contrary to many approaches, we do not require any buffer space to be available. Second, to precisely localize movable objects in the scene, we develop an integrated approach for robust multi-object workspace state estimation from a single uncalibrated RGB camera using a deep neural network trained only with synthetic data. We validate our multi-object visually guided manipulation pipeline with several experiments on a real

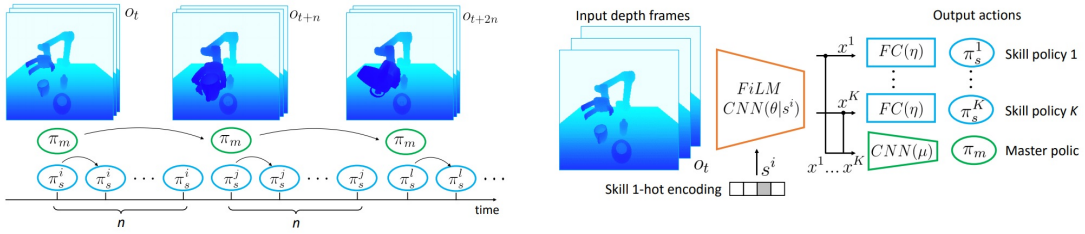


Figure 19. Illustration of our approach. (Left): Temporal hierarchy of master and skill policies. The master policy π_m is executed at a coarse interval of n time-steps to select among K skill policies $\pi_s^1 ... \pi_s^K$. Each skill policy generates control for a primitive action such as grasping or pouring. (Right): CNN architecture used for the skill and master policies.

UR-5 robotic arm by solving various rearrangement planning instances, requiring only 60 ms to compute the complete plan to rearrange 25 objects. In addition, we show that our system is insensitive to camera movements and can successfully recover from external perturbation. Figure 20 shows an example of the problems we consider. This work is under-review and an early pre-print is available [37].

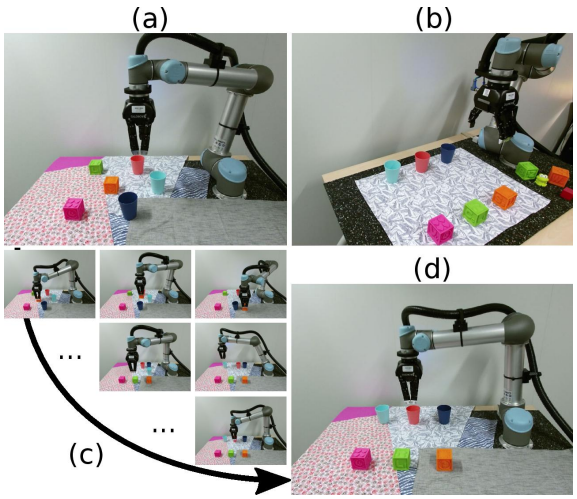


*Figure 20. **Visually guided rearrangement planning.** Given a source (a) and target (b) RGB images depicting a robot and multiple movable objects, our approach estimates the positions of objects in the scene without the need for explicit camera calibration and efficiently finds a sequence of robot actions (c) to re-arrange the scene into the target scene. Final object configuration after re-arrangement by the robot is shown in (d).*

7.5.6. Estimating the Center of Mass and the Angular Momentum Derivative for Legged Locomotion — A recursive approach

Participants: François Bailly, Justin Carpentier, Mehdi Benallegue, Bruno Watier, Philippe Soueres.

Estimating the center of mass position and the angular momentum derivative of legged systems is essential for both controlling legged robots and analyzing human motion. In this paper[4], a novel recursive approach to concurrently and accurately estimate these two quantities together is introduced. The proposed method employs kinetic and kinematic measurements from classic sensors available in robotics and biomechanics, to effectively exploits the accuracy of each measurement in the spectral domain. The soundness of the proposed approach is first validated on a simulated humanoid robot, where ground truth data is available, against an Extend Kalman Filter. The results demonstrate that the proposed method reduces the estimation error on the center of mass position with regard to kinematic estimation alone, whereas at the same time, it provides an accurate estimation of the derivative of angular momentum. Finally, the effectiveness of the proposed method is illustrated on real measurements, obtained from walking experiments with the HRP-2 humanoid robot. Figure 21 presents an illustration of the approach.

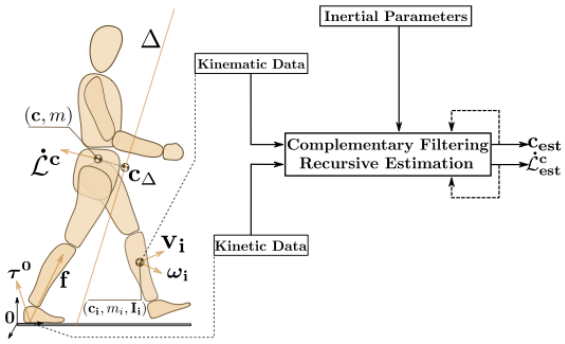


Figure 21. Illustration of the measurement apparatus. The several physical quantities involved in the estimation framework are displayed, as well as a simplified sketch of the estimation algorithm.

7.5.7. Dynamics Consensus between Centroidal and Whole-Body Models for Locomotion of Legged Robots

Participants: Rohan Budhiraja, Justin Carpentier, Nicolas Mansard.

It is nowadays well-established that locomotion can be written as a large and complex optimal control problem. Yet, current knowledge in numerical solver fails to directly solve it. A common approach is to cut the dimensionality by relying on reduced models (inverted pendulum, capture points, centroidal). However it is difficult both to account for whole-body constraints at the reduced level and also to define what is an acceptable trade-off at the whole-body level between tracking the reduced solution or searching for a new one. The main contribution of this paper [9] is to introduce a rigorous mathematical framework based on the Alternating Direction Method of Multipliers, to enforce the consensus between the centroidal state dynamics at reduced and whole-body level. We propose an exact splitting of the whole-body optimal control problem between the centroidal dynamics (under-actuation) and the manipulator dynamics (full actuation), corresponding to a rearrangement of the equations already stated in previous works. We then describe with details how alternating descent is a good solution to implement an effective locomotion solver. We validate this approach in simulation with walking experiments on the HRP-2 robot. Figure 22 presents a resulting motion of the proposed approach.

7.5.8. The Pinocchio C++ library – A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives

Participants: Justin Carpentier, Guilhem Saurel, Gabriele Buondonno, Joseph Mirabel, Florent Lamiraux, Olivier Stasse, Nicolas Mansard.

Figure 22. Walking sequence generated for HRP-2 robot using the proposed ADMM solver.

In this paper [10], we introduce Pinocchio, an open-source software framework that implements rigid body dynamics algorithms and their analytical derivatives. Pinocchio does not only include standard algorithms employed in robotics (e.g., forward and inverse dynamics) but provides additional features essential for the control, the planning and the simulation of robots. In this paper, we describe these features and detail the programming patterns and design which make Pinocchio efficient. We evaluate the performances against RBDL, another framework with broad dissemination inside the robotics community. We also demonstrate how the source code generation embedded in Pinocchio outperforms other approaches of state of the art. Figure 23 presents the logo of Pinocchio.

Figure 23. Logo of Pinocchio.

7.5.9. Crocoddyl: An Efficient and Versatile Framework for Multi-Contact Optimal Control

Participants: Carlos Mastalli, Rohan Budhiraja, Wolfgang Merkt, Guilhem Saurel, Bilal Hammoud, Maximilien Naveau, Justin Carpentier, Sethu Vijayakumar, Nicolas Mansard.

In this paper [31], we introduce Crocoddyl (Contact RObot COntrol by Differential DYnamic Library), an open-source framework tailored for efficient multi-contact optimal control. Crocoddyl efficiently computes the state trajectory and the control policy for a given predefined sequence of contacts. Its efficiency is due to the use of sparse analytical derivatives, exploitation of the problem structure, and data sharing. It employs differential geometry to properly describe the state of any geometrical system, e.g. floating-base systems. We have unified dynamics, costs, and constraints into a single concept-action-for greater efficiency and easy prototyping. Additionally, we propose a novel multiple-shooting method called Feasibility-prone Differential Dynamic Programming (FDDP). Our novel method shows a greater globalization strategy compared to classical Differential Dynamic Programming (DDP) algorithms, and it has similar numerical behavior to state-of-the-art multiple-shooting methods. However, our method does not increase the computational complexity typically encountered by adding extra variables to describe the gaps in the dynamics. Concretely, we propose two modifications to the classical DDP algorithm. First, the backward pass accepts infeasible state-control trajectories. Second, the rollout keeps the gaps open during the early "exploratory" iterations (as expected in multiple-shooting methods). We showcase the performance of our framework using different tasks. With our

method, we can compute highly-dynamic maneuvers for legged robots (e.g. jumping, front-flip) in the order of milliseconds. Figure 24 presents a resulting motion of the proposed approach.

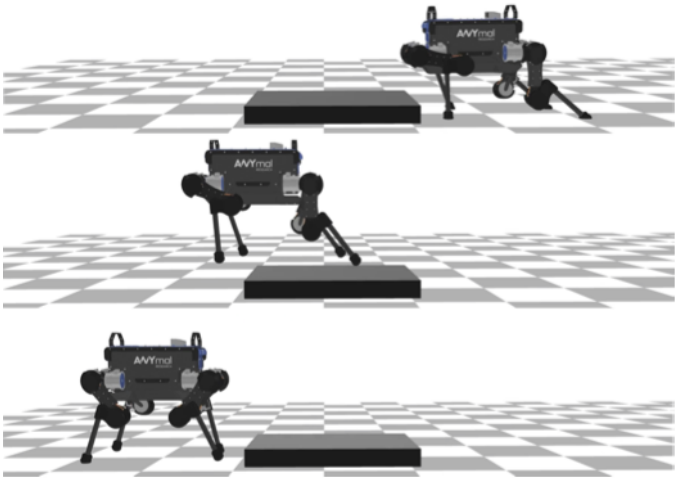


Figure 24. Crocoddyl: an efficient and versatile framework for multi-contact optimal control. Highly-dynamic maneuvers needed to traverse an obstacle with the ANYmal robot.