RESEARCH CENTER

FIELD
Perception, Cognition and Interaction

Activity Report 2018

Section New Results

Edition: 2019-03-07

CEDAR Project-Team

7. New Results

7.1. Interactive Data Exploration at Scale

Building upon our prior work in active learning-based interactive database exploration system, we improved this system in terms of efficiency and effectiveness. First, we formally defined the class of user interest queries to which our proposed Dual Space Model (DSM) can bring significant improvement in accuracy. Second, we generalized the DSM to arbitrary queries by forcing our system to fall back to the traditional active learning-based techniques if the requested query properties are not satisfied. Third, we launched a user study to collect real-world datasets and user interest patterns for comparison experiments. The evaluation results showed that our new system outperformed the start-of-the-art active learning techniques and data exploration systems. Fourth, to show the robustness of our system, we added some label noise into the experiments. It turned out that our system maintained a good performance and significantly outperformed traditional active learning-based system. These results have appeared in the prestigious PVLDB journal [10]. In addition, we have been working on integrating DSM with version space algorithms and designing more advanced methods to deal with label noise. In the near feature, a new software based on our proposed techniques will be put into use for interactive database exploration.

7.2. A learning-based approach to optimizing large-scale data analytics

As part of my PhD thesis of K. Zaouk, we have proposed two neural network architectures to support in-situ modeling of user objectives in large-scale data analytics. Although conceptually these architectures can work with any big data system, the modeling of user-objectives on analytics run was applied on Spark Streaming. In our problem settings where only few traces are run whenever a new workload is submitted to the cloud, we have proposed new optimizations to improve the accuracy and efficiency of the auto-encoder based architecture. Thus, we have developed a prototype that included these neural network architectures and optimizations. This prototype was then used to evaluate a benchmark of stream analytics that we developed and instrumented on top of two clusters that collect Spark Streaming workloads' traces.

We analyzed the performance of the proposed techniques and demonstrated their performance benefits over state of the art performance modeling techniques based on machine learning (such as Ottertune used in tuning traditional RDBMS). Our latest results show that we outperform Ottertune in robustness and in our problem settings. These results consolidated in a paper "Boosting Big Data Analytics with Deep Learning Models and Optimization Methods" submitted for publication, alongside with other scientific results in multi-objective optimization contributed by the co-author Fei Song. Work on this topic continues.

7.3. Event stream analysis

As enterprise information systems are collecting event streams from various sources, the ability of a system to automatically detect anomalous events and further provide human readable explanations is of paramount importance. In a position paper [19], we argue for the need of a new type of data stream analytics that can address anomaly detection and explanation discovery in a single, integrated system, which not only offers increased business intelligence, but also opens up opportunities for improved solutions. In particular, we propose a two-pass approach to building such a system, highlight the challenges, and offer initial directions for solutions.

7.4. Quotient summarization of RDF graphs

We have continued our work on efficiently computing informative summaries of large, heterogeneous RDF graphs.

First, we have noticed that type information, when available, can be used to group RDF nodes in interesting, pertinent equivalence classes. However, the integration of type in our quotient summarization framework (presented in ISWC 2017) is not straightforward, since an RDF node may have zero, one, or more than one types. In [15], we have identified a sufficient, flexible condition under which we are able to propose a form of quotient summarization based on types, even if a node has multiple types, and even if they are not organized in a tree-shape classification, but instead in a directed acyclic graph (DAG).

In parallel, we have finalized a comprehensive survey of RDF graph summarization techniques which appeared in the VLDB Journal [8]. We have also completely re-developed our RDF graph summarization platform, in order to ensure correctness, to factorize common elements across all the summarization methods, and to implement new, incremental summarization algorithms [21]. This work has attracted significant visibility through an invited keynote at the ESWC conference [25], and through an ISWC "Resource" publication where our summaries are integrated in a LOD visual exploration portal developed by the ILDA team of Inria [17].

7.5. Semantic integration of heterogeneous data

A large amount of data sources are publicly available in *heterogeneous formats* such as relational, RDF and JSON. These data sources can share information about common entities, which the users may want to query as a single dataset, possibly exploiting also a set of semantic constraints which serve as a common integration perspective. We proposed a new approach to query such *integration* of datasources in a *global RDF graph* using an *RDFS ontology* and user-specified entailment rules. Previous approaches to query answering in the presence of knowledge involve either the materialization of inferred data, or reformulation of the query; both approches have well-known drawbacks. We introduce a new way of query answering as a reduction to view-based answering in [11]. This approach avoids both materialization in the data and query reformulation.

We have also developed an RDF Schema *reformulation algorithm* taking into account the reasoning on the ontology. This algorithm reduces query answering on data in the presence of an ontology, to query evaluation (solely on the data). In particular, this reformulation algorithm can be used to speed up query answering in the integration system mentioned above.

7.6. Fact-checking: a content management perspective

Throughout the year, we have worked within the ANR ContentCheck project to analyze and systematize computational fact-checking as a discipline of computer science; we have analyzed and classified existing works in this area, proposed a generic architecture for computational fact-checking, and highlighted perspectives in a Web Conference (formerly known as WWW) article [18] and two tutorials, presented respectively at the Web conference [16] and the PVLDB conference [7]. This work has also been featured in an invited keynote at the BDA 2018 conference [24].

7.7. Novel fact-checking architectures and algorithms

Still part of our work in ContentCheck, we have worked to devise new algorithms and architectures for data journalism and journalistic fact checking.

First, we have considered the problem of making it easy to check the accuracy of a statistic claim, in the statistic database published by INSEE, the leading french statistic institute. In prior work, we had shown how the INSEE data can be converted into a collection of open data adherent to the best practices of the W3C (RDF graphs). Following up on that work, we have proposed a novel algorithm which allows to search these RDF datasets by means of user-friendly keyword queries. Our algorithm returns ranked answers at the granularity of the RDF dataset (corresponding to a spreadsheet in a statistic dataset published by INSEE) or, when possible, at the granularity of individual cells, or line/column in a spreadsheet that best matches the user query [13], [12].

Second, we have devised a new architecture for keyword search in a polystore systems, where users ask a set of keywords, and receive results showing how occurrences of these keywords across the set of data sources can be connected. This allows identifying possibly unforeseen connections across heterogeneous data sources. We have implemented this architecture in the ConnectionLens prototype, which we demonstrated in VLDB [9] and also informally at BDA [14].

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, Federico Ulliana.

Ontolology-mediated query answering (OMQA) is the issue of querying data while taking into account inferences enabled by ontological knowledge. This gives rise to *knowledge bases*, composed of a factbase (in database terms: an instance that contains incomplete data) and an ontology. Answers to queries are logically entailed from the knowledge base. Two families of formalisms for representing and reasoning with the ontological component have been considered in this context: *description logics* (DLs) and *existential rules* (aka Datalog+, or tuple-generating dependencies in database theory). Both frameworks correspond to fragments of first-order logic, which are incomparable in general but closely related in the context of OMQA: indeed, most DLs considered for OMQA, known as lighthweight DLs, are naturally translated into specific classes of existential rules. Importantly, the foundational work carried by the knowledge representation community led to the definition of several W3C standards for Semantic Web languages, namely the family of OWL 2 ontology languages, which can be used in combination with the RDF(S) Semantic Web language. This paradigm is also supported by commercial systems, such as Oracle.

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 (also known as the *chase* in databases), 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 can be divided into two steps: first, the query is *rewritten* using the rules into a first-order query (typically a union of conjunctive queries, but possibly a more compact form); then the rewritten query is evaluated against the factbase (again, as in a classical database system). Some classes of existential rules and lightweight description logics ensure the termination of the chase and/or query rewriting, but not all.

7.1.1. Revisiting the Chase

The interest for existential rules in the OMQA context brought again to light a fundamental tool in database theory, namely the chase. Several chase variants are known: they all yield logically equivalent results, but differ on how they handle redundancies possibly caused by the introduction of unknown individuals (often called nulls). Briefly, detecting redundancies leads to smaller saturated factbases, and prevents some infinite chase sequences, but it is costly. Given a chase variant, the (all-instances) 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 inclusion dependencies. We showed the decidability of the (all-instances) chase termination problem on linear rules for three main chase variants, namely *semi-oblivious*, *restricted* and *core* chase. The restricted chase is the most used variant of the chase, however it is notoriously tricky to study 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 these results, we introduced a novel approach based on so-called derivation trees and a single notion of forbidden pattern. Besides the theoretical interest of an unified approach and new proofs, we provided the first positive decidability results 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? [37] [27] (also to appear at ICDT 2019).

As part of Stathis Delivorias' PhD thesis, we considered the related problem of *boundedness*, which asks if a given set of existential rules is bounded, i.e., whether there is a predefined upper bound on the depth of the chase, independently from any factbase. This problem is already undecidable in the specific case of datalog rules (whose head has no existential variables). However, knowing that a set of rules is bounded for some chase variant does not help much in practice if the bound is unknown. Hence, we investigated the decidability of the k-boundedness problem, which asks whether a given set of rules is bounded by an integer k. We proved that k-boundedness is decidable for three main chase variants, namely the oblivious, semi-oblivious and restricted chase [23].

We investigated the combination of existential rules and answer set programming. The combination of the two formalisms requires to extend existential rules with nonmonotonic negation and to extend ASP with existential variables. To this aim, we introduced the syntax and semantics of existential non-monotonic rules using skolemization which join together the two frameworks. Building on our previous work published at ECAI and NMR, we presented syntactic conditions that ensure the termination of the chase for existential rules and discussed extension of these results in the nonmonotonic case [13].

7.1.2. Complexity of Ontology-Mediated Query Rewriting

Extending our previous work published at LICS, we carried out a systematic study on two fundamental problems in ontology-mediated query answering, in the context of the description logic OWL 2 QL. This dialect of the W3C standard ontology language OWL 2 is aimed towards efficient query answering on large data and ensures that every conjunctive ontology-mediated-query (OMQ) is rewritable into a first-order query. The first problem is the *succintness* of first-order rewritings of OMQs, which consists in understanding how difficult it is to built rewritings for queries in some OMQ class, and in particular to determine whether OMQs in the class have polynomial-size rewritings. The second problem is the *complexity* of OMQ answering. We classified OMQs according to the shape of their conjunctive queries (treewidth, the number of leaves) and the existential depth of their ontologies. For each of these classes, we determined the combined complexity of OMQ answering, and whether all OMQs in the class have polynomial-size first-order, positive existential and nonrecursive datalog rewritings. We obtained the succinctness results using hypergraph programs, a new computational model for Boolean functions, which makes it possible to connect the size of OMQ rewritings and circuit complexity [14].

7.1.3. Ontology-Based Data Access

In the above settings, data is supposed to be stored in a factbase built on the same vocabulary as the ontology. We now consider a more general setting, often called *Ontology-Based Data Access (OBDA)*, in which data is stored in one or several databases, which were generally built independently from the ontology. Hence, the ontological level acts as a mediating level, and a new component, namely *mappings*, allows to transfer the answers to queries over the data into facts expressed in the ontology vocabulary. Mappings may be triggered to actually materialize the factbase, but such materialization may be not possible nor desirable, in which case the factbase remains virtual.

OBDA is the core setting we consider in the Inria Project Lab iCODA on data journalism (https://project.inria.fr/icoda/). As part of Maxime Buron's PhD thesis (co-supervision shared between CEDAR and GraphIK teams), we investigate several frameworks and query answering techniques in the OBDA setting. We consider the Semantic Web language RDFS to express the (possibly virtual) factbase and the core ontology, RDF rules that include classical RDF entailment rules but possibly richer ontological knowledge, expressive mappings (namely global-local-as-view mappings, whereas most existing work in the area is restricted to global-as-view mappings), and queries which, in the spirit of RDF, can interrogate both the ontology and the data at the same time. In particular, we proposed a new way of answering queries by a reduction to database query rewriting with views [21]. Software development and experiments are under progress.

We also pursued our work on inconsistency-tolerant query answering, revisiting existing complexity results obtained for OMQA in the wider context of OBDA, i.e., considering mappings. We formalized the problem and performed a detailed analysis of the data complexity of inconsistency-tolerant OBDA for ontologies formulated in data-tractable description logics, considering different semantics, notions of repairs and classes

of GAV mappings. Our results imply that adding plain GAV mappings to the OMQA framework does not affect data complexity of inconsistency-tolerant query answering, but considering mappings with negated atoms leads to higher complexity [20].

Note that the latter work can also be seen as a contribution to maxi-consistent reasoning (see Section 7.2.2).

7.2. Reasoning with Inconsistency

Participants: Meghyn Bienvenu, Pierre Bisquert, Patrice Buche, Abdelraouf Hecham, Madalina Croitoru, Jérôme Fortin, Rallou Thomopoulos, Bruno Yun.

When reasoning about inconsistent logical KBs, one has to deploy reasoning mechanisms that do not follow the classical logical inference. This is due to the fact that, in classical logic, falsum implies everything. Alternative reasoning techniques are therefore needed in order to make sense of such KBs. In this section we present our results using two main classes of such techniques: defeasible reasoning and maxi-consistent reasoning.

7.2.1. Defeasible Reasoning

Defeasible reasoning is used to evaluate claims or statements in an inconsistent setting where the rules encoding the ontological knowledge may contradict each other. Unfortunately, there is no universally valid way to reason defeasibly. An inherent characteristic of defeasible reasoning is its systematic reliance on a set of intuitions and rules of thumb, which have been long debated between logicians. For example, could an information derived from a contested claim be used to contest another claim (i.e., ambiguity handling)? Could "chains" of reasoning for the same claim be combined to defend against challenging statements (i.e., team defeat)? Is circular reasoning allowed? Etc. We got interested in the task of a data engineer looking to select what existing tool to use to perform defeasible reasoning. To this end we proposed the first benchmark in the literature for first-order logic defeasible reasoning tools profiling and showed how to use the proposed benchmark in order to categorize existing tools based on their semantics (e.g. ambiguity handling), logical language (e.g. existential rules) and expressiveness (e.g. priorities) [25]. Furthermore, we proposed a new logical formalism called Statement Graphs (SGs) that captures the state-of-the-art defeasible reasoning features via a flexible labelling function [24].

7.2.2. Maxi-Consistent Reasoning

We now consider reasoning with inconsistent knowledge bases, when making the assumption that the ontological knowledge (here expressed by rules) is reliable, hence inconsistencies come from the data (or factbase), which may contradict ontological knowledge. We consider maximally consistent subsets of the factbase as the basis for inference (in short, "maxi-consistent" reasoning).

Repair semantics. One of the main challenges of reasoning with inconsistency is handling the inherent inconsistency that might occur amongst independently built data sources partially describing the same knowledge of interest. Inconsistency-tolerant semantics consider all maximally consistent subsets of a factbase, called repairs, that they manipulate using a modifier of these repairs (e.g. saturating them by the rules) and an inference strategy (e.g. answers have to be found in all repairs). However, using all repairs might be inappropriate for certain applications that would rather focus on particular data sources. For instance, when considering more reliable sources (i.e., sensor information, provenance data etc.) one could focus on repairs using mostly facts from such sources. When there is no given preference order on sources, we propose to use an intrinsic preference on facts based on their participation in inconsistencies, which generates a preference of repairs (i.e., those that contain less controversial facts are preferred). This led us to define a novel framework that takes into consideration the inconsistency on the facts and restricts the set of repairs to the "best" with respect to inconsistency values. We showed the significance and the practical interest of our approach using the real data collected in the framework of the Pack4Fresh project for reducing food wastes. During this project, we collected data using an online poll from a set of professionals of the food industry, including wholesalers, quality managers, floorwalkers and warehouse managers, about food packagings and their characteristics. The framework was able to rank the repairs efficiently and the results were then analysed and evaluated by experts from the packaging industry [35].

Argumentation. Argumentation is a reasoning under inconsistency technique, that allows to build arguments and attacks over an inconsistent data. The arguments represent the various inferences one can make. The attacks capture the inconsistency between the different pieces of knowledge. The set of arguments and the corresponding set of attacks is referred to as an argumentation framework (AF). AFs are visually represented using a directed graph where the nodes represent the arguments and the directed edges the attacks between the arguments. Classically, reasoning with argumentation systems consists of finding the maximal sets of arguments that (1) are not attacking each other and (2) defend themselves (as a group) from all incoming attacks. Such sets are called extensions.

Argumentation as a reasoning method over logic knowledge bases has the added value of providing better explanations to users than classical methods. However, one drawback of logic based argumentation frameworks is the large number of arguments generated. We provided a methodology for filtering semantically redundant arguments adapted for knowledge bases without rules or knowledge bases with rules. In the first case of knowledge bases without rules, we use the observation that free facts (i.e., facts that are not touched by any negative constraints) induce an exponential growth on the argumentation graph without any impact on its underlying structure. Therefore, we first generate the argumentation graph corresponding to the knowledge base without the free facts and then redo the whole graph including the arguments of the free facts in an efficient manner. In the second case, of the knowledge bases with rules, we introduce a new structure for the arguments and the attacks. In this new structure, we have significantly less arguments [28] (extended in [31]).

Furthermore, we provided a tool called Dagger that allows a knowledge engineer to (1) input a KB in a commonly used format and then (2) generate, (3) visualise or (4) export the argumentation graph [30]. Using the tool we were able to provide the first benchmark of logic based argumentation graphs in the litterature [32].

An alternative to the extension based semantics explained above are the ranking based semantics used mainly in the case where arguments are seen as abstract entities (and not necessarily logic derivation). There is a difference in the output format between these two approaches: when using a ranking based semantics, the output is a ranking on the arguments; in the case of extension based semantics, the output is a set of extensions. While the ranking and the scores (which are present in many ranking based semantics) allow to better assess the acceptability degree of each individual argument, the question "what are the different points of view of the argumentation framework?" stays unanswered when using a ranking based semantics. We have proposed a modular framework that is generic enough to be able to accommodate various application scenarios. In this case, one important property of the framework lies in its versatility and its capacity to yield different results according to various instantiations [33].

7.3. Decision Support Systems Applied to Agronomy

Participants: Pierre Bisquert, Patrice Buche, Abdelraouf Hecham, Madalina Croitoru, Jérôme Fortin, Rallou Thomopoulos, Bruno Yun.

High-level decision-making needs to take into account the often-conflicting interests of different stakeholders with the goal of finding solutions to provide trade-offs and build consensus towards the adoption of so-called win-win solutions. In order to enrich the deliberation process we have proposed several complementary approaches that combine various methods for an unified approach towards decision making. This has been applied in practical domains as explained below.

First, in [17] we presented a systematic method to assess possible options, based on the complementarity of argumentation modeling and system dynamics (SD) simulation, in conjunction with field experimentation. Taking advantage of the argument analysis, SD simulations are used to: 1) compare different cultural strategies available to farmers in current operating, market and regulatory conditions; 2) propose plausible what-if scenarios anticipating technological progress, and exploring the impact of adopting potential incentives and dissuasive regulatory measures.

Second, voting theory has been applied at the service of decision making. We employed Computational Social Choice (CSC) and Argumentation Framework (AF) as a combination to propose socially fair decisions which take into account both (1) the involved agents' preferences and (2) the justifications behind these preferences. Furthermore we implemented a software tool for decision-making which is composed of two main systems, i.e., the social choice system and the deliberation system [16]. This work was evaluated in practice [18]. Note that the use of argumentation in practice, when not considering fully formalised domains is very challenging. This specifically concerns decision support systems as shown in [34] where we focused on the following research question: "How to define an attack relation for argumentative decision making in socio-economic systems?" To address this question we proposed three kinds of attacks that could be defined in the context of a specific application (packaging selection) and studied how the non-computer-science experts evaluated, against a given set of decision tasks, each of these attacks.

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 domains of query optimization and machine learning.

7.1.1. Pattern Mining

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

- *Mining of novel types of patterns.* This includes mining of negative patterns [24], [14] and periodic patterns [18].
- *Data Mining for the masses.* In [11], we propose a communication model that bridges knowledge delivery between data miners and domain users in the field of library science. Our model proposes a five-steps process in order to achieve effective knowledge synthesis and delivery of insights to the domain users.
- *Efficient pattern mining.* In [10], we propose a method to sample itemsets efficiently on streaming data. This contribution tackles two limitations of the state of the art in pattern mining: (1) the so-called pattern explosion —the user is confronted to too many patterns—, and (2) the assumption of static data.
- *Data Mining for Data Science*. One of the most basic types of patterns is to know if the data makes one single group, i.e., is *unimodal*, or can be clustered into several groups. In [13], we propose a new statistical test of unimodality, that is both independent of the input distribution and computationally efficient.

7.1.2. Decision Support

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

- *Forecasting & prediction.* In [15], [12], we propose solutions to automate the task of capacity planning in the context of a large data network as the one available at Orange. The work in [19] offers a tool to predict the nutritional needs of sows in lactation.
- *Anomaly Detection.* The work in [20] tackles the problem of fraud detection under imbalanced data.

7.1.3. Others

- *Machine Learning.*[16] proposes a novel algorithm to weight the importance of classification errors when training a classifier. [8] proposes a classification algorithm optimized for highly imbalanced data.
- Query optimization. In [9] we propose a query-load-agnostic caching approach to speed-up provenance-aware queries in RDF data cubes.

7.2. Mining Periodic Patterns with a MDL Criterion

Participants: E. Galbrun, P. Cellier, N. Tatti, A. Termier, B. Crémilleux

The quantity of event logs available is increasing rapidly, be they produced by industrial processes, computing systems, or life tracking, for instance. It is thus important to design effective ways to uncover the information they contain. Because event logs often record repetitive phenomena, mining periodic patterns is especially relevant when considering such data. Indeed, capturing such regularities is instrumental in providing condensed representations of the event sequences. The work in [18] presents an approach for mining periodic patterns from event logs while relying on a Minimum Description Length (MDL) criterion to evaluate candidate patterns. Our goal is to extract a set of patterns that suitably characterises the periodic structure present in the data. We evaluate the interest of our approach on several real-world event log datasets.

7.3. NegPSpan: Efficient Extraction of Negative Sequential Patterns with Embedding Constraints

Participants: T. Guyet, R. Quinou

Mining frequent sequential patterns consists in extracting recurrent behaviors, modeled as patterns, in a big sequence dataset. Such patterns inform about which events are frequently observed in sequences, i.e., what does really happen. Sometimes, knowing that some specific event does not happen is more informative than extracting a lot of observed events. Negative sequential patterns (NSP) formulate recurrent behaviors by patterns containing both observed events and absent events. Few approaches have been proposed to mine such NSPs. In addition, the syntax and semantics of NSPs differ in the different methods which makes it difficult to compare them. [24] provides a unified framework for the formulation of the syntax and the semantics of NSPs. Then, it introduces a new algorithm, NegPSpan, that extracts NSPs using a PrefixSpan depth-first scheme and enabling maxgap constraints that other approaches do not take into account. The formal framework allows for highlighting the differences between the proposed approach w.r.t. to the methods from the literature, especially w.r.t. the state of the art approach eNSP. Intensive experiments on synthetic and real datasets show that NegPSpan can extract meaningful NSPs and that it can process bigger datasets than eNSP thanks to significantly lower memory requirements and better computation times.

7.4. NTGSP: Mining Negative Temporal Patterns

Participants: K. Tsesmeli, M. Boumghar, T. Guyet, R. Quiniou, L. Pierre

In [14] the authors study the problem of extracting frequent patterns containing positive events, negative events specifying the absence of events as well as temporal information on the delay between these events. [14] defines the semantics of such patterns and proposes the NTGSP method based on state-of-the-art approaches. The performance of the method is evaluated on commercial data provided by EDF (Électricité de France).

7.5. Accelerating Itemset Sampling using Satisfiability Constraints on FPGA

Participants: M. Gueguen, O. Sentieys, A. Termier

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 for analysis. Recent techniques, called *output space sampling* solve this problem by outputting only a sample of 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 [10], the authors propose a contribution to tackle both issues, by designing an FPGA accelerator for pattern mining with output space sampling. They show that their accelerator can outperform a state-of-the-art implementation on a server class CPU using a modest FPGA product.

7.6. Are your data data gathered? The Folding Test of Unimodality

Participants: A. Siffer, C. Largouët, A. Termier

Understanding data distributions is one of the most fundamental research topics in data analysis. The literature provides a great deal of powerful statistical learning algorithms to gain knowledge on the underlying distribution given multivariate observations. We are likely to find out a dependence between features, the appearance of clusters or the presence of outliers. Before such deep investigations, [13] proposes the folding test of unimodality. As a simple statistical description, it allows to detect whether data are gathered or not (unimodal or multimodal). To the best of our knowledge, this is the first multivariate and purely statistical unimodality test. It makes no distribution assumption and relies only on a straightforward p-value. Experiments on real world data show the relevance of the test and how to use it for the task of clustering.

7.7. Day-ahead Time Series Forecasting: Application to Capacity Planning

Participants: C. Leverger, V. Lemaire, S. Malinowski, T. Guyet, L. Rozé
In the context of capacity planning, forecasting the evolution of server usage enables companies to better manage their computational resources. The work in [12] addresses this problem by collecting key indicator time series. The article proposes a method to forecast the evolution of server usage one day-ahead. The method assumes that data is structured by a daily seasonality, but also that there is typical evolution of indicators within a day. Then, it uses the combination of a clustering algorithm and Markov Models to produce day-ahead forecasts. Our experiments on real datasets show that the data satisfies our assumption and that, in the case study, our method outperforms classical approaches (AR, Holt-Winters).

7.8. PerForecast: A Tool to Forecast the Evolution of Time Series for Capacity Planning.

Participants: C. Leverger, R. Marguerie, V. Lemaire, T. Guyet, S. Malinowski
The work published in [15] presents PerForecast, a tool for automatic capacity planning. The tool relies on univariate temporal data and automatically configured predictive models. The aim is to anticipate *capacity problems* in the infrastructure of Orange in order to ensure the delivery of services to customers. For example, PerForecast can predict the overhead of a server at the earliest possible stage, so that new machines can be ordered before the deterioration of the service in question. The purchase procedures being long and costly, the earlier they are done, the better the quality of service.

7.9. Tree-based Cost-Sensitive Methods for Fraud Detection in Imbalanced Data

Participants: G. Metzler, X. Badiche, B. Belkasmi, E. Fromont, A. Habrard, M. Sebban
Bank fraud detection is a difficult classification problem where the number of frauds is much smaller than the number of genuine transactions. The authors of [20] present cost sensitive tree-based learning strategies applied in this context of highly imbalanced data. The paper first proposes a cost sensitive splitting criterion for decision trees that takes into account the cost of each transaction. Then the criterion is extended with a decision rule for classification with tree ensembles. The authors then propose a new cost-sensitive loss for gradient boosting. Both methods have been shown to be particularly relevant in the context of imbalanced data. Experiments on a proprietary dataset of bank fraud detection in retail transactions show that the presented cost sensitive algorithms increase the retailer's benefits by 1,43% compared to non cost-sensitive ones and that the gradient boosting approach outperforms all its competitors.

7.10. An Algorithm to Optimize the F-measure by Proper Weighting of Classification Errors

Participants: K. Bascol, R. Emonet, E. Fromont, A. Habrard, G. Metzler, M. Sebban

[16] proposes an F-Measure optimization algorithm with theoretical guarantees that can be used with any error-weighting learning method. The algorithm, iteratively generates a set of costs from the training set so that the final classifier has an F-measure close to optimal. The optimality of the F-measure is expressed using a finer upper bound as presented in [31]. Furthermore, we show that the costs obtained at each iteration of our method can drastically reduce the search space and thus converge quickly to the optimal parameters. The efficiency of the method is shown both in terms of F-measurement but also in terms of speed of convergence on several unbalanced datasets.

7.11. Learning Maximum excluding Ellipsoids from Imbalanced Data with Theoretical Guarantees

Participants: G. Metzler, X. Badiche, B. Belkasmi, E. Fromont, A. Habrard, M. Sebban
[8] addresses the problem of learning from imbalanced data. The authors consider the scenario where the number of negative examples is much larger than the number of positive ones. This work proposes a theoretically-founded method, which learns a set of local ellipsoids centered at the minority class examples while excluding the negative examples of the majority class. This task is addressed from a Mahalanobis-like metric learning point of view, which allows deriving generalization guarantees on the learned metric using the uniform stability framework. The experimental evaluation on classic benchmarks and on a proprietary dataset in bank fraud detection shows the effectiveness of the approach, particularly when the imbalance is huge.

7.12. Answering Provenance-Aware Queries on RDF Data Cubes under Memory Budgets

Participants: L. Galárraga, K. Ahlstrøm, K. Hose, T. B. Pedersen
The steadily-growing popularity of semantic data on the Web and the support for aggregation queries in SPARQL 1.1 have propelled the interest in Online Analytical Processing (OLAP) and data cubes in RDF. Query processing in such settings is challenging because SPARQL OLAP queries usually contain many triple patterns with grouping and aggregation. Moreover, one important factor of query answering on Web data is its provenance, i.e., metadata about its origin. Some applications in data analytics and access control require to augment the data with provenance metadata and run queries that impose constraints on this provenance. This task is called provenance-aware query answering. The work in [9] investigates the benefit of caching some parts of an RDF cube augmented with provenance information when answering provenance-aware SPARQL queries. [9] proposes provenance-aware caching (PAC), a caching approach based on a provenance-aware partitioning of RDF graphs, and a benefit model for RDF cubes and SPARQL queries with aggregation. The results on real and synthetic data show that PAC outperforms significantly the LRU strategy (least recently used) and the Jena TDB native caching in terms of hit-rate and response time.

LINKS Project-Team

7. New Results

7.1. Querying Heterogeneous Linked Data

7.1.1. Data Integration

The PhD project of Lozano on relational to RDF data integration is progressing under the direction of Boneva, and Staworko. At **AMW** [9] they studied the *relational to RDF data exchange problem*. They focus in particular on a preliminary analysis of the consistency problem for relational to RDF data exchange with target ShEx schema.

7.1.2. Schema Validation

Shape Expression Language 2.0 (ShEx) is a language to describe the vocabulary and the structure of an RDF graph. It is base on the notion of shapes, a typing system supporting algebraic operations, recursive references to other shapes or Boolean combination.

In their **PODS** paper [7], Staworko studied the *containment problem* for ShEx (in cooperation with Wieczorek from Wrazlaw). Containement is a classical subject for schema-related issue in database theory. The authors proved that it is decidable for ShEx-schema, but with a untractable complexity (co-NEXP-hard). They also carefuly craft restriction of ShEx schema to design tractable-but-still-signifiant fragments.

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.

The topic of the PhD project of M. Sakho supervised by Niehren and Boneva is to generalize algorithms for querying streams to hyperstreams. These are collections of linked streams as naturally produced as intermediate results of complex events processing. Hyperstreams are incomplete descriptions of relational structures, so they can queried similarly to incomplete databases, for which the notion of a certain query answer is most appropriate.

In a paper published at **RP** [13], they studied certain query answering for hyperstreams with simple events. Such hyperstreams can be identified with compressed string patterns. They proved that the certain query answering for regular queries on compressed string patterns is PSPACE-complete, independently of whether the finite automata defining the regular queries are assumed deterministic or not, and independently of whether compression is permitted or not. They also showed that the problem is in PTIME when restricted to *linear* string patterns (possibly with compression) and to deterministic finite automata.

In a paper published at **LATA** [6], the studied certain query answering on hyperstreams of complex events. Such hyperstreams can be modeled by compressed tree pattern with context variables. They showed that certain query answering for regular queries on compressed tree pattern with context variables is EXP-complete, independently of whether the tree automata defining the regular queries are assumed deterministic or not, and independently of whether compression is permitted or not. They also showed that the problem is in PTIME when restricted to *linear* tree patterns (possibly with compression) and to deterministic tree automata.

7.2.2. Transformations

In his PhD project – belonging to the ANR Colis– Gallot with his supervisors Salvati and Lemay presented higher order tree transducers which extend macro tree transducers. Moreover they obtained nice properties such as the closure of the transducers under composition. Algorithms to compute such compositions are proposed. Those algorithms perform partial evaluation and are guided by semantic interpretations over finite domains.

Another virtue of higher-order transducers is that their *linear* syntactic restriction make them equivalent to logically defined MSO transductions. One of the composition algorithm proposed preserves the linearity. Furthermore, we have also showed that we can decrease the order of linear transducer (i.e. the complexity of the functions it handles) when this one is larger than 4.

These results are unpublished paper for now.

7.3. Foundations of AI

Various problems of databases and knowledge bases are closely related to foundational problems in artificial intelligence, since they are rooted in logic or graph theory.

7.3.1. Knowledge Compilation

Many problems in Artificial Intelligence boil down to the exploration of the solution set (called the models) of logical formulas. Such an exploration can be finding one model of the formula, counting the number of models or enumerating them all. However, even for simple quantifier-free formulas, those explorations are known be untractable (NP-hard).

Knowledge compilation encompasses methods that aim to change the representation of the set of models in order to get tractable algorithms for (some of) those tasks. A big computational cost is paid during the compilation time but then replying to queries become tractable on the new representation. More generally, the core of Knowledge compilation is the study of the trade-off between the size of the representation and the easiness of queries. This subject is of interest for both Artificial Intelligence and Database communities.

At **STACS** [15], Capelli, in cooperation with Mengel from CRIL (Lens), studied knowledge compilation techiques for quantified Boolean formulas. Deciding the existence of models for such formulas is known to climb arbitrarly high the polynomial time hierarchy. The authors provide an efficient compilation procedure for formulas having a *bounded tree-width* generalizing results from SAT solving.

7.3.2. Aggregation and Enumeration for Graphs

Aggregation and enumeration are not relevant for answer sets of database queries but equally for any kinds of sets, most typically defined by combinatoric problems on graphs.

In a paper published at **ICALP** [8], Paperman proposed (in cooperation with Amarilli from Telecom Paristech) to study the problem of finding so called *topological sort* satisfying constraints provided by regular expressions. Searching topological sort happens typically in situations where an order is *uncertain*. For instance, in relational database where users provides a partial preference order, or in concurrent and distributed programming where some tasks can be executed in an arbitrary order. A classical task in *preferential query answering* is to find a topological sort satisfying some global constrained. Typically, to find a total order satisfying all (or most) of the customers. The paper provides and proves sufficient conditions on the *shape of the constraints* to make the problem tractable (P-time) as well as sufficient condition to make the problem NP-hard. They also prove a complete dichotomy for an adapted and well chosen version of the constrained topological sort problem.

In an article in **JCSS** [2], Capelli (with Bergougnoux and Kanté from Bordeaux and Clérmont-Ferrand) propose an algorithm for counting the number of *transversal* in some *hypergraphs*. Here, a hypergraph is a collection of sets – called *hyperedges* over a *ground set* and a traversal is a subset intersecting all hyperedges. In full generality, counting the number of minimal traversals in a hypergraph is a hard problem: it is known to be $\#P$-complete. They proved that under the assumptions of β-acyclicity, it is possible to count all the minimal traversals can be done in polynomial times.

MAGNET Project-Team

7. New Results

7.1. On the Bernstein-Hoeffding Method

We consider extensions of Hoeffding's "exponential method" approach for obtaining upper estimates on the probability that a sum of independent and bounded random variables is significantly larger than its mean. We show that the exponential function in Hoeffding's approach can be replaced with any function which is non-negative, increasing and convex. As a result we generalize and improve upon Hoeffding's inequality. Our approach allows to obtain "missing factors" in Hoeffding's inequality. The later result is a rather weaker version of a theorem that is due to Michel Talagrand. Moreover, we characterize the class of functions with respect to which our method yields optimal concentration bounds. Finally, using ideas from the theory of Bernstein polynomials, we show that similar ideas apply under information on higher moments of the random variables ([4]).

7.2. IncGraph: Incremental graphlet counting for topology optimisation

Graphlets are small network patterns that can be counted in order to characterise the structure of a network (topology). As part of a topology optimisation process, one could use graphlet counts to iteratively modify a network and keep track of the graphlet counts, in order to achieve certain topological properties. Up until now, however, graphlets were not suited as a metric for performing topology optimisation; when millions of minor changes are made to the network structure it becomes computationally intractable to recalculate all the graphlet counts for each of the edge modifications. We propose IncGraph, a method for calculating the differences in graphlet counts with respect to the network in its previous state, which is much more efficient than calculating the graphlet occurrences from scratch at every edge modification made. In comparison to static counting approaches, our findings show IncGraph reduces the execution time by several orders of magnitude. The usefulness of this approach was demonstrated by developing a graphlet-based metric to optimise gene regulatory networks. IncGraph is able to quickly quantify the topological impact of small changes to a network, which opens novel research opportunities to study changes in topologies in evolving or online networks, or develop graphlet-based criteria for topology optimisation. IncGraph is freely available as an open-source R package on CRAN (incgraph). The development version is also available on GitHub (rcannood/incgraph) ([2]).

7.3. Graph sampling with applications to estimating the number of pattern embeddings and the parameters of a statistical relational model

Counting the number of times a pattern occurs in a database is a fundamental data mining problem. It is a subroutine in a diverse set of tasks ranging from pattern mining to supervised learning and probabilistic model learning. While a pattern and a database can take many forms, this paper focuses on the case where both the pattern and the database are graphs (networks). Unfortunately, in general, the problem of counting graph occurrences is #P-complete. In contrast to earlier work, which focused on exact counting for simple (i.e., very short) patterns, we present a sampling approach for estimating the statistics of larger graph pattern occurrences. We perform an empirical evaluation on synthetic and real-world data that validates the proposed algorithm, illustrates its practical behavior and provides insight into the trade-off between its accuracy of estimation and computational efficiency ([5]).

7.4. A machine learning based framework to identify and classify long terminal repeat retrotransposons

Transposable elements (TEs) are repetitive nucleotide sequences that make up a large portion of eukaryotic genomes. They can move and duplicate within a genome, increasing genome size and contributing to genetic diversity within and across species. Accurate identification and classification of TEs present in a genome is an important step towards understanding their effects on genes and their role in genome evolution. We introduce TE-LEARNER, a framework based on machine learning that automatically identifies TEs in a given genome and assigns a classification to them. We present an implementation of our framework towards LTR retrotransposons, a particular type of TEs characterized by having long terminal repeats (LTRs) at their boundaries. We evaluate the predictive performance of our framework on the well-annotated genomes of Drosophila melanogaster and Arabidopsis thaliana and we compare our results for three LTR retrotransposon superfamilies with the results of three widely used methods for TE identification or classification: REPEATMASKER, CENSOR and LTRDIGEST. In contrast to these methods, TE-LEARNER is the first to incorporate machine learning techniques, outperforming these methods in terms of predictive performance, while able to learn models and make predictions efficiently. Moreover, we show that our method was able to identify TEs that none of the above method could find, and we investigated TE-LEARNER's predictions which did not correspond to an official annotation. It turns out that many of these predictions are in fact strongly homologous to a known TE ([6]).

7.5. A Distributed Frank-Wolfe Framework for Learning Low-Rank Matrices with the Trace Norm

We consider the problem of learning a high-dimensional but low-rank matrix from a large-scale dataset distributed over several machines, where low-rankness is enforced by a convex trace norm constraint. We propose DFW-Trace, a distributed Frank-Wolfe algorithm which leverages the low-rank structure of its updates to achieve efficiency in time, memory and communication usage. The step at the heart of DFW-Trace is solved approximately using a distributed version of the power method. We provide a theoretical analysis of the convergence of DFW-Trace, showing that we can ensure sublinear convergence in expectation to an optimal solution with few power iterations per epoch. We implement DFW-Trace in the Apache Spark distributed programming framework and validate the usefulness of our approach on synthetic and real data, including the ImageNet dataset with high-dimensional features extracted from a deep neural network ([7]).

7.6. Personalized and Private Peer-to-Peer Machine Learning

The rise of connected personal devices together with privacy concerns call for machine learning algorithms capable of leveraging the data of a large number of agents to learn personalized models under strong privacy requirements. In this paper, we introduce an efficient algorithm to address the above problem in a fully decentralized (peer-to-peer) and asynchronous fashion, with provable convergence rate. We show how to make the algorithm differentially private to protect against the disclosure of information about the personal datasets, and formally analyze the trade-off between utility and privacy. Our experiments show that our approach dramatically outperforms previous work in the non-private case, and that under privacy constraints, we can significantly improve over models learned in isolation ([9]).

7.7. Hiding in the Crowd: A Massively Distributed Algorithm for Private Averaging with Malicious Adversaries

The amount of personal data collected in our everyday interactions with connected devices offers great opportunities for innovative services fueled by machine learning, as well as raises serious concerns for the privacy of individuals. In this paper, we propose a massively distributed protocol for a large set of users to privately compute averages over their joint data, which can then be used to learn predictive models. Our protocol can find a solution of arbitrary accuracy, does not rely on a third party and preserves the privacy of

users throughout the execution in both the honest-but-curious and malicious adversary models. Specifically, we prove that the information observed by the adversary (the set of malicious users) does not significantly reduce the uncertainty in its prediction of private values compared to its prior belief. The level of privacy protection depends on a quantity related to the Laplacian matrix of the network graph and generally improves with the size of the graph. Furthermore, we design a verification procedure which offers protection against malicious users joining the service with the goal of manipulating the outcome of the algorithm ([15]).

7.8. A Probabilistic Model for Joint Learning of Word Embeddings from Texts and Images

Several recent studies have shown the benefits of combining language and perception to infer word embeddings. These multimodal approaches either simply combine pre-trained textual and visual representations (e.g. features extracted from convolutional neural networks), or use the latter to bias the learning of textual word embeddings. In this work, we propose a novel probabilistic model to formalize how linguistic and perceptual inputs can work in concert to explain the observed word-context pairs in a text corpus. Our approach learns textual and visual representations jointly: latent visual factors couple together a skip-gram model for co-occurrence in linguistic data and a generative latent variable model for visual data. Extensive experimental studies validate the proposed model. Concretely, on the tasks of assessing pairwise word similarity and image/caption retrieval, our approach attains equally competitive or stronger results when compared to other state-of-the-art multimodal models ([8]).

7.9. A Framework for Understanding the Role of Morphology in Universal Dependency Parsing

We present a simple framework for characterizing morphological complexity and how it encodes syntactic information. In particular, we propose a new measure of morpho-syntactic complexity in terms of governor-dependent preferential attachment that explains parsing performance. Through experiments on dependency parsing with data from Universal Dependencies (UD), we show that representations derived from morphological attributes deliver important parsing performance improvements over standard word form embeddings when trained on the same datasets. We also show that the new morpho-syntactic complexity measure is predictive of the gains provided by using morphological attributes over plain forms on parsing scores, making it a tool to distinguish languages using morphology as a syntactic marker from others ([11]).

7.10. Online Reciprocal Recommendation with Theoretical Performance Guarantees

A reciprocal recommendation problem is one where the goal of learning is not just to predict a user's preference towards a passive item (e.g., a book), but to recommend the targeted user on one side another user from the other side such that a mutual interest between the two exists. The problem thus is sharply different from the more traditional items-to-users recommendation, since a good match requires meeting the preferences at both sides. We initiate a rigorous theoretical investigation of the reciprocal recommendation task in a specific framework of sequential learning. We point out general limitations, formulate reasonable assumptions enabling effective learning and, under these assumptions, we design and analyze a computationally efficient algorithm that uncovers mutual likes at a pace comparable to that achieved by a clairvoyant algorithm knowing all user preferences in advance. Finally, we validate our algorithm against synthetic and real-world datasets, showing improved empirical performance over simple baselines ([13]).

7.11. On Similarity Prediction and Pairwise Clustering

We consider the problem of clustering a finite set of items from pairwise similarity information. Unlike what is done in the literature on this subject, we do so in a passive learning setting, and with no specific constraints on the cluster shapes other than their size. We investigate the problem in different settings: i. an

online setting, where we provide a tight characterization of the prediction complexity in the mistake bound model, and ii. a standard stochastic batch setting, where we give tight upper and lower bounds on the achievable generalization error. Prediction performance is measured both in terms of the ability to recover the similarity function encoding the hidden clustering and in terms of how well we classify each item within the set. The proposed algorithms are time efficient ([12]).

7.12. A Probabilistic Theory of Supervised Similarity Learning for Pointwise ROC Curve Optimization

The performance of many machine learning techniques depends on the choice of an appropriate similarity or distance measure on the input space. Similarity learning (or metric learning) aims at building such a measure from training data so that observations with the same (resp. different) label are as close (resp. far) as possible. In this paper, similarity learning is investigated from the perspective of pairwise bipartite ranking, where the goal is to rank the elements of a database by decreasing order of the probability that they share the same label with some query data point, based on the similarity scores. A natural performance criterion in this setting is pointwise ROC optimization: maximize the true positive rate under a fixed false positive rate. We study this novel perspective on similarity learning through a rigorous probabilistic framework. The empirical version of the problem gives rise to a constrained optimization formulation involving U-statistics, for which we derive universal learning rates as well as faster rates under a noise assumption on the data distribution. We also address the large-scale setting by analyzing the effect of sampling-based approximations. Our theoretical results are supported by illustrative numerical experiments ([14]).

7.13. 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 this paper, 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 ([16]).

7.14. Nonstochastic Bandits with Composite Anonymous Feedback

We investigate a nonstochastic bandit setting in which the loss of an action is not immediately charged to the player, but rather spread over at most d consecutive steps in an adversarial way. This implies that the instantaneous loss observed by the player at the end of each round is a sum of as many as d loss components of previously played actions. Hence, unlike the standard bandit setting with delayed feedback, here the player cannot observe the individual delayed losses, but only their sum. Our main contribution is a general reduction transforming a standard bandit algorithm into one that can operate in this harder setting. We also show how the regret of the transformed algorithm can be bounded in terms of the regret of the original algorithm. Our reduction cannot be improved in general: we prove a lower bound on the regret of any bandit algorithm in this setting that matches (up to log factors) the upper bound obtained via our reduction. Finally, we show how our reduction can be extended to more complex bandit settings, such as combinatorial linear bandits and online bandit convex optimization ([10]).

MOEX Project-Team

4. New Results

4.1. 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 [7], 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 [2] and revealed that, by playing simple interaction games, agents can effectively repair random networks of ontologies or even create new alignments.

4.1.1. Strengthening modality for cultural alignment repair

Participants: Jérôme Euzenat [Correspondent], Iris Lohja.

Our previous work on cultural alignment repair achieved 100% precision for all adaptation operators, i.e., all the correspondences in the alignments were correct, but were still missing some correspondences, and did not achieve 100% recall. We had conjectured that this was due to a phenomenon called reverse shadowing [2], avoiding to find specific correspondences.

This year we introduced a new adaptation modality, strengthening, to test this hypothesis. The strengthening modality replaces a successful correspondence by one of its subsumed correspondences covering the current instance. This modality is different from those developed so far, because it leads agents to adapt their alignment when the game played has been a success (previously, it was always when a failure occurred). We defined three alternative definitions of this modality depending on if the agent chooses the most general, most specific or a random such correspondence.

The strengthening modality has been implemented in our *Lazy lavender* software. We experimentally showed that it was not interferring with the other modalities as soon as the *add* operator was used. This means that all properties of the previous adaptation operators are preserved. Moreover, as expected, recall was greatly increased, to the point that some operators achieve 99% F-measure. However, the agents still do not reach 100% recall.

4.1.2. Experiment reproducibility through container technology

Participants: Jérôme Euzenat [Correspondent], Bilal Lahmami.

Performing experiments and reporting them requires care in order for others to be able to repeat them.

We experimented with container technology in order to embed our experiments and offer to others to run them easily. To that extent, we developed scripts associated to the *Lazy lavender* software to specify, run, and analyse experiments. In particular, these scripts are able to generate a Docker container specification that can perform experiments in the same conditions or with updated software. The documentation of the experiments on our Wiki platform (https://gforge.inria.fr/plugins/mediawiki/wiki/lazylav/index.php/Lazy_Lavender) is also eased by this process.

4.2. Link keys

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

- Extracting link keys;
- Reasoning with link keys.

4.2.1. Link key extraction with relational concept analysis

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

We have further investigated link key extraction using relational concept analysis and the associated prototype implementation [8]. In particular, we showed that that link keys extracted by formal concept analysis are equivalent to an extension of those which were extracted by our former algorithm [1]

4.2.2. Link key extraction under ontological constraints

Participants: Jérôme David [Correspondent], Jérôme Euzenat, Khadija Jradeh.

We investigated the use of link keys taking advantage of ontologies. This can be carried out in two different directions: exploiting the ontologies under which data sets are published, and extracting link keys using ontology constructors for combining attribute and class names.

Following the first approach, we extended our existing algorithms to extract link keys involving inverse ($^{-1}$), union ($\sqcup$), intersection ($\sqcap$) and paths ($\circ$) of properties. This helps providing link keys when it is not possible otherwise (without inverse, there is no possible correspondence if one data set is using parents and the other is using children). We showed how the paths could be normalised to reduce the search space. Extracting link keys under these conditions required to introduce better indexing techniques to avoid unnecessary link key generation and even looping.

We implemented this method and evaluated it by running experiments on two real data sets, this resulted in finding the correct link keys that were not found without them.

4.2.3. 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. We further worked on the tableau method designed for the $\mathcal{ALC}$ description logic to support reasoning with link keys.

4.3. Semantic web queries

4.3.1. Evaluation of query transformations without data

Participants: Jérôme David, Jérôme Euzenat [Correspondent].

Query transformations are ubiquitous in semantic web query processing. For any situation in which transformations are not proved correct by construction, the quality of these transformations has to be evaluated. Usual evaluation measures are either overly syntactic and not very informative —the result being: correct or incorrect— or dependent from the evaluation sources. Moreover, both approaches do not necessarily yield the same result. We proposed to ground the evaluation on query containment [4]. This allows for a data-independent evaluation that is more informative than the usual syntactic evaluation. In addition, such evaluation modalities may take into account ontologies, alignments or different query languages as soon as they are relevant to query evaluation [6].

ORPAILLEUR Project-Team

7. New Results

7.1. Mining of Complex Data

Participants: Nacira Abbas, Guilherme Alves Da Silva, Alexandre Blansché, Lydia Boudjeloud-Assala, Quentin Brabant, Brieuc Conan-Guez, Miguel Couceiro, Adrien Coulet, Alain Gély, Laurine Huber, Nyoman Juniarta, Florence Le Ber, Joël Legrand, Pierre Monnin, Tatiana Makhalova, Amedeo Napoli, Abdelkader Ouali, François Pirot, Frédéric Pennerath, Justine Reynaud, Chedy Raïssi, Sébastien Da Silva, Yannick Toussaint.

Keywords: formal concept analysis, relational concept analysis, pattern structures, pattern mining, association rule, redescription mining, graph mining, sequence mining, biclustering, hybrid mining, meta-mining

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 data. In particular, FCA, which usually applies to binary data-tables, can be adapted to work on more complex data. In this way, we have contributed to two main extensions of FCA, namely Pattern Structures and Relational Concept Analysis. Pattern Structures (PS [73]) allow building a concept lattice from complex data, e.g. numbers, sequences, trees and graphs. Relational Concept Analysis (RCA) is able to analyze objects described both by binary and relational attributes [84] 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 [8], the mining of RDF data and the and visualization of the discovered patterns [1], and redescription mining (detailed later). Moreover, we have also investigated the use of the MDL principle ("Minimum Description Length") for the selection of interesting and diverse patterns [37], [39].

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 [32], [33]. The mining process is based on two approaches in the framework of Formal Concept Analysis (FCA). We focused on different types of sequences and more precisely on subsequences without any constraint and frequent contiguous subsequences. In parallel, we introduced a similarity measure allowing us to build a hierarchical classification which is used for interpretation and characterization of the trajectories. In addition, for completing the research work on the characterization of trajectories, we also studied how biclustering may be applied to trajectory recommendation [31], [52].

7.1.2. Redescription Mining

Among the mining methods developed in the team is redescription mining. Redescription mining aims to find distinct common characterizations of the same objects and, vice versa, to identify sets of objects that admit multiple shared descriptions [82]. It is motivated by the idea that in scientific investigations data oftentimes have different nature. For instance, 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 that exist between these different aspects.

A practical example in biology consists in finding geographical areas that admit two characterizations, one in terms of their climatic profile and one in terms of the occupying species. Discovering such redescriptions can contribute to better our understanding of the influence of climate over species distribution. Besides biology, applications of redescription mining can be envisaged in medicine or sociology, among other fields.

This year, we used redescription mining for analyzing and mining RDF 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 [49], [44].

7.1.3. Text Mining

In the context of the PractikPharma ANR Project, we study how cross-corpus training may guide the task of relationship extraction from texts, and especially, how large annotated corpora developed for alternative tasks may improve the performance of biomedical tasks, for which only a few annotated resources are available [34].

Transfer learning proposes to enhance machine learning performance on a problem, by reusing labeled data originally designed for a related problem. This is particularly relevant to the applications of deep learning in Natural Language Processing, because those usually require large annotated corpora that may not exist for the targeted domain, but exist for side domains. In a recent work, we experimented the extraction of relationships from biomedical texts with two deep learning models. The first model combines locally extracted features using a Multi Channel Convolutional Neural Network (MCCNN) model, while the second model exploits the syntactic structure of sentences using a Tree-LSTM (Long Short-Term Memory) architecture. The experiments show that the Tree-LSTM model benefits from a cross-corpus learning strategy, i.e. performances are improved when training data are enriched with off-target corpora, whereas it is not the case with MCCNN.

Indeed our approach leads to state of the art performances in four biomedical tasks for which only a few annotated resources are available (less than 400 manually annotated sentences) and even surpass state of the art performances in two of these four tasks. We particularly investigated how the syntactic structure of a sentence, which is domain independent, participates in the increase of performance when adding additional training data. This may have a particular impact in specialized domains in which training resources are scarce, because it means that these resources may be efficiently enriched with data from other domains for which large annotated corpora exist.

7.1.4. Mining subgroups as a single-player game

Discovering patterns that strongly distinguish one class label from another is a challenging data-mining task. The unsupervised discovery of such patterns would enable the construction of intelligible classifiers and to elicit interesting hypotheses from the data. Subgroup Discovery (SD) is one framework that formally defines this pattern mining task. However, SD still faces two major issues: (i) how to define appropriate quality measures to characterize the uniqueness of a pattern; (ii) how to select an accurate heuristic search technique when exhaustive enumeration of the pattern space is unfeasible. The first issue has been tackled by the Exceptional Model Mining (EMM) framework. This general framework aims to find patterns that cover tuples that locally induce a model that substantially differs from the model of the whole dataset. The second issue has been studied in SD and EMM mainly with the use of beam-search strategies and genetic algorithms for discovering a pattern set that is non-redundant, diverse and of high quality. Consequently,

In our current work [9], we proposed to formally define pattern mining as a single-player game, as in a puzzle, and to solve it with a Monte Carlo Tree Search (MCTS), a technique mainly used for artificial intelligence and planning problems. The exploitation/exploration trade-off and the power of random search of MCTS lead to an any-time mining approach, in which a solution is always available, and which tends towards an exhaustive search if given enough time and memory. Given a reasonable time and memory budget, MCTS quickly drives the search towards a diverse pattern set of high quality. MCTS does not need any knowledge of the pattern quality measure, and we show to what extent it is agnostic to the pattern language.

7.1.5. Consensus and Aggregation 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 [76], [78]. 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) [27], that is inspired by the frameworks of unsupervised random forests (URF) [85] and of extremely randomized trees (ET) [75]. 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. We are now working on the adaptation of UET for heterogeneous data (both numerical and symbolic), possibly, with missing values.

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 and 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]. For further results in the qualitative approach to decision making see, e.g., [10] [3]; and see also [24] for a survey chapter on new perspectives in ordinal evaluation.

7.2. Knowledge Discovery in Healthcare and Life Sciences

Participants: Miguel Couceiro, Adrien Coulet, Nicolas Jay, Joël Legrand, Pierre Monnin, Amedeo Napoli, Abdelkader Ouali, Chedy Raïssi, Malika Smaïl-Tabbone, 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 (GID), on the basis of their HPO annotations and of GO annotations of genes known for being responsible for these diseases [43]. We used clustering algorithms based on semantic similarities and enabling to compare sets of annotations. This experiment illustrates the fact that considering several ontologies provides better results, while selecting the best set of ontologies to combine is dependent on the dataset and on the classification task.

7.2.2. Validation of Pharmacogenomic Knowledge

State of the art knowledge in pharmacogenomics is heterogeneous w.r.t. validation. A part is 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 lead this validation, we define a minimal data schema for pharmacogenomic knowledge units (PGxO ontology), which is instantiated with data of various provenance (e.g. biomedical databases, literature and EHR). Such an instantiation produces a unique knowledge graph named PGxLOD (https://pgxlod.loria.fr/). We defined and applied a first set of reconciliation rules that compare and align whenever possible knowledge elements of various provenance. A journal article on the construction of PGxLOD and its use in knowledge comparison is currently under evaluation. We are continuing this effort by studying methods which enable a more flexible knowledge comparison.

In addition, we took part to the Biohackathon 2018 Paris (https://bh2018paris.info/) during which we worked on two tasks. Firstly we updated PGxLOD for improving its quality, completeness and interconnection with other resources. Secondly we mined PGxLOD and searched for explanations of the molecular mechanism of adverse drug responses. PGxLOD is under evaluation for being registered as a resource of the IBF (*French Institute for Bioinformatics*) and Elixir (an international organization that supports and structures bioinformatics efforts in Europe).

7.2.3. Mining Electronic Health Records

In the context of the Snowball Inria Associate Team, we developed an approach based on pattern structures to identify frequently associated ADRs (Adverse Drug Reactions) from patient data either in the form of EHR or ADR spontaneous reports. Pattern structures provide an expressive representation of ADR, taking into account the multiplicity of drugs and phenotypes involved in such reactions. Additionally, pattern structures allow considering diverse biomedical ontologies used to represent or annotate patient data, enabling a "semantic" comparison of ADRs. Up to now, this is one of the first research attempts considering such representations to mine rules between frequently associated ADRs. We illustrated the generality of the approach on two patient datasets, each of them linked to distinct biomedical ontologies. The first dataset corresponds to anonymized EHRs, extracted from "STRIDE", the EHR data warehouse of Stanford Hospital and Clinics. The second dataset is extracted from the U.S. FDA (for Food & Drug Administration) "Adverse Event Reporting System" (FAERS). Several significant association rules have been extracted, analyzed and may be used as a basis for a recommendation system.

In collaboration with Stanford University and the CHRU Nancy, we studied the use of Electronic Health Records to predict at first prescription the need for a patient to be prescribed with a reduced drug dose [4]. 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 predicting and preventing.

7.3. Knowledge Engineering and Web of Data

Participants: Nicolas Jay, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer, Justine Reynaud, Yannick Toussaint.

Keywords: knowledge engineering, web of data, definition mining, classification-based reasoning, case-based reasoning, belief revision, semantic web

7.3.1. Current Trends in Case-Based Reasoning

Case-based reasoning (CBR) aims at solving a new problem, called the target problem, by exploiting past experiences (i.e. source cases) as well as other knowledge sources: domain knowledge, similarity knowledge and adaptation knowledge.

Two research works were carried out about how exploiting at the best the source cases. A first work addresses the exploitation of negative cases for adaptation knowledge discovery. Usually CBR exploits positive source cases consisting of a source problem and its solution that is known to be correct for the problem. However, negative cases, i.e. problem-solution pairs where the solution is an incorrect answer to the problem, which can be acquired when CBR process fails, are useful, especially for adaptation knowledge discovery. In [29], we propose an adaptation knowledge discovery approach exploiting both type of cases (positive and negatives cases), using closed itemsets built on variations between cases. Experiments show that exploiting negative cases in addition to positive ones improves the quality of the adaptation knowledge being extracted and, so, improves the results of the CBR system.

A second work addresses the issue of the selection of source cases used to solve a target problem. Three approaches have been studied to better exploit source cases: (1) approximation, which considers the use of one source case (the most similar to the target problem) to solve the target problem, (2) interpolation, which considers the use of two source cases (such as the target problem is between these two similar source problems), and (3) extrapolation, which considers the use of three source cases, linked to the target problem by an analogical proportion, where the analogical proportion handles both similarity and dissimilarity between cases. Experiments show that interpolation and extrapolation techniques are of interest for reusing cases, either in an independent or in a combined way [36], [47].

Using analogical proportion has also been used to find relevant pathology-gene pairs [28]. This first study to infer pathology-gene relation is based on the following hypothesis: if a target pathology is in analogy with three other pathologies for which associated genes are known, then it is plausible that the gene to be associated with the target pathology is in analogy with the genes associated to the three pathologies involved in the analogical proportion.

Another use of analogical proportion is its application to machine translation and is based on a similar principle: if four sentences form an analogical proportion in a language, then it is plausible that their translations in another language also form an analogical proportion. This was the idea developed by Yves Lepage (Waseda University), a few years ago. Now, a starting work on case-based machine translation aims at developing these ideas by incorporation other knowledge sources to the CBR system than the cases (domain knowledge, retrieval knowledge and adaptation knowledge) [35].

Another work on CBR is its application to medical coding. Cancer registries are important tools in the fight against cancer. At the heart of these registries is the data collection and coding process. Ruled by complex international standards and numerous best practices, operators are easily overwhelmed. In [54], [55], a system is presented to assist operators in the interpretation of best medical coding practices.

There has been another work on CBR related to an application in agronomy developed some time ago that has been synthesized in [60].

7.3.2. Exploring and Classifying the Web of Data

A part of the research work in Knowledge Engineering is oriented towards knowledge discovery in the web of data, following 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 of the data viewed as their their potential to provide concept definitions in terms of necessary and sufficient conditions [69]. We have proposed a novel technique based on Formal Concept Analysis which classifies subsets of RDF data into a concept lattice [83]. 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. Moreover, this is a way of reconciling syntax and semantics in the LOD cloud. Experiments on the DBpedia knowledge base shows that this kind of approach is well-founded and effective [44].

In the same way, FCA can be used to improve ontologies associated with the Web of data. Accordingly, we proposed a method to build a concept lattice from linked data and compare the structure of this lattice with an ontology used to type the considered data. The result of this comparison makes clear some alternative axioms to be proposed to ontology developers. We extended and reused this work in ontology alignment tasks [41].

PETRUS Project-Team

7. New Results

7.1. Extensive and Secure PDMS Architecture (Axis 1)

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

The Personal Cloud paradigm is emerging through a myriad of solutions offered to users to let them gather and manage their whole digital life. This paradigm shift towards user empowerment raises fundamental questions with regards to the appropriateness of the data management functionalities and protection techniques which are offered by existing solutions to laymen users. This year, we reviewed, compared and analyzed personal cloud alternatives in terms of the functionalities they provide and the threat models they target. From this analysis, we derived a general set of security requirements that any Personal Data Management System (PDMS) should consider. We then identified the challenges of implementing such a PDMS and proposed a preliminary design for an extensive and secure PDMS reference architecture satisfying the considered requirements. Finally, we discussed several important research challenges remaining to be addressed to achieve a mature PDMS ecosystem. A first paper making the functionality and security standpoint in PDMS solutions, proposing five security goals and a preliminary architecture to fulfill these goal based on Trusted Execution Environments was published at IS'19 [12], and preliminary results on the case of a crowdsensing architecture was presented at Middleware'18 [15] and BDA'18 [18].

7.2. Data sharing model for the Personal Cloud (Axis 2)

Participants: Nicolas Anciaux [correspondent], Philippe Pucheral, Guillaume Scerri, Paul Tran Van, Baptiste Crepin.

In the PDMS context, new sharing models are needed to help end-users controlling the sharing policies under use. We proposed an architecture to produce authorizations satisfying users' sharing desires without having to trust the underlying producing these authorizations in the PhD thesis of Paul Tran-Van [11] and we demonstrated the solution at EDBT'18 [14]. We currently investigate the case of a data sharing system producing what we call 'zero-knowledge permissions', i.e., a set of authorizations produced by an untrusted sharing model which is supposed to reveal no information at all about a given subset of documents in the user space.

7.3. SEP2P: Secure and Efficient P2P Personal Data Processing (Axis 3)

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? We studied the feasibility of such a pure P2P personal data management system and provide efficient and scalable mechanisms to reduce the data leakage to its minimum with covert adversaries. In particular, we showed that data processing tasks can be assigned to nodes in a verifiable random way, which cannot be influenced by malicious colluding nodes. We proposed a generic solution which largely minimizes the verification cost. Our experimental evaluation shows that the proposed protocols lead to minimal private information leakage, while the cost of the security mechanisms remains very low even with a large number of colluding corrupted nodes. We illustrated our generic protocol proposal on three data-oriented use-cases, namely, participatory sensing, targeted data diffusion and more general distributed aggregate queries. The full protocol was simulated and evaluated. A first paper focusing on imposed randomness was published at EDBT'19 [13].

7.4. Mobile Participatory Sensing with Strong Privacy Guarantees (Axis 3)

Participant: Iulian Sandu Popa [correspondent].

Mobile participatory sensing could be used in many applications such as vehicular traffic monitoring, pollution tracking, or even health surveying. However, its 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 mobile participatory sensing. 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, we propose an enhanced version of the protocol, named PAMPAS$^+$, to make the system robust even against advanced hardware attacks on the SPs. Hence, the user location privacy leakage remains very low even for an attacker controlling the SSI and a few corrupted SPs. The leakage is proportional with the number of corrupted SPs and thus requires a massive SP corruption to break the system, which is extremely unlikely in practice. This work has been accomplished in collaboration with NJIT (see Section 9.2.1.1) and has been recently submitted as a journal paper.

7.5. Trustworthy Distributed Queries on Personal Data using TEEs (Axis 3)

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

The decentralized way of managing personal data in a PDMS provides a de facto protection against massive attacks usually performed on central servers. But this raises the question of how to preserve individuals' trust on their PDMS when performing global computations crossing data from multiple individuals? And how to guarantee the integrity of the final result when it has been computed by a myriad of collaborative but independent PDMSs? We study a secure decentralized computing framework where each participant gains the assurance that his data is only used for the purpose he consents to and that only the final result is disclosed. Conversely, the goal is to provides the querier with the guarantee that this result has been honesty computed, by the expected code on the expected data. A preliminary solution which capitalizes on the use of Trusted Execution Environments (TEE) at the edge of the network was presented at BDA'18 [19] and APVP'18 [20].

7.6. Performance of large scale data-oriented operations under TEE constraints (Axis 3)

Participants: Robin Carpentier [correspondent], Nicolas Anciaux, Iulian Sandu Popa, Guillaume Scerri.

The rise of Trusted Execution Environments like Intel SGX, and their more and more widespread use for data processing raises the question of their impact on performance, specifically for data oriented operations. While some works aim at embedding either the entirety of part of a database engine within a TEE, the direct impact of processing data with TEEs as opposed to more classical environment has not been studied yet. In particular, the cryptographic overhead of accessing persistent data outside the TEE enclave, the limited RAM amount of each TEE enclave, the cost of external function calls and memory access overheads, may slow the computing by orders of magnitude compared to a regular environment, and have to be taken into account. Preliminary results presenting both a benchmark of data operations within Intel SGX, together with optimisation of search algorithm dealing with the specific way of accessing external memory from inside SGX have been presented at BDA'18 [16].

TYREX Project-Team

6. New Results

6.1. On the Optimization of Recursive Relational Queries

Graph databases have received a lot of attention recently as they are particularly useful in many applications such as social networks or for the semantic web. Various languages have emerged to query such graph databases. At the heart of many of those query languages, there is a construction to navigate through the graph which allows some form of recursion.The relational model has benefited from a huge body of research in the last half century and that is why many graph databases either rely on, or have adopted the techniques of, relational-based 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. In this work, we introduce μ-RA, a variation of the Relational Algebra that allows for the expression of relational queries with recursion. μ-RA can express unions of conjunctive regular path queries as well as certain non-regular properties. We present its syntax, semantics and the rewriting rules we specifically devised to tackle the optimization of recursive queries. A prototype evaluator implementing these rewriting rules is shown to be more efficient than previous approaches.

These results were presented at the BDA 2018 conference [14].

6.2. A Multi-Criteria Experimental Ranking of Distributed SPARQL Evaluators

SPARQL is the standard language for querying RDF data. There exists a variety of SPARQL query evaluation systems implementing different architectures for the distribution of data and computations. Differences in architectures coupled with specific optimizations, for e.g. preprocessing and indexing, make these systems incomparable from a purely theoretical perspective. This results in many implementations solving the SPARQL query evaluation problem while exhibiting very different behaviours, not all of them being adapted to any context. We provide a new perspective on distributed SPARQL evaluators, based on multi-criteria experimental rankings. Our suggested set of 5 features (namely velocity, immediacy, dynamicity, parsimony, and resiliency) provides a more comprehensive description of the behaviours of distributed evaluators when compared to traditional runtime performance metrics. We show how these features help in more accurately evaluating to which extent a given system is appropriate for a given use case. For this purpose, we systematically benchmarked a panel of 10 state-of-the-art implementations. We ranked them using a reading grid that helps in pinpointing the advantages and limitations of current technologies for the distributed evaluation of SPARQL queries.

These results were presented at the IEEE Big Data 2018 conference [13].

6.3. SPARQL Query Containment under Schema

Query containment is defined as the problem of determining if the result of a query is included in the result of another query for any dataset. It has major applications in query optimization and knowledge base verification. The main objective of this work is to provide sound and complete procedures to determine containment of SPARQL queries under expressive description logic schema axioms. Beyond that, these procedures are experimentally evaluated. To date, testing query containment has been performed using different techniques: containment mapping, canonical databases, automata theory techniques and through a reduction to the validity problem in logic. In this work, we use the latter technique to test containment of SPARQL queries using an expressive modal logic called μ-calculus. For that purpose, we define an RDF graph encoding as a transition system which preserves its characteristics. In addition, queries and schema axioms are encoded as μ-calculus formulae. Thereby, query containment can be reduced to testing validity in the logic. We identify

various fragments of SPARQL and description logic schema languages for which containment is decidable. Additionally, we provide theoretically and experimentally proven procedures to check containment of these decidable fragments. Finally, we propose a benchmark for containment solvers which is used to test and compare the current state-of-the-art containment solvers.

These results were published in the Journal on Data Semantics [4].

6.4. Selectivity Estimation for SPARQL Triple Patterns with Shape Expressions

ShEx (Shape Expressions) is a language for expressing constraints on RDF graphs. In this work we optimize the evaluation of conjunctive SPARQL queries, on RDF graphs, by taking advantage of ShEx constraints. Our optimization is based on computing and assigning ranks to query triple patterns, dictating their order of execution. We first define a set of well-formed ShEx schemas that possess interesting characteristics for SPARQL query optimization. We then define our optimization method by exploiting information extracted from a ShEx schema. We finally report on evaluation results performed showing the advantages of applying our optimization on the top of an existing state-of-the-art query evaluation system.

These results were presented at the 2018 International Conference on Web Engineering [9].

6.5. Evaluation of Query Transformations without Data

Query transformations are ubiquitous in semantic web query processing. For any situation in which transformations are not proved correct by construction, the quality of these transformations has to be evaluated. Usual evaluation measures are either overly syntactic and not very informative — the result being: correct or incorrect — or dependent from the evaluation sources. Moreover, both approaches do not necessarily yield the same result. We suggest that grounding the evaluation on query containment allows for a data-independent evaluation that is more informative than the usual syntactic evaluation. In addition, such evaluation modalities may take into account ontologies, alignments or different query languages as soon as they are relevant to query evaluation.

These results were presented at a workshop of the 2018 International Conference on World Wide Web [10].

6.6. Graph Queries: From Theory to Practice

In this work, we review various graph query language fragments that are both theoretically tractable and practically relevant. We focus on the most expressive one that retains these properties and use it as a stepping stone to examine the underpinnings of graph query evaluation along graph view maintenance. Further broadening the scope of the discussion, we then consider alternative processing techniques for graph queries, based on graph summarization and path query learning. We conclude by pinpointing the open research directions in this emerging area. These results were published in Sigmod Record Journal [3].

6.7. Query-based Linked Data Anonymization

In this work, we introduce and develop a declarative framework for privacy-preserving Linked Data publishing in which privacy and utility policies are specified as SPARQL queries. Our approach is data independent and leads to inspect only the privacy and utility policies in order to determine the sequence of anonymization operations applicable to any graph instance for satisfying the policies. We prove the soundness of our algorithms and gauge their performance through experiments.

These results were presented in the International Semantic Web Conference (ISWC 2018) [11].

6.8. Querying Graphs

Graph data modeling and querying arises in many practical application domains such as social and biological networks where the primary focus is on concepts and their relationships and the rich patterns in these complex webs of interconnectivity. In this book, we present a concise unified view on the basic challenges which arise over the complete life cycle of formulating and processing queries on graph databases. To that purpose, we present all major concepts relevant to this life cycle, formulated in terms of a common and unifying ground: the property graph data model — the predominant data model adopted by modern graph database systems.

In this book [17], we aim especially to give a coherent and in-depth perspective on current graph querying and an outlook for future developments. Our presentation is self-contained, covering the relevant topics from: graph data models, graph query languages and graph query specification, graph constraints, and graph query processing. We conclude by indicating major open research challenges towards the next generation of graph data management systems.

6.9. 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 describe only the subtree structure of the given node. Recently, Giuseppe Castagna and our team independently proposed in 2015 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. In this work, 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 [19].

6.10. 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 develop distributed models that are scalable and interpretable. Key insights include analysing 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.

These results were presented at the 2018 International Conference on Data Science and Applications [12].

6.11. Scalable Machine Learning for Predicting At-Risk Profiles Upon Hospital Admission

We show how the analysis of very large amounts of drug prescription data make it possible to detect, on the day of hospital admission, patients at risk of developing complications during their hospital stay. We explore, for the first time, to which extent volume and variety of big prescription data help in constructing predictive models for the automatic detection of at-risk profiles. Our methodology is designed to validate our claims that: (1) drug prescription data on the day of admission contain rich information about the patient's situation and perspectives of evolution, and (2) the various perspectives of big medical data (such as veracity, volume, variety) help in extracting this information.We build binary classification models to identify at-risk patient

profiles. We use a distributed architecture to ensure scalability of model construction with large volumes of medical records and clinical data. We report on practical experiments with real data of millions of patients and hundreds of hospitals. We demonstrate how the fine-grained analysis of such big data can improve the detection of at-risk patients, making it possible to construct more accurate predictive models that significantly benefit from volume and variety, while satisfying important criteria to be deployed in hospitals.

These results were published in the Big Data Research journal [6].

6.12. ProvSQL: Provenance and Probability Management in PostgreSQL

This demonstration showcases ProvSQL, an open-source module for the PostgreSQL database management system that adds support for computation of provenance and probabilities of query results. A large range of provenance formalisms are supported, including all those captured by provenance semirings, provenance semirings with monus, as well as where-provenance. Probabilistic query evaluation is made possible through the use of knowledge compilation tools, in addition to standard approaches such as enumeration of possible worlds and Monte-Carlo sampling. ProvSQL supports a large subset of non-aggregate SQL queries.

These results were published in the PVLDB journal [8].

6.13. A Method to Quantitatively Evaluate Geo Augmented Reality Applications

We propose a method for quantitatively assessing the quality of Geo AR browsers. Our method aims at measuring the impact of attitude and position estimations on the rendering precision of virtual features. We report on lessons learned by applying our method on various AR use cases with real data. Our measurement technique allows shedding light on the limits of what can be achieved in Geo AR with current technologies. This also helps in identifying interesting perspectives for the further development of high-quality Geo AR applications.

These results were presented at the ISMAR 2018 conference [15].

6.14. Attitude Estimation for Indoor Navigation and Augmented Reality with Smartphones

We investigate the precision of attitude estimation algorithms in the particular context of pedestrian navigation with commodity smartphones and their inertial/magnetic sensors. We report on an extensive comparison and experimental analysis of existing algorithms. We focus on typical motions of smartphones when carried by pedestrians. We use a precise ground truth obtained from a motion capture system. We test state-of-the-art and built-in attitude estimation techniques with several smartphones, in the presence of magnetic perturbations typically found in buildings. We discuss the obtained results, analyze advantages and limits of current technologies for attitude estimation in this context. Furthermore, we propose a new technique for limiting the impact of magnetic perturbations with any attitude estimation algorithm used in this context. We show how our technique compares and improves over previous works. A particular attention was paid to the study of attitude estimation in the context of augmented reality motions when using smartphones.

These results were published in the Pervasive and Mobile Computing journal [7].

6.15. A Hybrid Approach for Spatio-Temporal Validation of Declarative Multimedia

Declarative multimedia documents represent the description of multimedia applications in terms of media items and relationships among them. Relationships specify how media items are dynamically arranged in time and space during runtime. Although a declarative approach usually facilitates the authoring task, authors can still make mistakes due to incorrect use of language constructs or inconsistent or missing relationships in a document. In order to properly support multimedia application authoring, it is important to provide tools with

validation capabilities. Document validation can indicate possible inconsistencies in a given document to an author so that it can be revised before deployment. Although very useful, multimedia validation tools are not often provided by authoring tools. This work proposes a multimedia validation approach that relies on a formal model called Simple Hyper-media Model (SHM). SHM is used for representing a document for the purpose of validation. An SHM document is validated using a hybrid approach based on two complementary techniques. The first one captures the document's spatio-temporal layout in terms of its state throughout its execution by means of a rewrite theory, and validation is performed through model checking. The second one captures the document's layout in terms of intervals and event occurrences by means of Satisfiability Modulo Theories (SMT) formulas, and validation is performed through SMT solving. Due to different characteristics of both approaches, each validation technique complements the other in terms of expressiveness of SHM and tests to be checked. We briefly present validation tools that use our approach. They were evaluated with real NCL documents and by usability tests.

These results were published in the ACM Transactions on Multimedia Computing, Communications and Applications journal [5].

VALDA Project-Team

6. New Results

6.1. Query Enumeration

Query enumeration is the problem of enumerating the results of a query over a database one by one; the goal is to obtain, after some initial low preprocessing time (e.g., linear in the data), one solution after the other with low delay (e.g., constant-time) in between.

In a first work [26], we consider the enumeration of MSO queries over strings under updates. For each MSO query we build an index structure enjoying the following properties: The index structure can be constructed in linear time, it can be updated in logarithmic time and it allows for constant delay time enumeration. This improves from the previous known index structures allowing for constant delay enumeration that would need to be reconstructed from scratch, hence in linear time, in the presence of updates. We allow relabeling updates, insertion of individual labels and removal of individual labels.

In a second work [29], we consider the evaluation of first-order queries over classes of databases that are nowhere dense. The notion of nowhere dense classes was introduced by Nešetřil and Ossona de Mendez as a formalization of classes of "sparse" graphs and generalizes many well-known classes of graphs, such as classes of bounded degree, bounded treewidth, or bounded expansion. It has recently been shown by Grohe, Kreutzer, and Siebertz that over nowhere dense classes of databases, first-order sentences can be evaluated in pseudo-linear time (pseudo-linear time means that for all ε there exists an algorithm working in time $O(n^{1+\varepsilon})$, where n is the size of the database). For first-order queries of higher arities, we show that over any nowhere dense class of databases, the set of their solutions can be enumerated with constant delay after a pseudo-linear time preprocessing. In the same context, we also show that after a pseudo-linear time preprocessing we can, on input of a tuple, test in constant time whether it is a solution to the query.

6.2. Provenance Circuits

We are interested in obtaining efficiently compact representation of the provenance of a query over a database.

In [28], we generalize three existing graph algorithms to compute the provenance of regular path queries over graph databases, in the framework of provenance semirings – algebraic structures that can capture different forms of provenance. Each algorithm yields a different trade-off between time complexity and generality, as each requires different properties over the semiring. Together, these algorithms cover a large class of semirings used for provenance (top-k, security, etc.). Experimental results suggest these approaches are complementary and practical for various kinds of provenance indications, even on a relatively large transport network.

In [16], we showcase ProvSQL, an open-source module for the PostgreSQL database management system that adds support for computation of provenance and probabilities of query results. A large range of provenance formalisms are supported, including all those captured by provenance semirings, provenance semirings with monus, as well as where-provenance. Probabilistic query evaluation is made possible through the use of knowledge compilation tools, in addition to standard approaches such as enumeration of possible worlds and Monte-Carlo sampling. ProvSQL supports a large subset of non-aggregate SQL queries.

Finally, in [20], [35], we focus on knowledge compilation, which can be used to obtain compact circuit-based representations of (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. We show how to connect the width of circuits to the size of their structured representation, through upper and lower bounds. For the upper bound, we show how bounded-treewidth circuits can be converted to a d-SDNNF, in time linear in the circuit size. Our bound, unlike existing results, is constructive and only singly exponential in the treewidth. We show a related lower bound on monotone DNF or CNF formulas, assuming a constant bound on the arity (size of clauses) and degree (number of

occurrences of each variable). Specifically, any d-SDNNF (resp., SDNNF) for such a DNF (resp., CNF) must be of exponential size in its treewidth; and the same holds for pathwidth when compiling to OBDDs. Our lower bounds, in contrast with most previous work, apply to any formula of this class, not just a well-chosen family. Hence, for our language of DNF and CNF, pathwidth and treewidth respectively characterize the efficiency of compiling to OBDDs and (d-)SDNNFs, that is, compilation is singly exponential in the width parameter.

6.3. Exploiting Content from the Web

One of our main domain of application is that of Web content. We investigate methods to acquire and exploit content from the Web.

In [30], we analyze form-based websites to discover sequences of actions and values that result in a valid form submission. Rather than looking at the text or DOM structure of the form, our method is driven by solving constraints involving the underlying client-side JavaScript code. In order to deal with the complexity of client-side code, we adapt a method from program analysis and testing, concolic testing, which mixes concrete code execution, symbolic code tracing, and constraint solving to find values that lead to new code paths. While concolic testing is commonly used for detecting bugs in stand-alone code with developer support, we show how it can be applied to the very different problem of filling Web forms. We evaluate our system on a benchmark of both real and synthetic Web forms.

In [21], we investigate *focused crawling*: collecting as many Web pages relevant to a target topic as possible while avoiding irrelevant pages, reflecting limited resources available to a Web crawler. We improve on the efficiency of focused crawling by proposing an approach based on reinforcement learning. Our algorithm evaluates hyperlinks most profitable to follow over the long run, and selects the most promising link based on this estimation. To properly model the crawling environment as a Markov decision process, we propose new representations of states and actions considering both content information and the link structure. The size of the state-action space is reduced by a generalization process. Based on this generalization, we use a linear-function approximation to update value functions. We investigate the trade-off between synchronous and asynchronous methods. In experiments, we compare the performance of a crawling task with and without learning; crawlers based on reinforcement learning show better performance for various target topics.

Finally, in [23], [24] we propose a framework to follow the dynamics of vanished Web communities, based on the exploration of corpora of Web archives. To achieve this goal, we define a new unit of analysis called Web fragment: a semantic and syntactic subset of a given Web page, designed to increase historical accuracy. This contribution has practical value for those who conduct large-scale archive exploration (in terms of time range and volume) or are interested in computational approaches to Web history and social science.

6.4. Knowledge Bases

Knowledge bases are collection of semantic facts (typically of the form subject–predicate–object) along with possible logical rules (e.g., in the form of existential rules) that apply to these facts. We investigate querying, data integration, and inference in such knowledge bases.

In [27], we focus on autocompletion of SPARQL queries over knowledge bases. We analyze several autocompletion features proposed by the main editors, highlighting the needs currently not taken into account while met by a user community we work with, scientists. Second, we introduce the first (to our knowledge) autocompletion approach able to consider snippets (fragments of SPARQL query) based on queries expressed by previous users, enriching the user experience. Third, we introduce a usable, open and concrete solution able to consider a large panel of SPARQL autocompletion features that we have implemented in an editor. Last but not least, we demonstrate the interest of our approach on real biomedical queries involving services offered by the Wikidata collaborative knowledge base.

In [25], we introduce a novel open-source framework for integrating the data of a user from different sources into a single knowledge base. Our framework integrates data of different kinds into a coherent whole, starting with email messages, calendar, contacts, and location history. We show how event periods in the user's location data can be detected and how they can be aligned with events from the calendar. This allows users to query their personal information within and across different dimensions, and to perform analytics over their emails, events, and locations. Our system models data using RDF, extending the schema.org vocabulary and providing a SPARQL interface.

Finally, in [22], [32], we view knowledge bases as composed of an instance that contains incomplete data and a set of existential rules, and investigate ontology-based query answering: answers to queries 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. It is well-known that the problem of determining, given a chase variant and a set of existential rules, whether the chase will halt on a given instance / on any instance, is undecidable. Hence, a crucial issue is whether it becomes decidable for known subclasses of existential rules. We consider linear existential rules, a simple yet important subclass of existential rules. We study the decidability of the associated chase termination problem for different chase variants, with a novel approach based on a single graph and a single notion of forbidden pattern. Besides the theoretical interest of a unified approach, an original result is the decidability of the restricted chase termination for linear existential rules.

6.5. Transparency and Bias

In this last set of results, we investigate transparency and bias in data management.

Bias in online information has recently become a pressing issue, with search engines, social networks and recommendation services being accused of exhibiting some form of bias. In [15], we make the case for a systematic approach towards measuring bias. To this end, we discuss formal measures for quantifying the various types of bias, we outline the system components necessary for realizing them, and we highlight the related research challenges and open problems.

In [19], we pursue an investigation of data-driven collaborative work-flows. In the model, peers can access and update local data, causing side-effects on other peers' data. In this paper, we study means of explaining to a peer her local view of a global run, both at runtime and statically. We consider the notion of "scenario for a given peer" that is a subrun observationally equivalent to the original run for that peer. Because such a scenario can sometimes differ significantly from what happens in the actual run, thus providing a misleading explanation , we introduce and study a faithfulness requirement that ensures closer adherence to the global run. We show that there is a unique minimal faithful scenario, that explains what is happening in the global run by extracting only the portion relevant to the peer. With regard to static explanations, we consider the problem of synthesizing, for each peer, a "view program" whose runs generate exactly the peer's observations of the global runs. Assuming some conditions desirable in their own right, namely transparency and boundedness, we show that such a view program exists and can be synthesized. As an added benefit, the view program rules provide provenance information for the updates observed by the peer.

Finally, in two articles oriented towards applications and policy, we discuss bias and neutrality and their impact on regulation. In [18] we discuss the different forms of neutrality in the digital world, from the neutrality of networks to neutrality of content. In [17], we investigate the impact of bias and neutrality concerns on algorithms used by businesses.

WIMMICS Project-Team

7. New Results

7.1. Users Modeling and Designing Interaction

7.1.1. User-centered Heuristics for the Control of Personal Data

Participants: Alain Giboin, Patrice Pena, Fabien Gandon.

This work (done in collaboration with Karima Boudaoud and Yoann Bertrand, SPARKS, I3S, in the context of the PadDOC FUI project) led to the elaboration and the evaluation of a set of user-centered heuristics and a procedure for designing and evaluating systems allowing the control of personal data. The elaboration of the heuristics was based on: (1) the transposal of Nielsen's heuristics and of Scapin and Bastien's ergonomic criteria to the control of personal data ; (2) the user centering of the Privacy-by-Design notion of integrated privacy; and (3) the integration of Altman's interaction approach to privacy.

7.1.2. Needs Analysis of the Target Users of the WASABI musical search platform

Participants: Alain Giboin, Isabelle Mirbel, Michel Buffa, Elmahdi Korfed.

In the context of the ANR project WASABI, we performed an analysis of the needs of the target users of the future WASABI platform. This analysis has been reported in an internal report.

7.1.3. Modeling the Users of Collaborative Ontology Building Environments

Participant: Alain Giboin.

We undertook a study on the evolution of the user model of collaborative ontology building environments (COBEs). By a user model – or a contributor model – we refer to the representation that COBEs designers have of the users of their systems and more generally of the actors contributing to the building of ontologies. This study aimed at emphasizing the importance to get a better knowledge of potential COBE contributors in order to design collaborative tools better suited to these contributors. The study was published in [55]. In this paper, we describe: (1) the method we used to study the evolution of the user/contributor model; (2) the evolution of the model (in terms of user types, user characterizations, and user's environment characterizations); (3) the parallel evolutions of: (a) the methods of COBEs design, (b) the systems themselves, and (c) the methods of collaborative ontology building; we mention some evolution perspectives envisioned by the designers.

7.1.4. Design of a User-Centered Evaluation Method for Exploratory Search Systems

Participants: Emilie Palagi, Alain Giboin, Fabien Gandon.

This work was undertaken in the context of the PhD of Emilie Palagi, in cooperation with Raphaël Troncy (EURECOM). Our method takes into account users' exploratory search (ES) behavior and is based on a cognitive model of an ES task. We specially work on Discovery Hub (Wimmics project – Inria) and 3cixty (EURECOM project) ESSs. During the third year of the PhD, we continued the evaluation of our model of exploratory search by comparing it to video records of seven other ES sessions on Discovery Hub, Frankenplace and 3cixty. We analyzed the videos with the same methodology: we wrote down the different chains of the different model's features used by the users in their ES session. For all the records we were able to identify the features of our model and extend our table of observed possible transitions between the model's features. From this analysis, we conclude that our model of ES can express the users' activity during an ES task. This work was partially published in [49].

Based on the ES model's features and the possible transitions between them, we designed two different evaluation and design methods of ES systems which do not necessarily involve users:

- Without users: Heuristics of ES and a procedure to use them. These heuristics are principles for the interaction design. The Heuristics of ES can be used several times along the design process of the ES system (in the design and evaluation phases). We presented the heuristics and evaluated them. This work was published in [48].
- With users: a guide for the elaboration of a customizable test protocol. The goal of the test is to analyze ES session records in order to find the model's features. In this guide, we give indications to customize the protocol and prepare users tests. We focused on two model-based elements of this customizable test protocol: a protocol for the elaboration of exploratory search tasks, and a video analysis grid for the evaluation of recorded exploratory search sessions.

7.1.5. Supporting Learning Communities with Intelligent services

Participants: Oscar Rodríguez Rocha, Catherine Faron Zucker.

The *Système Intelligent d'Enseignement en Santé 3.0* (SIDES 3.0), (Intelligent Health Education System 3.0), is a 3 years project funded by the French National Agency for Research (ANR) within the framework of the call for projects DUNE 2016. It builds upon a national Web platform, the *Système Informatique D'Evaluation en Santé (SIDES)* (Health Assessment Information System), used since 2013 by the faculties of medicine in France which enables them to perform all of their validation exams on tablets, providing them with automatic corrections. It contributes to the preparation of medical students to perform the *Epreuves Classantes Nationales (ECN) informatisées (ECNi)* (Computerized National Qualifying Events) which have been successfully held in France in June 2016 (8000 candidates simultaneously throughout France). The SIDES platform is administered by the 35 medicine faculties in France and is used by more than 70,000 students throughout their training. The system is also used to prepare students for *ECNi*. Over the last 3 years, more than 4 million clinical cases (made up of 15 questions each) have been performed by students (all activities combined).

Building on this success, the SIDES 3.0 project aims to upgrade the SIDES solution to an innovative solution providing the user with intelligent learning services based on a modelization of the pedagogical resources with Semantic Web models and technologies. It is coordinated by the *Université Numérique Thématique (UNT) en Santé et Sport*[0]. This structure offers an ideal national positioning for support and coordination of training centers (UFR) and also offers long-term financial sustainability. In this framework, we focus on developing and applying adaptive learning approaches to automatic quiz generation from existing questions, and quiz recommendation adapted to user profiles and learning contexts, to allow medical students to better achieve their educational objectives by answering quizzes [50], [51].

7.1.6. Explainable Predictions Using Product Reviews

Participants: Elena Cabrio, Fabien Gandon, Nicholas Halliwell, Freddy Lecue, Serena Villata.

This is a joint work between Accenture and Wimmics team, funded by Accenture. The goal of this project is to design a recommender system that returns explainable predictions to the user, incorporating text from the product reviews in the explanation. To start, we have replicated results from current state of the art methods. We then gathered a dataset of Amazon books and corresponding reviews, and ran the current state of the art algorithm on our dataset. The next steps will be to build a deep learning model to outperform the current state of the art algorithm, and develop a method to explain the predictions to the user using the product reviews.

7.1.7. Argument Mining

Participants: Elena Cabrio, Fabien Gandon, Claude Frasson, Andrea Tettamanzi.

[0] http://www.uness.fr

We have published a survey paper about Argument Mining at IJCAI [61]. Argument mining is the research area aiming at extracting natural language arguments and their relations from text, with the final goal of providing machine-processable structured data for computational models of argument. This research topic has started to attract the attention of a small community of researchers around 2014, and it is nowadays counted as one of the most promising research areas in Artificial Intelligence in terms of growing of the community, funded projects, and involvement of companies. In this paper, we presented the argument mining tasks and we discussed the obtained results in the area from a data-driven perspective. An open discussion highlights the main weaknesses suffered by the existing work in the literature and proposes open challenges to be faced in the future.

Together with two colleagues from FBK Trento (Italy), we applied argumentation mining techniques, in particular relation prediction, to study political speeches in monological form, where there is no direct interaction between opponents. We argued that this kind of technique can effectively support researchers in history, social and political sciences, which must deal with an increasing amount of data in digital form and need ways to automatically extract and analyse argumentation patterns. We tested and discussed our approach based on the analysis of documents issued by R. Nixon and J. F. Kennedy during 1960 presidential campaign. We relied on a supervised classifier to predict argument relations (i.e., support and attack), obtaining an accuracy of 0.72 on a dataset of 1,462 argument pairs. The application of argument mining to such data allowed not only to highlight the main points of agreement and disagreement between the candidates' arguments over the campaign issues such as Cuba, disarmament and health-care, but also an in-depth argumentative analysis of the respective viewpoints on these topics. The results of this research have been published at AAAI [58].

In this direction, we have also, in collaboration with the Heron Lab of the University of Montreal, presented an empirical study about the relation between argumentative persuasion and emotions. Argumentative persuasion usually employs one of the three persuasion strategies: Ethos, Pathos or Logos. Several approaches have been proposed to model persuasive agents, however, none of them explored how the choice of a strategy impacts the mental states of the debaters and the argumentation process. We conducted a field experiment with real debaters to assess the impact of the mental engagement and emotions of the participants, as well as of the persuasiveness power of the arguments exchanged during the debate. Our results showed that the Pathos strategy is the most effective in terms of mental engagement. The results of this research have been published at FLAIRS [60].

Together with Souhila Kaci (LIRMM) and Leendert van der Torre (University of Luxembourg), we have proposed a formal framework to reason about preferences in abstract argumentation. Consider an argument A that is attacked by an argument B, while A is preferred to B. Existing approaches will either ignore the attack or reverse it. We introduced a new reduction of preference and attack to defeat, based on the idea that in such a case, instead of ignoring the attack, the preference is ignored. We compared this new reduction with the two existing ones using a principle-based approach for the four Dung semantics. The principle-based or axiomatic approach is a methodology to choose an argumentation semantics for a particular application, and to guide the search for new argumentation semantics. For this analysis, we also introduced a fourth reduction, and a semantics for preference-based argumentation based on extension selection. Our classification of twenty alternatives for preference-based abstract argumentation semantics using six principles suggests that our new reduction has some advantages over the existing ones, in the sense that if the set of preferences increases, the sets of accepted arguments increase as well. The results of this research have been published at COMMA [36].

Together with Celia da Costa Pereira (I3S) and Mauro Dragoni (FBK Trento), we presented SMACk, an opinion summary system built on top of an argumentation framework with the aim to exchange, communicate and resolve possibly conflicting viewpoints. SMACk allows the user to extract debated opinions from a set of documents containing user-generated content from online commercial websites, and to automatically identify the mostly debated positive aspects of the issue of the debate, as well as the mostly debated negative ones. The key advantage of such a framework is the combination of different methods, i.e., formal argumentation theory and natural language processing, to support users in making more informed decisions, e.g., in the context of online purchases. The results of this research have been published in the AI Communications journal [14].

7.2. Communities and Social Interactions Analysis

7.2.1. Argumentation and Emotion Detection with Adaptive Sentiment Analysis

Participants: Vorakit Vorakitphan, Serena Villata, Elena Cabrio.

This PhD work just started in the context of the ANSWER project with Qwant search engine. One of the main objectives of the ANSWER project is to use emotion detection algorithms within text inquiries and sentiment analysis to provide powerful enhancements in the search results from Qwant search engine. The final goal is to extract effective and scalable indicators of sentiment, emotions, and argumentative relations in order to offer the users additional means to filter the results selected by the search engine. Powerful algorithms in state-of-art will be focused to define new criteria for filtering search results, i.e., the expression of a feeling in the answers found by the search engine. By doing as mentioned, textual elements to which we wish to associate a polarity will no longer be considered in their individuality but connected to each other by polarized relations to be analyzed in a higher level setting. Currently, the work progress is in the survey of state-of-the-art based on emotion detection algorithms and implementation of sentiment analysis. Then the next target, classification models with multi-label features based on emotion detection, will be deeply explored as a starting point of this research. Moreover, NLP related to emotional news content will be taken into account to build a novel dataset based on emotion annotation from news articles in sentence-level.

7.2.2. Cyberbullying Events Prevention

Participants: Pinar Arslan, Michele Corazza, Elena Cabrio, Serena Villata.

In the CREEP EIT project, we built an emotion detection classifier to automatically identify the emotion for user-generated texts such as Twitter and Instagram posts. The correlation analysis that we carried out to get a better understanding of the associations between emotions and cyberbullying instances unveiled that certain emotions (e.g., anger, joy) would be good indicative features to detect cyberbullying instances. Hence, our pipeline firstly reveals automatically detected emotion labels for social media texts to be used to detect cyberbullying instances. The automatically predicted emotion labels were used as one of the predictors for our cyberbullying detection classifier. As part of the project, we successfully built a classifier for offensive language in social media interactions for English, Italian and German using neural networks. This classifier was evaluated by participating in two shared tasks: Germeval (German offensive language detection) and Evalita (Italian hate speech detection). For the Germeval Challenge [29], two systems for predicting message-level offensive language in German tweets were used: one discriminates between offensive and not offensive messages, and the second performs a fine-grained classification by recognizing also classes of offense. Both systems are based on the same approach, which builds upon Recurrent Neural Networks used with the following features: word embeddings, emoji embeddings and social-network specific features. The model combines word-level information and tweet-level information to perform the classification tasks. Our best performing model ranked 7th out of 51 submitted runs on the binary classification task, 5th out of 25 for the fine-grained classification task. For the Evalita Challenge shared tasks [28], our submissions were based on three separate classes of models: a model using a recurrent layer, an ngram-based neural network and a LinearSVC. For the Facebook task and the two cross-domain tasks we used the recurrent model and obtained promising results, especially in the cross-domain setting. For Twitter, we used an ngram-based neural network and the Linear SVC-based model. Our system ranked 1st in the Facebook to Twitter dataset, 2nd in the Twitter to Facebook dataset, 3rd in the Facebook dataset and 4th on the Twitter dataset.

7.2.3. Modeling of a Social Network of Service Providers

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

In the framework of a collaborative project with Silex France company and the CIFRE PhD thesis of Molka Dhouib, our aim is to model the social network of service providers and companies registered in the *software as a service* sourcing tool developed by Silex for the recommendation of the service providers that are best suited to meet the service requests expressed by companies. Our aim is to automate the matching of service requests and offers by reasoning on the social network of service providers and companies. We developed an automatic categorization of companies, service requests and service offers based on their textual descriptions. We conducted some experiments using state-of-the-art supervised Machine Learning techniques to classify

Silex textual data into predefined categories, and to choose the best vector representations of the textual descriptions of service offers and requests in the Silex platform, and the best Machine Learning algorithm. This work has been presented at the French conference on applications of Artificial Intelligence APIA2018 [31].

7.3. Vocabularies, Semantic Web and Linked Data based Knowledge Representation

7.3.1. Modeling a Vocabulary of Professional Skills and Fields of Activities

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 a dedicated vocabulary of competences and fields of activities to semantically annotate B2B service offers. We started with the study of existing reference taxonomies representing skills, professions and fields of activities and we formalized them in SKOS. Then we built a SKOS vocabulary from the internal Silex repositories. Finally we performed a semi-automatic alignment of these vocabularies. This work has been presented at the French conference on Knowledge Engineering IC 2018 [53].

7.3.2. Representing and Querying a Knowledge Graph on Pedagogical Resources

Participants: Géraud Fokou Pelap, Catherine Faron Zucker, Fabien Gandon, Olivier Corby.

In the framework of the EduMICS (Educative Models Interactions Communities with Semantics) joint laboratory (LabCom) between the Wimmics team and the Educlever company, we built a knowledge graph from the database of the Educlever platform describing learning resources, and related knowledge and skills. We deployed our proposed Semantic Web based solution within the industrial environment of Educlever, using Web services, and we showed the added value of Semantic Web modelling enabling to implement new functionalities with SPARQL queries on the knowledge graph. This work has been presented at the SemWeb.Pro 2018 day [56] and at the WEBIST conference [34].

7.3.3. A Learnable Crawler for Linked Open Data

Participants: Hai Huang, Fabien Gandon.

This work is supported by the ANSWER project in cooperation with Qwant company. It consists of designing a learnable Linked Data crawler featured by a prediction component which is able to predict whether a newly discovered URI contains RDF data or not.

As the Web of Linked Open Data is growing exponentially, crawling for Linked Data has become increasingly important. Unlike normal Web crawlers, a Linked Data crawler performs selectively to collect linked RDF (including RDFa) data on the Web. From the perspectives of throughput and coverage, given a newly discovered URI, the key issue of Linked Data crawlers is to decide whether this URI is desirable to download (if it contains RDF data). Current solutions adopt heuristic rules aiming to filter irrelevant URIs. Unfortunately, it would hurt the coverage of crawling. In this work, we developed a learnable Linked Data crawler featured by a prediction component which is able to predict whether a newly discovered URI contains RDF data or not. We extracted useful features from the context RDF graph of the URI. The prediction model is based on FTRL-proximal [0] online learning algorithm. We evaluated it through extensive experiments in comparison with a number of baseline methods and demonstrated its efficiency.

7.3.4. Argument Mining on Clinical Trials

Participants: Tobias Mayer, Serena Villata, Elena Cabrio.

[0] FTRL: Follow The Regularized Leader

This work was done in the context of the PhD of Tobias Mayer, which is situated in the IADB project, "Intégration et Apprentissage sur les Données Biomédicales". We created a new annotated dataset of Randomized Controlled Tirals (RCT) about four different diseases (glaucoma, diabetes, hepatitis B, and hypertension), containing 976 argument components (697 containing evidence, 279 claims) together with a first approach for the argumentative component detection [39]. Empirical results are promising and show the portability of the proposed approach over different branches of medicine. Furthermore, we proposed a new sub-task of the argument component identification task: evidence type classification, which distinguishes the provided evidence on a more fine-grained level. To address it, we proposed a supervised approach and we tested it on our data set [40].

As a collaboration with "Base, Corpus, Language" (BCL) at UCA within the IADB project, we anonymized and cleaned clinical reports (from CHU Nice), built a "raw" French corpus from it and are currently working on transferring the above mentioned annotations and models to this data set.

7.3.5. Structure Detection in Song Lyrics

Participants: Michael Fell, Elena Cabrio, Fabien Gandon.

In the context of the WASABI ANR project, we work on the estimation of the structure of song lyrics. For this, we have built a predictive model that successfully segments song texts into their underlying paragraphs - a task called "Lyrics Segmentation". We have augmented existing state-of-the-art models for Lyrics Segmentation in two ways: (i) by applying convolutional neural networks to the task alongside of novel feature representations. This work resulted in a publication at the COLING conference [33]; (ii) by extending the feature representation with time-synchronized audio features, we improve the segmentation model performance. It can now also use audio cues when text cues are non-indicative; this improves segmentation performance. Our current endeavors aim at summarizing song texts so that journalists and musicologists can perform efficient searches under different perspectives (e.g. structure and semantic content).

7.3.6. Legal Information, Privacy

Participants: Elena Cabrio, Serena Villata.

Together with Valentina Leone and Luigi di Caro (University of Torino), we presented the *InvestigatiOnt* tool which aims to ease the interaction of end users with legal ontologies in order to spread the use of machine-processable legal information as well as its understanding. This research is addressed in the context of the EU H2020 MIREL project. The results of this research have been published as demo paper at ISWC [71].

Together with Sabrina Kirrane (Vienna University of Economics and Business) and Matthieu d'Aquin (National University of Ireland Galway), we examined 78 articles from dedicated venues, the Privacy Online workshop series, two SPOT workshops, as well as the broader literature that connects the Semantic Web research domain with issues relating to privacy, security and/or policies. Specifically, we classified each paper according to three taxonomies (one for each of the aforementioned areas), in order to identify common trends and research gaps. We concluded by summarising the strong focus on relevant topics in Semantic Web research (e.g. information collection, information processing, policies and access control), and by highlighting the need to further explore under-represented topics (e.g., malware detection, fraud detection, and supporting policy validation by data consumers). The results of this research have been published in the Semantic Web journal [16].

7.3.7. Semantic Web for Biodiversity

Participants: Franck Michel, Catherine Faron Zucker.

The collaboration initiated with the French National Museum of Natural History of Paris (MNHN) is now giving rise to the development of an activity related to biodiversity data sharing and integration.

The TAXREF-LD linked data dataset, that we produced jointly with the MNHN, now appears in the Linked Open Data cloud [0] and is published on AgroPortal [0], the ontology Web portal for agronomy and agriculture. At the Biodiversity Information Standards conference (TDWG 2018), we presented some insights in the modelling of biodiversity Linked Data [45], we demonstrated how SPARQL Micro-Services can help in the integration of heterogeneous biodiversity-related data sources [43]. We also presented a poster on the Bioschemas.org initiative [46], a W3C community group that seeks the definition and adoption of common biology-related markup. In this context, we have proposed a first specification of the Taxon term [0] whose adoption as part of the official Schema.org vocabulary is currently being discussed with Google.

We took part in the D2KAB ANR project submission that aims to turn agronomy and biodiversity data into semantically described, interoperable, actionable open-knowledge. The project has been accepted and is due to start in June 2019.

7.3.8. 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 published two contributions this year, together with Johan Montagnat from I3S.

First, we proposed a generic method to bridge the gap between the Semantic Web and NoSQL worlds [42]. To avoid defining yet another SPARQL translation method for each and every database, a SPARQL query is translated into a pivot abstract query, spanning all database-independent steps. Only then, the abstract query is translated into the target database query language while taking into account the specific database capabilities and constraints.

Second, we defined the SPARQL Micro-Service architecture that harnesses the Semantic Web standards to enable automatic combination of Linked Data and data residing in Web APIs (aka. REST Web services). A SPARQL micro-service is a lightweight, task-specific SPARQL endpoint that provides access to a small, resource-centric virtual graph, while dynamically assigning dereferenceable URIs to Web API resources that do not have URIs beforehand. The graph is delineated by the Web API service being wrapped, the arguments passed to this service, and the restricted types of RDF triples that this SPARQL micro-service is designed to spawn.

This work was presented at the ESWC conference [42] and the LDOW workshop at the Web Conference [44]. We also conducted an experimentation where we dynamically augment biodiversity Linked Data with data from multiple Web APIs: Flickr, Biodiversity Heritage Library, Encyclopedia of Life, Macauley scientific media archive, and MusicBrainz [43].

7.3.9. Linked Data Script Language

Participant: Olivier Corby.

We have designed and implemented LDScript, a programming language compatible with SPARQL that enables users to write extension functions that are directly executable in SPARQL queries.

We have leveraged pattern matching for structured objects such as lists where we can retrieve first elements, intermediate sublist and last elements. We have defined event driven processing where the SPARQL interpreter emits events which are processed by LDScript functions. The function definitions are annotated with event names. This enables users to trace query execution, to overload SPARQL statements such as "*order by, distinct*" and to extend SPARQL with new statements implemented as functions. In particular we are able to overload SPARQL operators for extension datatypes such as romain numbers or values with units. We are also able to trap and overload SPARQL execution errors with specific LDScript functions. In addition, we have introduced a second order "eval" function that enables us to evaluate the arguments of expressions that caused an error.

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

LDScript has been extended in order to process SPARQL Update in addition to SPARQL Query. Hence LDScript can be used to implement Semantic Web services with the following statements: SPARQL Query and Update, OWL RL entailment, RDF transformation to HTML.

This a follow up work on the formalism that was originally published at ISWC 2017 [72].

7.3.10. Graphic Display for RDF Graphs

Participants: Olivier Corby, Erwan Demairy.

This work has been done in the context of an Inria funding for software development (ADT).

In order to perform Linked Data visualisation, we connected the D3.js graphic display library to the Corese Web server. We designed an STTL transformation that generates D3 graph format with stylesheet from RDF graph. The graph display is performed thanks to SVG code generated by D3. The graph display can be interactive, that is hypertext navigation can be associated with a click on graph nodes. We have setup a demo with HAL open data server [0], see figure 1 .

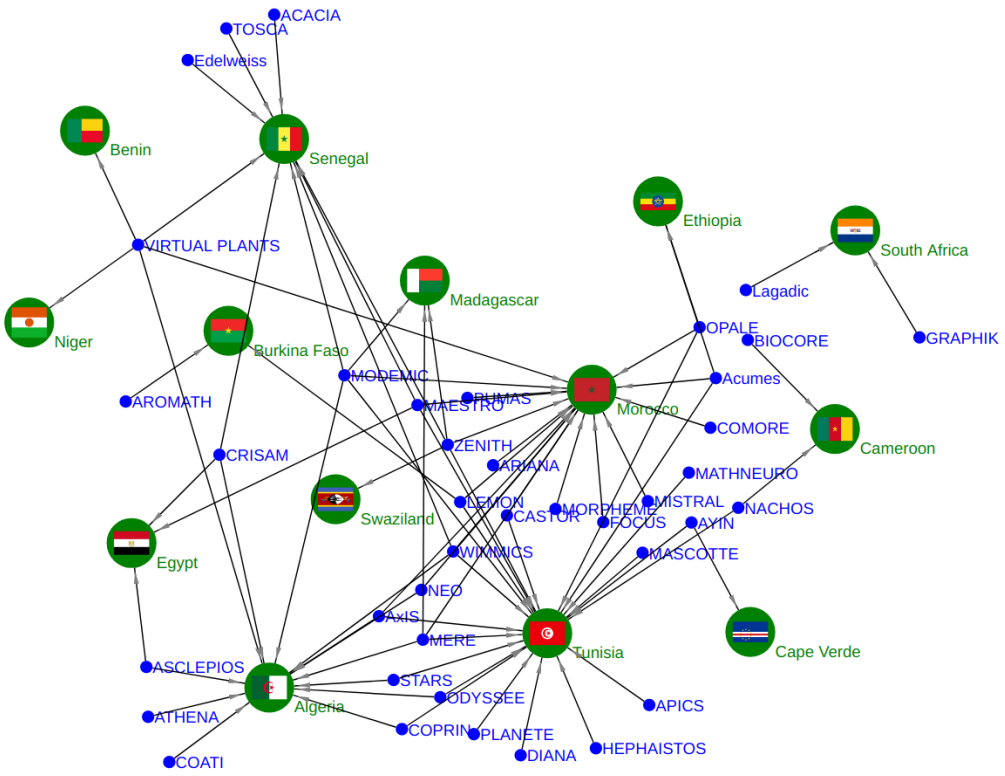


Figure 1. Inria Sophia Antipolis teams publishing with south American countries.

This work has been presented at the software development day at Inria Sophia Antipolis, November 14th.

[0] http://corese.inria.fr

7.3.11. Federated Query Scaler

Participant: Olivier Corby.

This work is done in the context of the *Federated Query Scaler* Inria exploratory research project (PRE) together with Olivier Dameron and Vijay Ingalalli from Dyliss team at Inria Rennes.

In this project, focused on SPARQL federated queries, Vijay Ingalalli designed a graph index for distributed SPARQL endpoints that enables us to predict whether joins between patterns can be performed within endpoints. We also wrote a compiler that generates a SPARQL query with service clauses from a federated query, that is a query annotated with several SPARQL endpoints URL.

We welcomed Vijay Ingalalli at Inria Sophia Antipolis, January 15-19, and Olivier Corby visited the Dyliss team in Rennes, March 4-6.

7.4. Analyzing and Reasoning on Heterogeneous Semantic Graphs

7.4.1. Distributed Artificial Intelligence for Revisable Linked Data Management

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

The aim of this PhD thesis is to study and to propose original solutions to many key aspects: Knowledge Representation in case of uncertain, incomplete and reviewable data; Uncertainty Representation in a data source, with provenance; Distributed Knowledge Revision and Propagation; Reasoning over Uncertain, Incomplete and distributed data-sources. Starting from an open Web of Data, this work tries to give the users more objective, exhaustive and certain views and information about their queries, based on distributed data sources with different levels of certainty and trustworthiness. We proposed a vocabulary to formalize uncertainty representation, and a framework to handle uncertainty mapping to sentences and contexts. This work has been presented as a poster at ISWS [68].

7.4.2. Learning Class Disjointness Axioms using Grammatical Evolution

Participants: Thu Huong Nguyen, Andrea Tettamanzi.

The aim of this research is to discover automatically class disjointness axioms from recorded RDF facts on the Web of Data. This may be regarded as a case of inductive reasoning and ontology learning. The instances, represented by RDF triples, play the role of specific observations, from which axioms can be extracted by generalization. We proposed the use of Grammatical Evolution, one type of evolutionary algorithm, for mining disjointness OWL2 axioms from an RDF data repository such as DBpedia. For the evaluation of candidate axioms against the DBpedia dataset, we adopt an approach based on possibility theory. We have submitted a paper to the conference EuroGP 2019.

7.4.3. Semantic Data for Image Recognition

Participants: Anna Bobasheva, Fabien Gandon.

This work is done in the context of the MonaLIA project with French Ministry of Culture, in collaboration with Frédéric Precioso, I3S, UCA. It consists of a preliminary study on image recognition of the Joconde database in connection with semantic data (JocondeLab).

The goal of this project is to exploit the cross-fertilization of recent advances in image recognition and semantic indexing on annotated image databases in order to improve the accuracy and the details of the annotation. The idea is, at first, to assess the potential of machine learning (including deep learning) and the semantic annotations on the Joconde database (350 000 illustrated artwork records from French museums). Joconde also contains metadata based on a thesaurus. In a previous project (JocondeLab) these metadata were formalized in Semantic Web formalism and were linking the iconographic Garnier thesaurus and DBpedia to the data of the Joconde database.

We developed SPARQL queries on Joconde database to extract the subset of images to train the Deep Learning classifier. We identified class subsets with enough labeled images for training, we balance number of images per class and we avoid images with intersected classes.

We tuned the pre-trained VGG16 implementation of the CNN classifier to classify the artwork images using well-known VGG16 with batch normalization [75] to train the classifier for the artwork images. We learned transfer from the training of the network on the ImageNet dataset to decrease the training time and we ran the classifier on many datasets and in different modes.

We developed another set of queries on the metadata to find the dependencies between the classification outcome and the artwork properties by applying statistical methods. We identified the usable (populated enough with reasonable number of categorical values) properties of the metadata. We used Recursive Feature Elimination (RFE) and Decision Tree to identify the top most statistically significant dependent variables and decision splitting values.

Results have been presented at a workshop of Ministry of Culture and Inria, November 22nd, at Bibliothèque Nationale de France in Paris.

7.4.4. Hospitalization Prediction

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

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. The first results of this project will be presented at the French EGC 2019 conference [54]. In this paper, we report and discuss the results of our first experiments on the database PRIMEGE PACA that contains more than 350,000 consultations carried out by 16 general practitioners. 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.

7.4.5. Fake News Detection

Participants: Jérôme Delobelle, Elena Cabrio, Serena Villata.

This work is part of the RAPID CONFIRMA (COntre argumentation contre les Fausses InfoRMAtion) DGA project aiming to automatically detect fake news and limit their diffusion. For this purpose, a framework will be 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, we will use methods from Natural Language Processing and Argumentation Theory to propose automatically extracted counter-arguments (adapted to target audience) from the existing reference press articles. These arguments allow to attack the false information detected in the fake news. Argument Mining techniques will make it possible to (1) analyse the argumentation in natural language, for example by looking for the argumentative structures, identifying the relations of support or attack between the arguments; (2) locate the data related to specific information (related to fake news) on the Web.

7.4.6. Mining and Reasoning on Legal Documents

Participants: Cristian Cardellino, Milagro Teruel, Serena Villata.

Together with Cristian Cardellino, Fernando Cardellino, Milagro Teruel and Laura Alonso Alemany from Univ. of Cordoba, we have proposed a methodology to improve argument annotation guidelines by exploiting inter-annotator agreement measures. After a first stage of the annotation effort, we have detected problematic issues via an analysis of inter-annotator agreement. We have detected ill-defined concepts, which we have addressed by redefining high-level annotation goals. For other concepts, that are well-delimited but complex, the annotation protocol has been extended and detailed. Moreover, as can be expected, we showed that distinctions where human annotators have less agreement are also those where automatic analyzers perform worse. Thus, the reproducibility of results of Argument Mining systems can be addressed by improving inter-annotator agreement in the training material. Following this methodology, we are enhancing a corpus annotated

with argumentation, available online [0] together with guidelines and analyses of agreement. These analyses can be used to filter performance figures of automated systems, with lower penalties for cases where human annotators agree less. This research is addressed in the context of the EU H2020 MIREL project. The results of this research have been published at LREC [59].

Together with some colleagues from Data61 Queensland (Australia) and Antonino Rotolo (University of Bologna), we proposed a formal framework that can instantiate in agents' dialogues moral/rational criteria, such as the maximin principle, Pareto efficiency, and impartiality, which were used, e.g., by John Rawls' theory or rule utilitarianism. Most ethical systems define how the individuals ought, morally, act being part of a society. The process of elicitation of a moral theory governing the agents in a society requires them to express their own norms with the aim to find a moral theory on which all may agree upon. This research is addressed in the context of the EU H2020 MIREL project. The results of this research have been published at DEON [57].

7.4.7. Argumentation

Participants: Serena Villata, Andrea Tettamanzi.

In collaboration with Mauro Dragoni of FBK and Célia da Costa Pereira of I3S, we have proposed the SMACk System, combining argumentation and aspect-based opinion mining [14].

7.4.8. Agent-Based Recommender Systems

Participants: Amel Ben Othmane, Nhan Le Thanh, Andrea Tettamanzi, Serena Villata.

We have proposed a spatio-temporal extension for our multi-context framework for agent-based recommender systems (CARS), to which we have added representation and algorithms to manage uncertainty, imprecision, and approximate reasoning in time and space [47].

7.4.9. RDF Mining

Participants: Duc Minh Tran, Andrea Tettamanzi.

In collaboration with Dario Malchiodi of the University of Milan and Célia da Costa Pereira of I3S, we have studied the use of a prediction model as a surrogate of a possibilistic score for OWL axioms [38], [37].

In collaboration with Claudia d'Amato of the University of Bari, 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 [52].

[0] https://github.com/PLN-FaMAF/ArgumentMiningECHR

ZENITH Project-Team

7. New Results

7.1. Query Processing

7.1.1. 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 this work, we address the problem of top-k query processing over encrypted data. Top-k queries are important for many applications such as information retrieval, spatial data analysis, temporal databases, graph databases, etc. We consider two cases for top-k query processing over encrypted data: 1) centralized: the encrypted data are stored at a single node of a data center, which is useful if the database can fit at one node; 2) distributed: the encrypted data are partitioned and the partitions are encrypted and distributed across multiple nodes, which is useful if the database is very big.

In [52], we address the distributed case, and propose a system, called SD-TOPK, for top-k query processing over encrypted data distributed across several nodes of the cloud. SD-TOPK comes with a distributed top-k query processing algorithm that is executed in the nodes, and finds a set including the encrypted top-k data items. It also has an efficient filtering algorithm that removes most of the false positives included in the set returned by the top-k query processing algorithm. This filtering is done without needing to decrypt the data in the cloud.

In [51], we propose a complete system, called *BuckTop*, for the centralized case. BuckTop is able to efficiently evaluate top-k queries over encrypted data outsourced to a singe node, without having to decrypt it in that node. It includes a top-k query processing algorithm that works on the encrypted data stored in the cloud node, and returns a set that is proved to contain the encrypted data corresponding to the top-k results. We implemented BuckTop and compared its performance for processing top-k queries over encrypted data with that of the popular threshold algorithm (TA) over original (plaintext) data. The results show the effectiveness of BuckTop for outsourcing sensitive data in the cloud and answering top-k queries.

7.1.2. Privacy Preserving Index for Range Query Processing in the Clouds

Participants: Reza Akbarinia, Esther Pacitti.

During the last decade, a large body of academic work has tackled the problem of outsourcing databases to an untrusted cloud while maintaining both privacy and SQL-like querying functionality (at least partially). Range query is an important kind of query that expresses a bounded restriction over the retrieved records. In the database management systems, these queries are usually answered by using efficient indexes. However, developing privacy preserving indexes for untrusted environments is very challenging.

In [55], we propose a differentially private index to an outsourced encrypted dataset. Efficiency is enabled by using a plaintext index structure to perform range queries. Security relies on both differential privacy (of the index) and semantic security (of the encrypted dataset). Our solution, called PINED-RQ, develops algorithms for building and updating the differentially private index. Compared to state-of-the-art secure index based range query processing approaches, PINED-RQ executes queries in the order of at least one magnitude faster. The security of PINED-RQ is proved and its efficiency is assessed by an extensive experimental validation.

7.1.3. Constellation Queries to Analyze Geometrical Patterns

Participants: Dennis Shasha, Patrick Valduriez.

Constellation queries are useful to analyze geometrical patterns. A geometrical pattern is a set of points with all pairwise distances (or, more generally, relative distances) specified. Finding matches to such patterns, i.e. constellations, has applications to spatial data in seismic, astronomical, and transportation contexts. Finding geometric patterns is a challenging problem as the potential number of sets of elements that compose shapes is exponentially large in the size of the dataset and the pattern. In [53], we propose algorithms to find patterns in large data applications using constellation queries. Our methods combine quadtrees, matrix multiplication, and bucket join processing. Our distributed experiments show that the choice of the composition algorithm (matrix multiplication or nested loops) depends on the freedom introduced in the query geometry through the distance additive factor. Three clearly identified blocks of threshold values guide the choice of the best composition algorithm. Answering complex constellation queries, i.e. isotropic and non-isotropic queries, is challenging because scale factors and stretch factors may take any of an infinite number of values. In [53], we propose practically efficient sequential and distributed algorithms for pure, isotropic, and non-isotropic constellation queries. As far as we know, this is the first work to address isotropic and non-isotropic queries.

7.1.4. Parallel Polyglot Query Processing

Participants: Boyan Kolev, Oleksandra Levchenko, Esther Pacitti, Patrick Valduriez.

The blooming of different cloud data stores has turned polystore systems to a major topic in the nowadays cloud landscape. Especially, as the amount of processed data grows rapidly each year, much attention is being paid on taking advantage of the parallel processing capabilities of the underlying data stores. To provide data federation, a typical polystore solution defines a common data model and query language with translations to API calls or queries to each data store. However, this may lead to losing important querying capabilities. The polyglot approach of the CloudMdsQL query language allows data store native queries to be expressed as inline scripts and combined with regular SQL statements in ad-hoc integration queries. Moreover, efficient optimization techniques, such as bind join, can still take place to improve the performance of selective joins. In [47], we introduce the distributed architecture of the LeanXcale query engine that processes polyglot queries in the CloudMdsQL query language, yet allowing native scripts to be handled in parallel at data store shards, so that efficient and scalable parallel joins take place at the query engine level. The experimental evaluation of the LeanXcale parallel query engine on various join queries illustrates well the performance benefits of exploiting the parallelism of the underlying data management technologies in combination with the high expressivity provided by their scripting/querying frameworks

7.2. Scientific Workflows

7.2.1. In Situ Analysis of Simulation Data

Participants: Vitor Silva, Patrick Valduriez.

In situ analysis and visualization have been used successfully in large-scale computational simulations to visualize scientific data of interest, while data is in memory. Such data are obtained from intermediate (or final) simulation results, and once analyzed are typically stored in raw data files. However, existing in situ data analysis and visualization solutions (e.g. ParaView/Catalyst, VisIt) have limited online query processing and no support for dataflow analysis. The latter is a challenge for exploratory raw data analysis. In the context of the SciDISC associate team with Brazil [38], we propose a solution that integrates dataflow analysis with ParaView Catalyst for performing in-situ data analysis and monitoring dataflow from simulation runs [25].

In [21], we propose a solution (architecture and algorithms), called Armful, to combine the advantages of a dataflow-aware SWMS and raw data file analysis techniques to allow for queries on raw data file elements that are related but reside in separate files. Its main components are a raw data extractor, a provenance gatherer and a query processing interface, which are all dataflow-aware.

An instantiation of Armful is DfAnalyzer [34], a library of components to support online in-situ and in-transit data analysis. DfAnalyzer components are plugged directly in the simulation code of highly optimized parallel applications with negligible overhead. With support of sophisticated online data analysis, scientists get a detailed view of the execution, providing insights to determine when and how to tune parameters or reduce data that does not need to be processed [35]. The source code of the DfAnalyzer implementation for Spark is available on github (github.com/hpcdb/RFA-Spark).

7.2.2. Scheduling of Scientific Workflows in Multisite Cloud

Participants: Esther Pacitti, Patrick Valduriez.

In [30], we consider the problem of efficient scheduling of a large SWf in a multisite cloud, i.e. a cloud with geo-distributed cloud data centers (sites). The reasons for using multiple cloud sites to run a SWf are that data is already distributed , the necessary resources exceed the limits at a single site, or the monetary cost is lower. In a multisite cloud, metadata management has a critical impact on the efficiency of SWf scheduling as it provides a global view of data location and enables task tracking during execution. Thus, it should be readily available to the system at any given time. While it has been shown that efficient metadata handling plays a key role in performance, little research has targeted this issue in multisite cloud. Then we propose to identify and exploit hot metadata (frequently accessed metadata) for efficient SWf scheduling in a multisite cloud, using a distributed approach. We implemented our approach within a scientific workflow management system, which shows that our approach reduces the execution time of highly parallel jobs up to 64% and that of the whole SWfs up to 55%.

7.2.3. Distributed Management of Scientific Workflows for Plant Phenotyping

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

In the last decade, high-throughput phenotyping platforms have allowed acquisition of quantitative data on thousands of plants required for genetic analyses in well-controlled environmental conditions.The seven facilities of Phenome produce 200 terabytes of data annually, which are heterogeneous (images, time courses), multiscale (from the organ to the field) and originate from different sites. Hence, the major problem becomes the automatic analysis of these massive datasets and the ability to reproduce large and complex in-silico experiments.

In [31], we propose a solution (infrastructure) to distribute the computation of scientific workflows on very large grid computing facilities (EGI/France Grilles) to the 3D reconstruction, segmentation and tracking of plant organs. This infrastructure, InfraPhenoGrid, is based on OpenAlea, SciFloware and SON, a set of software and technology developed in the team. We have used this solution in [27] to dissect the genetic and environmental influence of biomass accumulation in complex multi-genotype maize canopies.

7.3. Data Analytics

7.3.1. 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 [36], we consider the problem of finding highly correlated pairs of time series across multiple sliding windows. Doing this efficiently and in parallel could help in applications such as sensor fusion, financial trading, or communications network monitoring, to name a few. We have developed a parallel incremental random vector/sketching approach, called ParCorr, to this problem and compared it with the state-of-the-art nearest neighbor method iSAX. Whereas iSAX achieves 100% recall and precision for Euclidean distance, the sketching approach is, empirically, at least 10 times faster and achieves 95% recall and 100% precision on real and simulated data. For many applications this speedup is worth the minor reduction in recall. Our method scales up to 100 million time series and scales linearly in its expensive steps (but quadratic in the less expensive ones).

In [48], we propose a demonstration of our sketch-based solution to efficiently perform both the parallel indexing of large sets of time series and a similarity search on them. Because our method is approximate, we explore the tradeoff between time and precision. A video showing the dynamics of the demonstration can be found at http://parsketch.gforge.inria.fr/video/parSketchdemo_720p.mov.

7.3.2. Parallel Mining of Maximally Informative k-Itemsets in Data Streams

Participants: Mehdi Zitouni, Reza Akbarinia, Florent Masseglia.

The discovery of informative itemsets is a fundamental building block in data analytics and information retrieval. While the problem has been widely studied, only few solutions scale. This is particularly the case when the dataset is massive, or the length k of the informative itemset to be discovered is high.

In [63], we address the problem of mining maximally informative k-itemsets (miki) in data streams based on joint entropy. We propose PentroS, a highly scalable parallel miki mining algorithm. PentroS renders the mining process of large volumes of incoming data very efficient. It is designed to take into account the continuous aspect of data streams, particularly by reducing the computations of need for updating the miki results after arrival/departure of transactions to/from the sliding window. PentroS has been extensively evaluated using massive real-world data streams. Our experimental results confirm the effectiveness of our proposal which allows excellent throughput with high itemset length.

7.3.3. Spatio-Temporal Data Mining

Participants: Esther Pacitti, Florent Masseglia.

The problem of discovering spatiotemporal sequential patterns affects a broad range of applications. Many initiatives find sequences constrained by space and time. We address in [40] an appealing new challenge for this domain: find tight space-time sequences, i.e., find within the same process: i) frequent sequences constrained in space and time that may not be frequent in the entire dataset and ii) the time interval and space range where these sequences are frequent. The discovery of such patterns along with their constraints may lead to extract valuable knowledge that can remain hidden using traditional methods since their support is extremely low over the entire dataset. Our contribution is a new Spatio-Temporal Sequence Miner (STSM) algorithm to discover tight space-time sequences.

7.4. Machine Learning for High-dimensional Data

7.4.1. Uncertainty in Fine-grained Classification

Participants: Titouan Lorieul, Alexis Joly.

Uncertainty is critical when considering classification problems that involve thousands of domain specific labels. A picture of a plant, for instance, contains only a partial information that is usually not sufficient to determine its scientific name with certainty. We first work on the modelling of such uncertainty in the context of crowdsourcing systems involving experts as well as non expert annotators. We rely on Bayesian inference to learn the annotators' confusion and to optimally assign them new items to be validated. In particular, we work on a non-parametric version of this model allowing to combine annotators' suggestions even when the number of possible labels is undetermined and might change over time [33]. In mirror to this research, we also work on the uncertainty of automatic classifiers, in particular deep convolutional neural networks trained on massive amounts of plant images. We conduct an experimental study aimed at evaluating quantitatively the intrinsic data ambiguity of image-based plant observations [64], and we started working on new methods for estimating the uncertainty of ensembles of deep neural networks by fitting a Dirichlet distribution on the set of their predictions. Besides, we study the use of different taxonomic levels as a source of potential reduction in prediction uncertainties [66].

7.4.2. Species Distribution Modelling based on Citizen Science Data

Participants: Christophe Botella, Alexis Joly.

Species distribution models (SDM) are widely used for ecological research and conservation purposes. Given a set of species occurrence, the aim is to infer its spatial distribution over a given territory. Because of the limited number of occurrences of specimens, this is usually achieved through environmental niche modeling approaches, i.e. by predicting the distribution in the geographic space on the basis of a mathematical representation of their known distribution in environmental space (= realized ecological niche). The environment is in most cases represented by climate data (such as temperature, and precipitation), but other variables such as soil type or land cover can also be used. In [24], we study for the first time the relevance of a species distribution model computed from automatically identified plant observations made by citizens rather than from classical inventories made by experts. The results show that the resulting models have a great potential for the early detection of new invasions. In [65] and [60], we propose a deep learning approach to species distribution modelling in order to improve the predictive effectiveness in the context of massive amount of occurrence data. Non-linear prediction models have been of interest for SDM for more than a decade but our study is the first one bringing empirical evidence that deep, convolutional and multilabel models might participate to resolve the limitations of SDM.

7.4.3. Evaluation of Species Identification and Prediction Algorithms

Participants: Alexis Joly, Hervé Goëau, Christophe Botella, Jean-Christophe Lombardo.

We ran a new edition of the LifeCLEF evaluation campaign [45] with the involvement of 13 research teams worldwide. The main novelties and outcomes of the 2018-th edition are the following:

- **GeoLifeCLEF**: a new challenge [71] dedicated to the location-based prediction of species based on spatial occurrences and environmental data tensors. The evaluation concludes that deep environmental convolutional neural networks perform better than spatial models or ponctual environmental models.
- **Man vs. Machine plant identification**: To evaluate how far automated identification systems are from the best possible performance, we organize a challenge involving 19 deep-learning systems implemented by 4 different research teams and 9 of the best expert botanists of the French flora. The main outcome of this work is that the performance of state-of-the-art deep learning models is now very close to the most advanced human expertise.
- **Bird sounds identification**: the 2018-the edition of the BirdCLEF challenge reveals impressive identification performance when considering bird sounds recorded by the Xeno-Canto community. Identifying birds in raw, multi-directional soundscapes, however, remains a very challenging task.

7.4.4. Towards the Recognition of The World's Flora: When HPC Meets Deep Learning

Participants: Hervé Goëau, Jean-Christophe Lombardo, Alexis Joly.

Automated identification of plants and animals have improved considerably in the last few years, in particular thanks to the recent advances in deep learning. In 2017, a challenge on 10,000 plant species (PlantCLEF) resulted in impressive performances with accuracy values reaching 90%. One of the most popular plant identification application, Pl@ntNet, nowadays works on 18K plant species. It accounts for million of users all over the world and already has a strong societal impact in several domains including education, landscape management and agriculture. Now, the big challenge is to train such systems at the scale of the world's biodiversity. Therefore, we built a training set of about 12M images illustrating 300K species of plants. Training a convolutional neural network on such a large dataset can take up to several months on a single node equipped with four recent GPUs. Moreover, to select the best performing architecture and optimize the hyper-parameters, it is often necessary to train several of such networks. Overall, this becomes a highly intensive computational task that has to be distributed on large HPC infrastructures. Therefore, we experiment two french national supercomputers through an access offered by GENCI (Occigen@CINES, a 3.5 Pflop/s Tier-1 cluster based on Broadwell-14cores@2.6Ghz nodes and Joliot-Curie»@TGCC, a BULL-Sequana-X1000 cluster integrating 1656 nodes Intel Skylake8168-24cores@2.7GHz). To implement the synchronized stochastic gradient descent on the CPU cluster Joliot-Curie, we are using the deep learning framework Intel CAFFE coupled with Intel MLSL library (in the context of a collaboration with Intel).

7.4.5. Evaluation of Music Separation Techniques

Participants: Antoine Liutkus, Fabian-Robert Stöter.

After the groundbreaking advent of deep learning, we feel the music processing community needs to step back and think about what had been accomplished and what remains challenging in the problems of musical signal processing and filtering. Therefore, we give a complete overview of the state of the art in music demixing in [32] comprising more than 350 references, as well as two chapters in dedicated books [68], [67]. These references may be considered as complete overviews of the state of the art in music demixing. Furthermore, we introduce the topic to non-expert researchers and engineers in [26].

Apart from this effort in presenting the most recent advances in music processing to the community, we organize yearly a systematic evaluation of state of the art. We report the results of the 2018 Signal Separation Evaluation Campaign in [58], gathering a record number of participants. A perceptual evaluation of the results obtained through this campaign is presented in [59], in collaboration with researchers from the Surrey University.

7.4.6. Robust Probabilistic Models for Time-series

Participants: Antoine Liutkus, Fabian-Robert Stöter.

Processing large amounts of data for denoising or analysis comes with the need to devise models that are robust to outliers and that permit efficient inference. For this purpose, we advocate the use of non-Gaussian models for this purpose, which are less sensitive to data-uncertainty. Most of our effort on this topic is split in two subtasks.

First, we develop new filtering methods that go beyond least-squares estimation. In collaboration with researchers from RWTH, Aachen, Germany, we introduce a new model based on mixtures of Gaussians for filtering in [50]. It combines tractability with a better account of phase consistency for complex data. Along with researchers from IRISA, Rennes and Telecom ParisTech, we also work on filtering α-stable processes [44], [46], [57], which enjoy important applications in robust signal processing.

Second, we work on large amounts of musical archives. This includes an original way to scale up interference reduction in live musical recordings in collaboration with the managers of the Montreux Jazz Festival data at EPFL (Switzerland).

ALICE Project-Team

7. New Results

7.1. Hex-dominant meshing: Mind the gap!

Participants: Nicolas Ray, Dmitry Sokolov, Maxence Reberol, Franck Ledoux, Bruno Lévy.

We proposed a robust pipeline that can generate hex-dominant meshes from any global parameterization of a tetrahedral mesh (Figure 1). We focus on robustness in order to be able to benchmark different parameterizations on a large database. Our main contribution is a new method that integrates the hexahedra (extracted from the parameterization) into the original object. The main difficulty is to produce the boundary of the result, composed of both faces of hexahedra and tetrahedra. Obviously, this surface must be a good approximation of the original object but, more importantly, it must be possible to remesh the volume bounded by this surface minus the extracted hexahedra (called void). We enforce these properties by carefully tracking and eliminating all possibilities of failure at each step of our pipeline.

We tested our method on a large collection of objects (200+) with different settings. In most cases, we obtained results of very good quality as compared to the state-of-the-art solutions.

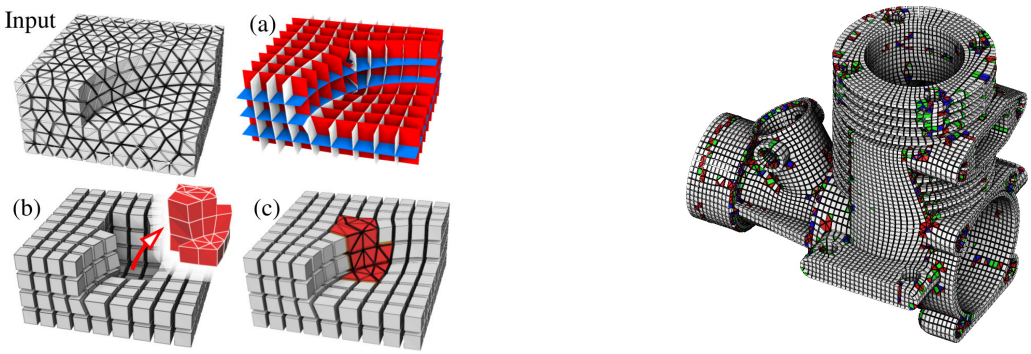


Figure 1. Our hexahedral-dominant meshing procedure: Start from an input tetrahedral mesh. Compute a global parameterization ***(a)****. Extract hexahedra by contouring the isovalues of the parameterization. Isolate the boundary of the void (in red), i.e., the volume with a degenerate / singular parameterization* ***(b)*** *(also called "gap" or "cavity"), shown in red. Remesh the void and stitch it into the hexahedral mesh* ***(c)****.*

7.2. Meshless Voronoi on the GPU

Participants: Nicolas Ray, Dmitry Sokolov, Sylvain Lefebvre, Bruno Lévy.

We proposed a GPU algorithm that computes a 3D Voronoi diagram (Figure 2). Our algorithm is tailored for applications that solely make use of the geometry of the Voronoi cells, such as Lloyd's relaxation used in meshing, or some numerical schemes used in fluid simulations and astrophysics. Since these applications only require the geometry of the Voronoi cells, they do not need the combinatorial mesh data structure computed by the classical algorithms (Bowyer-Watson). Thus, by exploiting the specific spatial distribution of the point-sets used in this type of applications, our algorithm computes each cell independently, in parallel, based on its nearest neighbors. In addition, we show how to compute integrals over the Voronoi cells by decomposing them on the fly into tetrahedra, without needing to compute any combinatorial information. The advantages

of our algorithm is that it is fast, very simple to implement, has constant memory usage per thread and does not need any synchronization primitive. These specificities make it particularly efficient on the GPU: it gains one order of magnitude as compared to the fastest state-of-the-art multicore CPU implementations. To ease the reproducibility of our results, the full documented source code is included in the supplemental material.

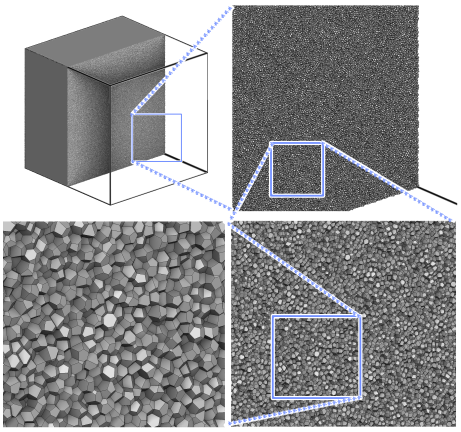


Figure 2. The 3D Voronoi diagram of 10 million points computed on the GPU in 800 ms (NVidia V100). We do not compute the tetrahedra, but in terms of equivalent computation speed, this corresponds to 84 million Delaunay tetrahedra per second.

7.3. Computational Optimal Transport

Participants: Bruno Lévy, Erica Schwindt.

We continued working on Optimal Transportation and its applications in fluid simulation and astrophysics [21], [20]. We developed an efficient and robust algorithm to compute Laguerre diagrams and intersections with tetrahedralized domains, that is, the geometric structure involved in a specific form of optimal transport that we are interested in. In addition, we developed an efficient parallel algorithm to compute Laguerre diagrams, with the possibility of handling periodic boundaries (3-torus), that is to say that the domain is a unit cube with opposite faces that are identified (if one leaves the domain from the left, it enters the domain from the right, etc..., like in the PacMan game). Such a topology is interesting for some simulations in astrophysics, or in material science, that consider a huge domain with homogeneous behavior and replace it with a tiny fraction and periodic boundary conditions (equivalent to a periodic material). We made the algorithms available in the geogram programming library (http://alice.loria.fr/software/geogram/doc/html/index.html). In cooperation with Roya Mohayaee (Institut d'Astrophysique de Paris) and Jean-Michel Alimi (Observatoire de Paris), we started applying the method to some inverse problems in astrophysics (Early Universe Reconstruction), that is reconstructing the past history of the universe from a 3D map of the galaxy clusters. Under some simplifying assumptions, the problem is precisely an instance of semi-discrete optimal transport that our algorithm solves efficiently. Our algorithm does the computation on a desktop PC within hours for several tenths of million points. With Quentin Merigot and Hugo Leclerc (U. Paris Sud), we are designing a new algorithm with the aim of scaling up to billions points (as requested by our astrophysicist colleagues).

AVIZ Project-Team

6. New Results

6.1. Declarative Rendering Model for Multiclass Density Maps

Participants: Jaemin Jo [Dept. of Computer Science and Engineering, Seoul National University, South Korea], Pierre Dragicevic, Jean-Daniel Fekete [correspondent].

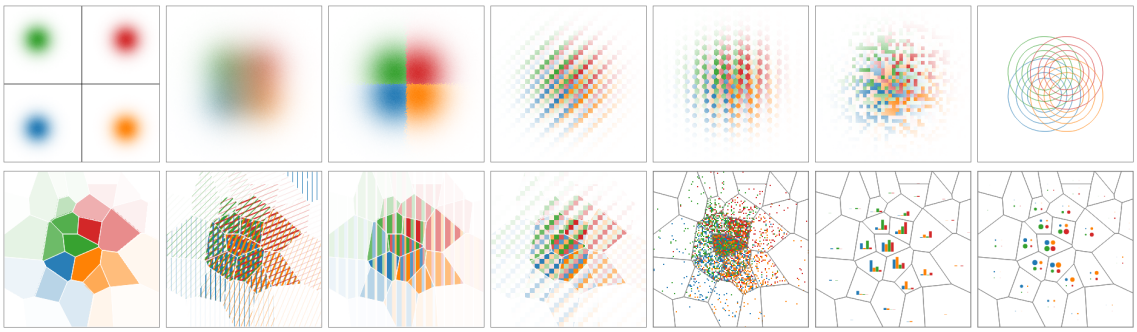


Figure 3. Design alternatives for a four-class density map.

Multiclass maps are scatterplots, multidimensional projections, or thematic geographic maps where data points have a categorical attribute in addition to two quantitative attributes. This categorical attribute is often rendered using shape or color, which does not scale when overplotting occurs. When the number of data points increases, multiclass maps must resort to data aggregation to remain readable. We use a novel model called *multiclass density maps*: multiple 2D histograms computed for each of the category values. Multiclass density maps are meant as a building block to improve the expressiveness and scalability of multiclass map visualization. This library implements our declarative model: a simple yet expressive JSON grammar associated with visual semantics, that specifies a wide design space of visualizations for multiclass density maps. Our declarative model is expressive and can be efficiently implemented in visualization front-ends such as modern web browsers. Furthermore, it can be reconfigured dynamically to support data exploration tasks without recomputing the raw data. Finally, we demonstrate how our model can be used to reproduce examples from the past and support exploring data at scale.

More on the project page: Multiclass Density Maps.

6.2. Reducing Affective Responses to Surgical Images through Color Manipulation and Stylization

Participants: Lonni Besançon [Linköping University Norrköping, Sweden], Amir Semmo [Hasso Plattner Institute, University of Potsdam, Germany], David Biau [Assistance Publique – Hôpitaux de Paris, France], Bruno Frachet [Assistance Publique – Hôpitaux de Paris, France], Virginie Pineau [Institut Curie, France], El Hadi Sariali [Assistance Publique – Hôpitaux de Paris, France], Rabah Taouachi [Institut Curie, France], Tobias Isenberg, Pierre Dragicevic [correspondant].

Figure 4. One of the surgery filters used in our study.

We presented the first empirical study on using color manipulation and stylization to make surgery images more palatable [38]. While aversion to such images is natural, it limits many people's ability to satisfy their curiosity, educate themselves, and make informed decisions. We selected a diverse set of image processing techniques, and tested them both on surgeons and lay people. While many artistic methods were found unusable by surgeons, edge-preserving image smoothing gave good results both in terms of preserving information (as judged by surgeons) and reducing repulsiveness (as judged by lay people). Color manipulation turned out to be not as effective.

This study is an initial investigation but opens up exciting avenues for future research. These include supporting surgery videos, other types of medical images than open surgery (e.g., skin diseases), as well as disturbing imagery outside the medical domain, such as offensive user-generated content that can psychologically impact professionals who monitor it.

All supplemental material is on the OSF page: osf.io/4pfes/.

6.3. Conceptual and Methodological Issues in Evaluating Multidimensional Visualizations for Decision Support

Participants: Evanthia Dimara [ISIR, Sorbonne Université, France], Anastasia Bezerianos [ISIR, Sorbonne Université, France], Pierre Dragicevic [correspondant].

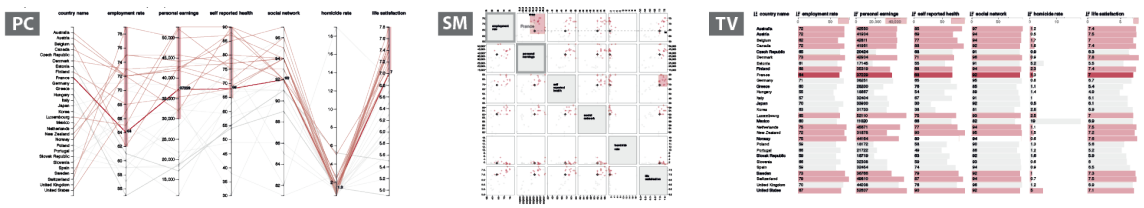


Figure 5. The three visualization techniques tested in our study.

We explored how to rigorously evaluate multidimensional visualizations for their ability to support decision making [22]. We first defined multi-attribute choice tasks, a type of decision task commonly performed with such visualizations. We then identified which of the existing multidimensional visualizations are compatible with such tasks, and evaluated three elementary visualizations: parallel coordinates, scatterplot matrices and tabular visualizations. Our method consisted in first giving participants low-level analytic tasks, in order to ensure that they properly understood the visualizations and their interactions. Participants were then given multi-attribute choice tasks consisting of choosing holiday packages. We assessed decision support through multiple objective and subjective metrics, including a decision accuracy metric based on the consistency between the choice made and self-reported preferences for attributes. We found the three visualizations to be comparable on most metrics, with a slight advantage for tabular visualizations. In particular, tabular visualizations allowed participants to reach decisions faster. Thus, although decision time is typically not central in assessing decision support, it can be used as a tie-breaker when visualizations achieve similar decision accuracy. Our results also suggest that indirect methods for assessing choice confidence may allow to better distinguish between visualizations than direct ones.

All supplemental material is on the project web page: aviz.fr/dm.

6.4. Blinded with Science or Informed by Charts? A Replication Study

Participants: Pierre Dragicevic [correspondant], Yvonne Jansen [ISIR, Sorbonne Université, France].

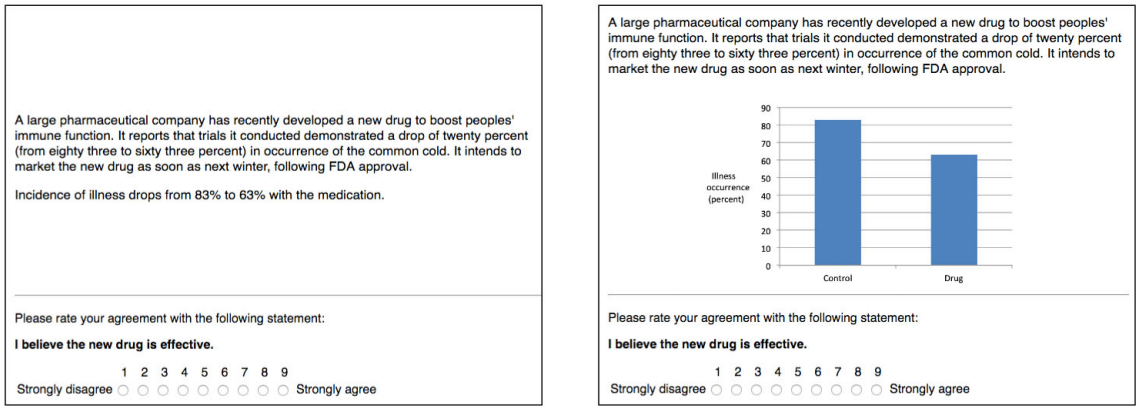


Figure 6. a) text without chart, b) text with "trivial" chart.

We provided a reappraisal of Tal and Wansink's study "Blinded with Science", where seemingly trivial charts were shown to increase belief in drug efficacy, presumably because charts are associated with science. Through a series of four replications conducted on two crowdsourcing platforms, we investigated an alternative explanation, namely, that the charts allowed participants to better assess the drug's efficacy [24]. Considered together, our experiments suggested that the chart seems to have indeed promoted understanding, although the effect is likely very small. Meanwhile, we were unable to replicate the original study's findings, as text with chart appeared to be no more persuasive – and sometimes less persuasive – than text alone. This suggests that the effect may not be as robust as claimed and may need specific conditions to be reproduced. Regardless, within our experimental settings and considering our study as a whole (N = 623), the chart's contribution to understanding was clearly larger than its contribution to persuasion.

The main lesson from our study is that with charts, the peripheral route of persuasion cannot be studied independently from the central route: in order to establish that a chart biases judgment, it is necessary to also rigorously establish that it does not aid comprehension. Our replication also opens many relevant questions for infovis. Are charts really associated with science? More generally, what associations do charts or visualizations trigger depending on their visual design? When exactly is a chart trivial?

All supplemental material is on the project web page: aviz.fr/blinded.

6.5. A Model of Spatial Directness in Interactive Visualization

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

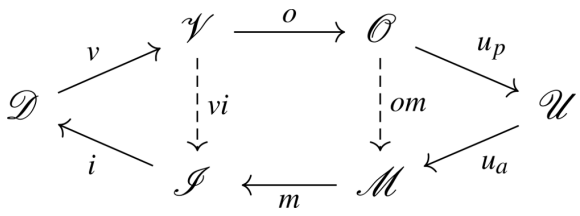


Figure 7. Illustration of the model of spatial directness.

We discussed the concept of directness in the context of spatial interaction with visualization. In particular, we proposeed a model (see Figure 7) 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 distinguished between different types of directness. 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 showed how to apply it to several real-world interaction scenarios in visualization, and thus discussed the resulting types of spatial directness, without recommending either more direct or more indirect interaction techniques. In particular, we demonstrated descriptive and evaluative usage of the proposed model, and also briefly discussed its generative usage.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Bruckner2018MSD.

6.6. Multiscale Visualization and Scale-Adaptive Modification of DNA Nanostructures

Participants: Haichao Miao [TU Wien, Austria, and Austrian Institute of Technology, Vienna, Austria], Elisa de Llano [Austrian Institute of Technology, Vienna, Austria], Johannes Sorger [Complexity Science Hub Vienna, Austria], Yasaman Ahmadi [Austrian Institute of Technology, Vienna, Austria], Tadija Kekic [Austrian Institute of Technology, Vienna, Austria], Tobias Isenberg [correspondant], M. Eduard Gröller [TU Wien, Austria], Ivan Barišic [Austrian Institute of Technology, Vienna, Austria], Ivan Viola [TU Wien, Austria and KAUST, Kingdom of Saudi Arabia].

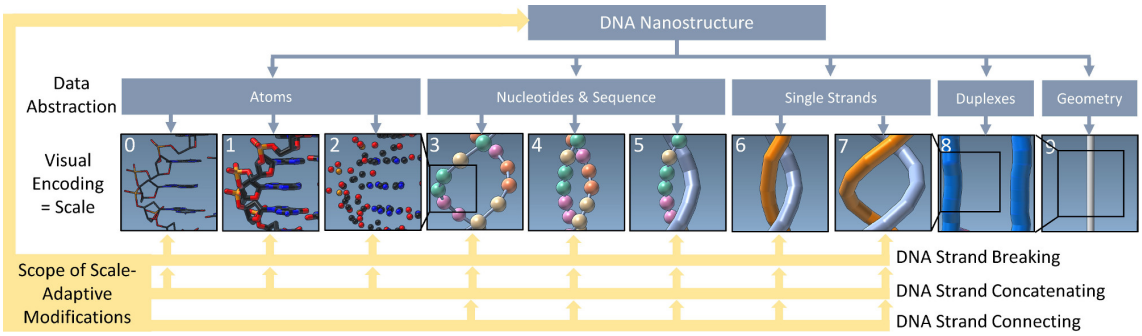


Figure 8. Illustration of the abstraction space.

We presented an approach to represent DNA nanostructures in varying forms of semantic abstraction, describe ways to smoothly transition between them, and thus create a continuous multiscale visualization and interaction space for applications in DNA nanotechnology. This new way of observing, interacting with, and creating DNA nanostructures enables domain experts to approach their work in any of the semantic abstraction levels, supporting both low-level manipulations and high-level visualization and modifications. Our approach allows them to deal with the increasingly complex DNA objects that they are designing, to improve their features, and to add novel functions in a way that no existing single-scale approach offers today. For this purpose we collaborated with DNA nanotechnology experts to design a set of ten semantic scales (see Figure 8). These scales take the DNA's chemical and structural behavior into account and depict it from atoms to the targeted architecture with increasing levels of abstraction. To create coherence between the discrete scales, we seamlessly transition between them in a well-defined manner. We used special encodings to allow experts to estimate the nanoscale object's stability. We also added scale-adaptive interactions that facilitate the intuitive modification of complex structures at multiple scales. We demonstrate the applicability of our approach on an experimental use case. Moreover, feedback from our collaborating domain experts confirmed an increased time efficiency and certainty for analysis and modification tasks on complex DNA structures. Our method thus offers exciting new opportunities with promising applications in medicine and biotechnology.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Miao2018MVS.

6.7. DimSUM: Dimension and Scale Unifying Maps for Visual Abstraction of DNA Origami Structures

Participants: Haichao Miao [TU Wien, Austria, and Austrian Institute of Technology, Vienna, Austria], Elisa de Llano [Austrian Institute of Technology, Vienna, Austria], Tobias Isenberg [correspondant], M. Eduard Gröller [TU Wien, Austria], Ivan Barišic [Austrian Institute of Technology, Vienna, Austria], Ivan Viola [TU Wien, Austria and KAUST, Kingdom of Saudi Arabia].

We presented a novel visualization concept for DNA origami structures that integrates a multitude of representations into a DimSUM. This novel abstraction map (see Figure 9) provides means to analyze, smoothly transition between, and interact with many visual representations of the DNA origami structures in an effective way that was not possible before. DNA origami structures are nanoscale objects, which are challenging to model in silico. In our holistic approach we seamlessly combined three-dimensional realistic shape models, two-dimensional diagrammatic representations, and ordered alignments in one-dimensional arrangements, with semantic transitions across many scales. To navigate through this large, two-dimensional abstraction map we highlighted locations that users frequently visit for certain tasks and datasets. Particularly interesting viewpoints can be explicitly saved to optimize the workflow. We have developed DimSUM together

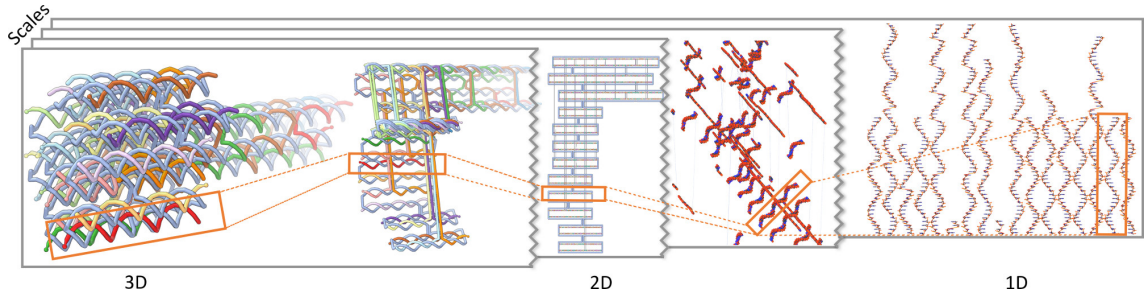


Figure 9. Illustration of the DimSUM space.

with domain scientists specialized in DNA nanotechnology. In the paper we discussed our design decisions for both the visualization and the interaction techniques. We demonstrateed two practical use cases in which our approach increases the specialists' understanding and improves their effectiveness in the analysis. Finally, we discussed the implications of our concept for the use of controlled abstraction in visualization in general.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Miao2018DDS.

6.8. Pondering the Concept of Abstraction in (Illustrative) Visualization

Participants: Ivan Viola [TU Wien, Austria and KAUST, Kingdom of Saudi Arabia], Tobias Isenberg [correspondant].

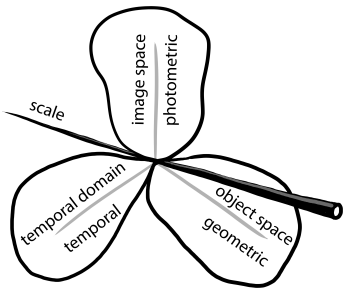


Figure 10. Illustration of the abstraction concept.

We discussed the concept of directness in the context of spatial interaction with visualization (Figure 10). In particular, we proposed a model (autoreffig:directness) 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 distinguished between different types of directness. 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 showed how to apply it to several real-world interaction scenarios in visualization, and thus discussed the resulting types of spatial directness, without recommending either more direct or more indirect interaction techniques. In particular, we demonstrated descriptive and evaluative usage of the proposed model, and also briefly discussed its generative usage.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Bruckner2018MSD.

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

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

Those of us who study large effects may believe ourselves to be unaffected by the reproducibility problems that plague other areas. 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.

6.10. Visualizing Ranges over Time on Mobile Phones: A Task-Based Crowdsourced Evaluation

Participants: Matthew Brehmer [Microsoft Research, USA], Bongshin Lee [Microsoft Research, USA], Petra Isenberg [correspondant], Eun Kyoung Choe [University of Maryland, USA].

In the first crowdsourced visualization experiment conducted exclusively on mobile phones, we experimentally compare approaches to visualizing ranges over time on small displays. People routinely consume such data via a mobile phone, from temperatures in weather forecasting apps to sleep and blood pressure readings in personal health apps. However, we lack guidance on how to effectively visualize ranges on small displays in the context of different value retrieval and comparison tasks, or with respect to different data characteristics such as periodicity, seasonality, or the cardinality of ranges. Central to our experiment is a comparison between two ways to lay out ranges: a more conventional linear layout strikes a balance between quantitative and chronological scale resolution, while a less conventional radial layout emphasizes the cyclicality of time and may prioritize discrimination between values at its periphery. With results from 87 crowd workers, we found that while participants completed tasks more quickly with linear layouts than with radial ones, there were few differences in terms of error rate between layout conditions. We also found that participants performed similarly with both layouts in tasks that involved comparing superimposed observed and average ranges.

More on the project Web page.

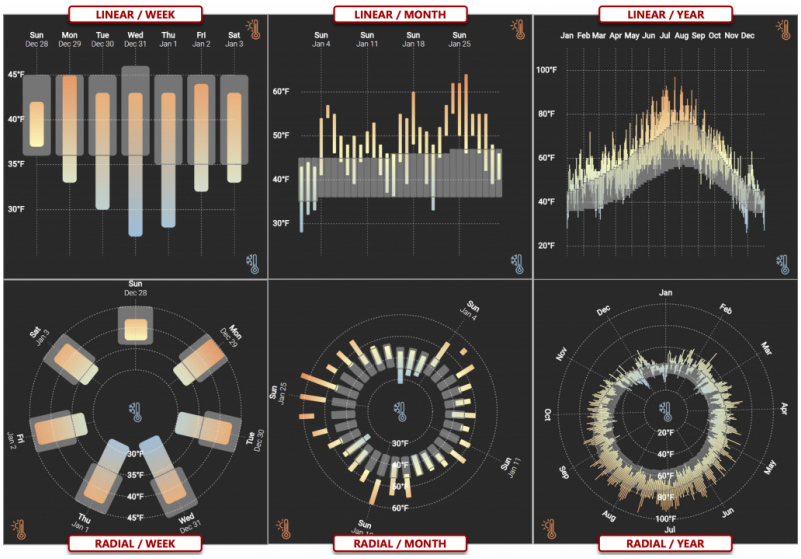


Figure 11. Linear and Radial temperature range charts designed for mobile phone displays, representative of the stimuli used in our crowdsourced experiment. The gradient bars encode observed temperature ranges and are superimposed on gray bars encoding average temperature ranges. Corresponding Week, Month, and Year charts display the same data.

EX-SITU Project-Team

7. New Results

7.1. Fundamentals of Interaction

Participants: Michel Beaudouin-Lafon [correspondant], Wendy Mackay, Cédric Fleury, Theophanis Tsandilas, Dimitrios Christaras Papageorgiou, Han Han, Germán Leiva, Nolwenn Maudet, Yujiro Okuya, Miguel Renom, Philip Tchernavskij, Andrew Webb.

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 reanalyze existing theories and methodologies in order to gain new insights.

Continuing our long-standing exploration of Fitts' law, we demonstrated the dangers of confounding factors in Fitts'-like experimental designs and recommended how to avoid them [20]. Confounds come from the fact that traditional Fitts'-like experiments use geometric progressions of the two main factors (target distance D and amplitude W) and aggregate data points per $ID = log(1 + D/W)$. This typically leads to a strong confound between D and ID, whereby an effect attributed to ID may in fact be due solely to D. We showed evidence of published results where this confound led to the misinterpretation of experimental results, and proposed stochastic sampling of D and W as a technique to avoid such problems.

We also reviewed statistical methods for the analysis of user-elicited gestural vocabularies [16] and argued that current statistics for assessing agreement across participants are problematic. First, we showed that raw agreement rates disregard agreement that occurs by chance and do not reliably capture how participants distinguish among referents. Second, we explained why current recommendations on how to interpret agreement scores rely on incorrect assumptions. Third, we demonstrated that significance tests for comparing agreement rates, either within or between participants, yield large Type I error rates ($> 40\%$ for $\alpha = .05$). As alternatives, we presented agreement indices that are routinely used in inter-rater reliability studies. We discussed how to apply them to gesture elicitation studies. We also demonstrated how to use common resampling techniques to support statistical inference with interval estimates. We applied these methods to reanalyze and reinterpret the findings of four gesture elicitation studies. We also participated in an invited formal debate at ACM/CHI 2018 to discuss the issue of replicability in HCI experiments, specifically whether or not the community should adopt the TOP (Transparency and Openness) guidelines for data and code transparency, citation, experiment preregistration and replication of experiments.

In order to explore novel forms of interaction based on the concepts of *interaction instruments* and *interactive substrates*, we conducted several studies and developed prototypes in three main areas:

First, we challenged the notion of application as the main organizing principle of digital environments. Most of our current interactions with the digital world are mediated by applications that impose artificial limits on collaboration among users and distribution across devices, and the constantly changing procedures that disrupt everyday use. These limitations are due partly to the engineering principles of encapsulation and program-data separation, which highlight the needs for appropriate conceptual models of interaction [18]. We proposed new architectural principles [28], [17] that address these issues by considering interactions as first-class objects that can be dynamically created, added to and removed from an interactive system.

Second, we addressed the needs of designers and developers of interactive systems through a series of studies and prototypes. Current prototyping tools do not adequately support the early stages of design, nor the necessary communication between designers and developers. We created and evaluated VideoClipper and Montage [21], two tools that facilitate video prototyping for the early sketching of ideas. VideoClipper facilitates the planning and capturing of video brainstorming ideas and video prototypes, while Montage (fig. 2) uses chroma-keying to create more advanced video prototypes and facilitating their reuse in different

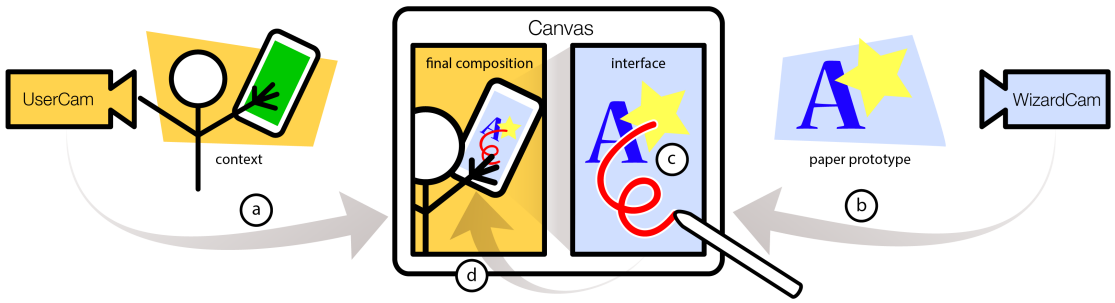


Figure 2. Montage: the UserCam captures the context (a) and the WizardCam captures the paper prototype (b); Both live-stream video to the Canvas, where the designer can add digital sketches (c). Montage replaces the green screen with the interface to create the final composition (d).

contexts. We also created Enact (under submission), a prototyping tool that lets designers and developers work in the same environment to create novel touch-based interaction techniques. Germán Leiva, supervised by Michel Beaudouin-Lafon, successfully defended his Ph.D. thesis *Interactive Prototyping of Interactions: From Throwaway Prototypes to Takeaway Prototyping* [34] on this topic.

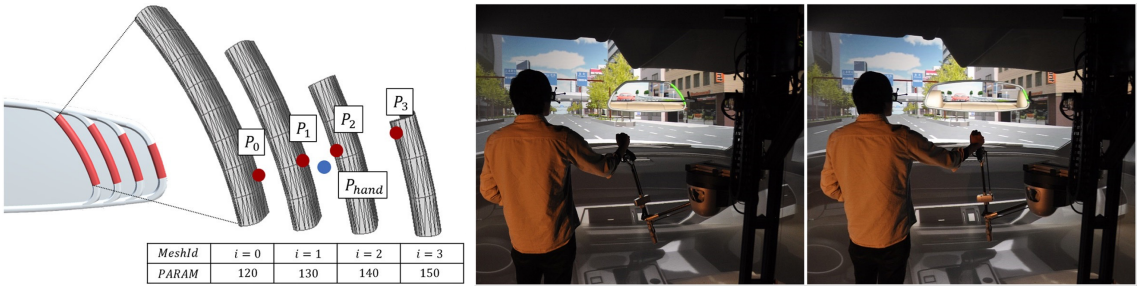


Figure 3. Pre-computed meshes of a rear-view mirror while modifying the right part: the user's hand position (P_{hand}) determines the selected shape (left). A virtual car cockpit where the user modifies the rear-view mirror shape in real time, using haptic force feedback (right).

Third, in the context of Computer Aided Design (CAD), we explored solutions for modifying parametric CAD objects in an immersive virtual reality system. In particular, we developed *ShapeGuide* [14], a technique that lets users modify parameter values by directly pushing or pulling the surface of a CAD object (Figure 3). Including force feedback increases the precision of the users' hand motions in the 3D space. In a controlled experiment, we compared *ShapeGuide* to a standard one-dimensional scroll technique to measure its added value for parametric CAD data modification on a simple industrial object. We also evaluated the effect of force feedback assistance on both techniques. We demonstrated that *ShapeGuide* is significantly faster and more efficient than the scroll technique. In addition, we showed that force feedback assistance enhances the precision of both techniques.

7.2. Human-Computer Partnerships

Participants: Wendy Mackay [correspondant], Baptiste Caramiaux, Téo Sanchez, Marianela Ciolfi Felice, Carla Griggio, Shu Yuan Hsueh, Wanyu Liu, John Maccallum, Nolwenn Maudet, Joanna Mcgrenere, Midas Nouwens, Andrew Webb.

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.

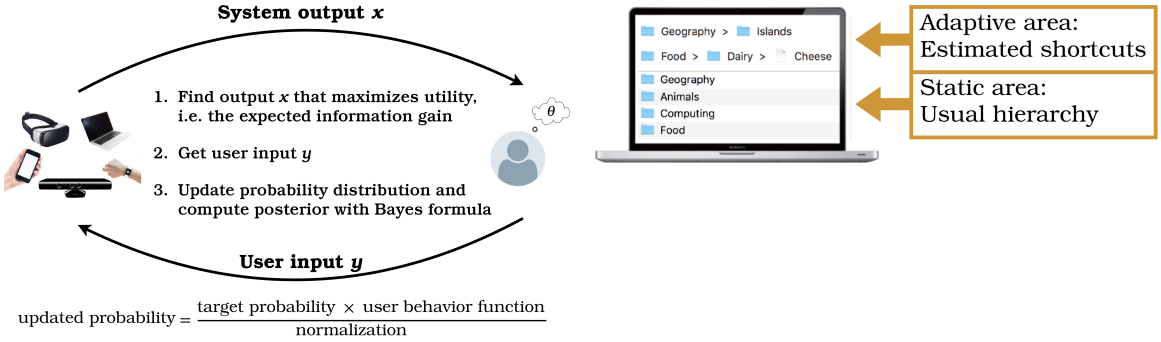


Figure 4. The BIG framework (left) and its application to the BIGFile split-adaptive interface for file navigation (right).

The *Bayesian Information Gain* (BIG) project uses Bayesian Experimental Design, where the criterion is to maximize the information-theoretic concept of mutual information, also known as information gain (fig. 4 -left). The resulting interactive system "runs experiments" on the user in order to maximize the information gain from the user's next input and get to the user's goal more efficiently. *BIGnav* applies BIG to multiscale navigation [7]. Rather than simply executing the navigation commands issued by the user, *BIGnav* interprets them to update its knowledge about the user's intended target, and then computes a new view that maximizes the expected information gain provided by the user's next input. This view is located such that, from the system's perspective, the possible navigation commands are uniformly probable, to the extent possible. *BIGFile* [22] (ACM CHI Honorable Mention award) uses a similar approach for file navigation, with a split interface (fig. 4 -right) that combines a classical area where users can navigate the file system as usual and an adaptive area with a set of shortcuts calculated with BIG. *BIGnav* and *BIGFile* create a novel form of human-computer partnership, where the computer challenges the user in order to extract more information from the user's input, making interaction more efficient. We showed that both techniques are significantly faster (40% and more) than conventional navigation techniques. Wanyu Liu, supervised by Michel Beaudouin-Lafon, successfully defended her Ph.D. thesis *Information theory as a unified tool for understanding and designing human-computer interaction* [35] on this topic.

In the area of visualization, we studied the common challenge faced by domain experts when identifying and comparing patterns in time series data. While automatic measures exist to compute time series similarity, human intervention is often required to visually inspect these automatically generated results. In collaboration with the ILDA Inria team and Univ. Paris-Descartes, we studied how different visualization techniques affect similarity perception in EEG signals [12], [31]. Our goal was to understand if the time series results returned from automatic similarity measures are perceived in a similar manner, irrespective of the visualization technique; and if what people perceive as similar with each visualization aligns with different automatic

measures and their similarity constraints. Overall, our work indicates that the choice of visualization affects which temporal patterns we consider to be similar, i.e., the notion of similarity in a time series is not visualization independent. This demonstrates the need for effective human-computer partnerships in which the computer complements, rather than replaces, human skills and expertise.

We began to explore *human-centred machine learning*, which takes advantage of *active machine learning* to facilitate personalization of an interactive system. We developed a gesture-based recognition system where the user iteratively provides instances and also answers the system's queries. Our results demonstrated the phenomenon of co-adaptation between the human user and the system, which challenges the state of the art in conventional active learning. We further explored interactive reinforcement learning as a way to explore high-dimensional parametric space efficiently [24].

7.3. Creativity

Participants: Sarah Fdili Alaoui [correspondant], Marianela Ciolfi Felice, Carla Griggio, Shu Yuan Hsueh, Germán Leiva, John Maccallum, Wendy Mackay, Baptiste Caramiaux, Nolwenn Maudet, Joanna Mcgrenere, Midas Nouwens, Jean-Philippe Rivière, Nicolas Taffin, Philip Tchernavskij, Theophanis Tsandilas, Andrew Webb, Michael Wessely.

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 identified diverse strategies for recording choreographic fragments and, influenced by the concept of *information substrates*, designed *Knotation* [19], a mobile pen-based tool where choreographers sketch representations of their choreographic ideas and make them interactive (Figure 5). Subsequent studies showed that *Knotation* supports both dance-then-record and record-then-dance strategies. Marianela Ciolfi Felice, supervised by Wendy Mackay and Sarah Fdili Alaoui, successfully defended her Ph.D. thesis *Supporting Expert Creative Practice* on this topic [32].

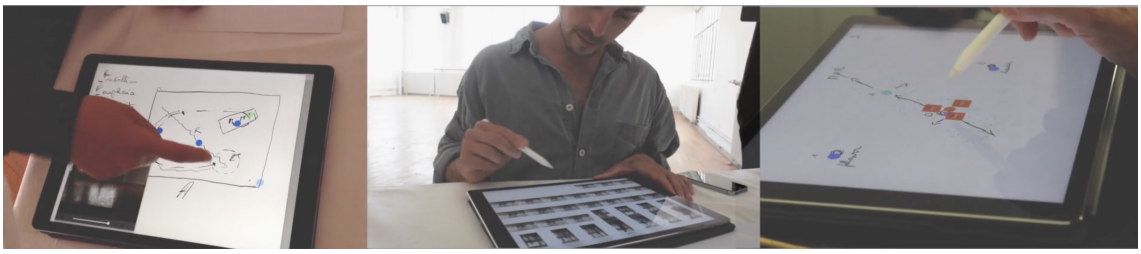


Figure 5. A choreographer uses Knotation to specify and interact with the spatial and temporal layout of a piece.

We are also developing a *Choreographer's Workbench*, a full-body interactive system that helps choreographers explore dance movements by linking previously recorded movement ideas and revealing their underlying relationships. The system emphasizes discoverability and appropriation of movement ideas, using feedforward to visualize movement characteristics. We studied how dancers learn complex expressive movements [23], and studied how variability during practice affects learning motor and timing skills [11]. We contributed to soma-based design, i.e. movement-based designs and design practices specifically engaging with aesthetics [13]. We also collaborated with Ircam on a tool that uses reinforcement learning to explore high-dimensional sound spaces [24]. Users enter likes and dislikes to guide navigation within the sound space, shifting from a parameter-based to a reward-based exploration strategy.

We also are interested in how makers transition between physical and digital designs. Makers often create both physical and digital prototypes to explore a design, taking advantage of the subtle feel of physical materials and the precision and power of digital models. We developed *ShapeMe* [25], a novel smart material that captures its own geometry as it is physically cut by an artist or designer. *ShapeMe* includes a software toolkit that lets its users generate customized, embeddable sensors that can accommodate various object shapes. As the designer works on a physical prototype, the toolkit streams the artist's physical changes to its digital counterpart in a 3D CAD environment (Figure 6). We used a rapid, inexpensive and simple-to-manufacture inkjet printing technique to create embedded sensors. We successfully created a linear predictive model of the sensors' lengths, and our empirical tests of *ShapeMe* showed an average accuracy of 2 to 3 mm. We further presented an application scenario for modeling multi-object constructions, such as architectural models, and 3D models consisting of multiple layers stacked one on top of each other.

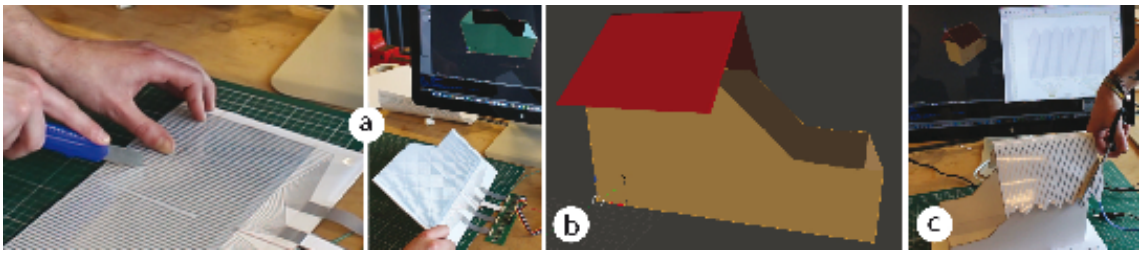


Figure 6. ShapeMe is a novel sensing technology that enables physical modeling with shape-aware material: (a) The maker cuts a foamcore piece to reshape the walls of a house model. The updated shape is captured by a grid of length-aware sensors and is communicated to 3D modeling software. (b) The makers digitally creates the pieces of the roof and then produces its physical model. (c) The maker explores variations of the roof by cutting its side with scissors, while its shape is continuously captured.

We also presented *Interactive Tangrami* [29], a method for prototyping interactive physical interfaces from functional paper-folded building blocks (Tangramis). *Interactive Tangrami* can contain various sensor input and visual output capabilities. Our digital design tool lets makers design the shape and interactive behavior of custom user interfaces. The software manages the communication with the paper-folded blocks and streams the interaction data via the Open Sound protocol (OSC) to an application prototyping environment, such as MaxMSP. The building blocks are fabricated digitally with a rapid and inexpensive ink-jet printing method. Our systems allows to prototype physical user interfaces within minutes and without knowledge of the underlying technologies. Finally, we continued our work with Saarland University, TU Berlin and MIT on digitally fabricated *directional screens* [15]. Michael Wessely, supervised by Theophanis Tsandilas and Wendy Mackay, successfully defended his Ph.D. thesis *Fabricating Malleable Interaction-Aware Material* [36] on these topics.

7.4. Collaboration

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

ExSitu is interested in exploring new ways of supporting collaborative interaction and remote communication. We investigated how large interactive spaces such as wall-sized displays or immersive virtual reality systems can foster collaboration in both co-located and remote situations in the context of Digiscope (http://digiscope.fr/). We also conducted in-depth studies to better understand communication through social networks.

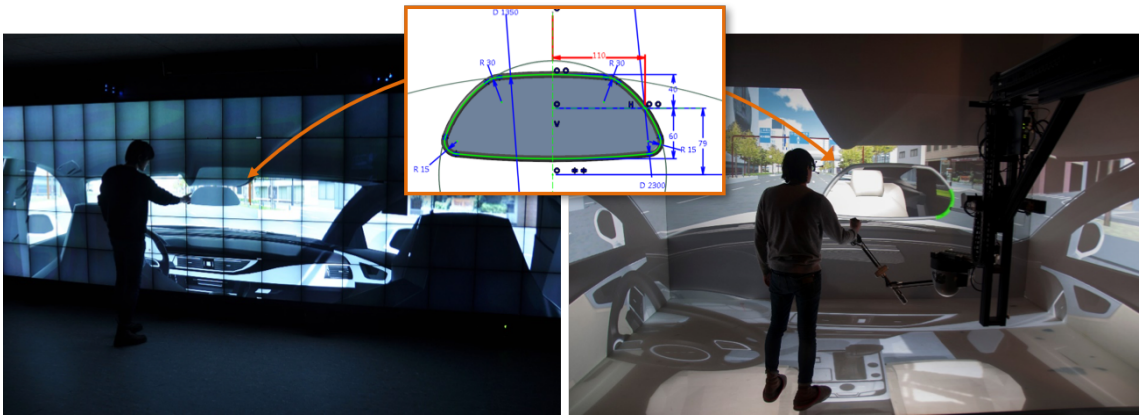


Figure 7. Collaborative CAD data modification between a wall-sized display (left) and a CAVE system (right).

Remote users can have significantly different display and interaction capabilities, such as a wall-size display v.s. an immersive CAVE. We started to explore how such asymmetric interaction capabilities provide interesting opportunities for new collaboration strategies. In particular, we developed a distributed architecture allowing the collaborative modifications of CAD data across heterogeneous platforms [27] and tested it between the EVE and WILDER platforms of Digiscope (CAVE vs. wall-sized touch display – Figure 7).

Remote collaboration across large interactive spaces also requires telepresence systems which support audio-video communication among users as they move in front of the display or inside of the immersive virtual reality system. We have added 3D audio to improve spatial awareness of remote users [26]: 3D audio lets us position a sound source for each remote participant at the virtual position occupied by this participant in the local space. When using video as well as audio, this lets us position the audio feed so that it is congruent with the position of the video feed.

Finally, we conducted an in-depth study of how users communicate via multiple social network apps that offer almost identical functionality. We studied how and why users distribute their contacts within their app ecosystem. We found that users appropriate the features and technical constraints of their apps to create idiosyncratic “communication places”, each with its own recursively defined membership rules, perceived purposes, and emotional connotations. Users also shift the boundaries of their communication places to accommodate changes in their contacts’ behavior, the dynamics of their relationships, and the restrictions of the technology. We argue that communication apps should support creating multiple “communication places” within the same app, relocating conversations across apps, and accessing functionality from other apps. Carla Griggio, supervised by Wendy Mackay, successfully defended her Ph.D. thesis *Designing for Ecosystems of Communication Apps* [33] on this topic.

GRAPHDECO Project-Team

6. New Results

6.1. Computer-Assisted Design with Heterogeneous Representations

6.1.1. 3D Sketching using Multi-View Deep Volumetric Prediction

Participants: Johanna Delanoy, Adrien Bousseau.

Drawing is the most direct way for people to express their visual thoughts. However, while humans are extremely good are perceiving 3D objects from line drawings, this task remains very challenging for computers as many 3D shapes can yield the same drawing. Existing sketch-based 3D modeling systems rely on heuristics to reconstruct simple shapes, require extensive user interaction, or exploit specific drawing techniques and shape priors. Our goal is to lift these restrictions and offer a minimal interface to quickly model general 3D shapes with contour drawings . While our approach can produce approximate 3D shapes from a single drawing, it achieves its full potential once integrated into an interactive modeling system, which allows users to visualize the shape and refine it by drawing from several viewpoints (Figure 4). At the core of our approach is a deep convolutional neural network (CNN) that processes a line drawing to predict occupancy in a voxel grid. The use of deep learning results in a flexible and robust 3D reconstruction engine that allows us to treat sketchy bitmap drawings without requiring complex, hand-crafted optimizations. While similar architectures have been proposed in the computer vision community, our originality is to extend this architecture to a multiview context by training an updater network that iteratively refines the prediction as novel drawings are provided.

This work is a collaboration with Mathieu Aubry from Ecole des Ponts ParisTech and Alexei Efros and Philip Isola from UC Berkeley. The work was published in Proceedings of the ACM on Computer Graphics and Interactive Techniques and presented at the ACM SIGGRAPH I3D Symposium on Interactive Computer Graphics and Games [12].

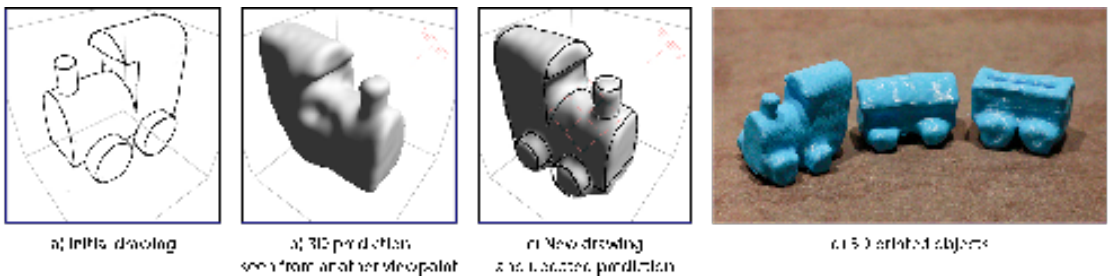


Figure 4. Our sketch-based modeling system can process as little as a single perspective drawing (a) to predict a volumetric object (b). Users can refine this prediction and complete it with novel parts by providing additional drawings from other viewpoints (c). This iterative sketching workflow allows quick 3D concept exploration and rapid prototyping (d).

6.1.2. Procedural Modeling of a Building from a Single Image

Participant: Adrien Bousseau.

Creating a virtual city is demanded for computer games, movies, and urban planning, but it takes a lot of time to create numerous 3D building models. Procedural modeling has become popular in recent years to overcome this issue, but creating a grammar to get a desired output is difficult and time consuming even for expert users. In this paper, we present an interactive tool that allows users to automatically generate such a grammar from a single image of a building. The user selects a photograph and highlights the silhouette of the target building as input to our method. Our pipeline automatically generates the building components, from large-scale building mass to fine-scale windows and doors geometry. Each stage of our pipeline combines convolutional neural networks (CNNs) and optimization to select and parameterize procedural grammars that reproduce the building elements of the picture. In the first stage, our method jointly estimates camera parameters and building mass shape. Once known, the building mass enables the rectification of the facades, which are given as input to the second stage that recovers the facade layout. This layout allows us to extract individual windows and doors that are subsequently fed to the last stage of the pipeline that selects procedural grammars for windows and doors. Finally, the grammars are combined to generate a complete procedural building as output. We devise a common methodology to make each stage of this pipeline tractable. This methodology consists in simplifying the input image to match the visual appearance of synthetic training data, and in using optimization to refine the parameters estimated by CNNs. We used our method to generate a variety of procedural models of buildings from existing photographs.

The work was published in Computer Graphics Forum, presented at Eurographics 2018 [15].

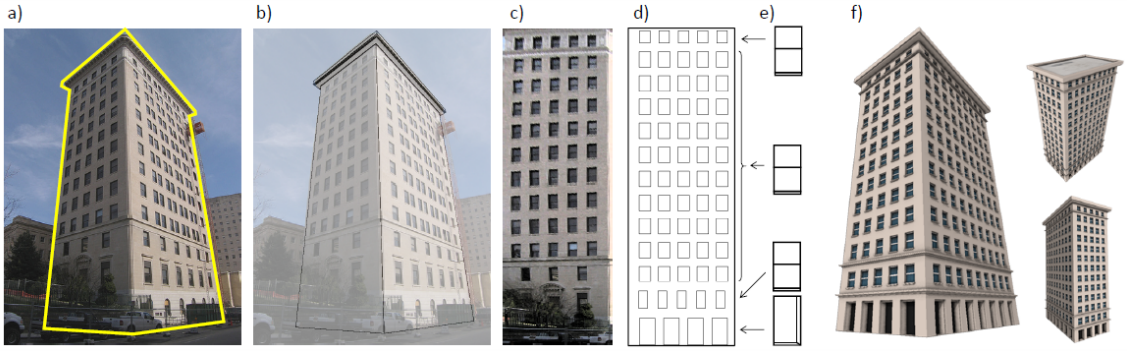


Figure 5. (a) Given an image and a silhouette of a building, (b) our approach automatically estimates the camera parameters and generates a building mass grammar as a first step. Then, (c) the facade image is rectified, and (d) the facade grammar is generated. (e) For each window non-terminal, the best window grammar is selected by maximum vote. (f) Finally the output grammar is constructed and a corresponding 3D geometry is generated.

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

Participants: Yulia Gryaditskaya, Frédéric Durand, Adrien Bousseau.

We collected a dataset of more than 400 product design sketches, representing 12 man-made objects drawn from two different view points by 7 to 15 product designers of varying expertise. Together with industrial design teachers, we distilled a taxonomy of the methods designers use to accurately sketch in perspective and used it to label each stroke of the 214 sketches drawn from one of the two viewpoints. We registered each sketch to its reference 3D model by annotating sparse correspondences. We made an analysis of our annotated sketches, which reveals systematic drawing strategies over time and shapes. We also developed several applications of our dataset for sketch-based modeling and sketch filtering. We will distribute our dataset under the Creative Commons CC0 license to foster research in digital sketching.

This work is a collaboration with Mark Sypesteyn, Jan Willem Hoftijzer and Sylvia Pont from TU Delft, Netherlands. It is currently under review.

6.1.4. Line Drawing Vectorization using a Global Parameterization

Participants: Tibor Stanko, Adrien Bousseau.

Despite the progress made in recent years, automatic vectorization of line drawings remains a difficult task. For drawings containing noise, holes and oversketched strokes, the main challenges are the correct classification of curve junctions, filling the missing information, and clustering multiple strokes corresponding to a single curve. We propose a new line drawing vectorization method, which addresses the above challenges in a global manner. Inspired by the quad meshing literature, we compute a global parametrization of the input drawing, such that nearby strokes are mapped to a single straight line in the parametric domain, while junctions are mapped to straight line intersections. The vectorization is obtained by following the straight lines in the parametric domain, and mapping them back to the original space. This allows us to process both clean and sketchy drawings.

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

6.1.5. Image-Space Motion Rigidification for Video Stylization

Participants: Johanna Delanoy, Adrien Bousseau.

Existing video stylization methods often retain the 3D motion of the original video, making the result look like a 3D scene covered in paint rather than the 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. 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 is currently under review.

6.1.6. Computational Design of Tensile Structures

Participants: David Jourdan, Adrien Bousseau.

Tensile structures are architectural shapes made of stretched elastic material that can be used to create large-span roofs. Their elastic properties make it quite challenging to obtain a specific shape, and the final shape of a tensile structure is usually found rather than imposed. We created a design tool for tensile structures that, unlike existing software, lets the user specify the shape they want and finds the closest fit.

This work is an ongoing collaboration with Melina Skouras from IMAGINE (Inria Rhone Alpes). A preliminary version was presented at JFIG (Journées Françaises d'Informatique Graphique) 2018.

6.2. Graphics with Uncertainty and Heterogeneous Content

6.2.1. Single-Image SVBRDF Capture with a Rendering-Aware Deep Network

Participants: Valentin Deschaintre, Aittala Miika, Frédéric Durand, George Drettakis, Adrien Bousseau.

Texture, highlights, and shading are some of many visual cues that allow humans to perceive material appearance in single pictures. Yet, recovering spatially-varying bi-directional reflectance distribution functions (SVBRDFs) from a single image based on such cues has challenged researchers in computer graphics for decades. We tackle lightweight appearance capture by training a deep neural network to automatically extract and make sense of these visual cues. Once trained, our network is capable of recovering per-pixel normal, diffuse albedo, specular albedo and specular roughness from a single picture of a flat surface lit by a hand-held flash. We achieve this goal by introducing several innovations on training data acquisition and network design. For training, we leverage a large dataset of artist-created, procedural SVBRDFs which we sample and render under multiple lighting directions. We further amplify the data by material mixing to cover a wide diversity of shading effects, which allows our network to work across many material classes. Motivated by the observation that distant regions of a material sample often offer complementary visual cues, we design a network that combines an encoder-decoder convolutional track for local feature extraction with a fully-connected track for *global feature* extraction and propagation. Many important material effects are view-dependent, and as such ambiguous when observed in a single image. We tackle this challenge by defining the loss as a differentiable SVBRDF similarity metric that compares the *renderings* of the predicted maps against renderings of the ground truth from several lighting and viewing directions. Combined together, these novel ingredients bring clear improvement over state of the art methods for single-shot capture of spatially varying BRDFs.

The work was published in ACM Transactions on Graphics and presented at SIGGRAPH 2018 [13], and was cited by several popular online ressources (https://venturebeat.com/2018/08/15/researchers-develop-ai-that-can-re-create-real-world-lighting-and-reflections/, https://www.youtube.com/watch?v=UkWnExEFADI).

Figure 6. From a single flash photograph of a material sample (insets), our deep learning approach predicts a spatially-varying BRDF. See supplemental materials for animations with a moving light.

6.2.2. Material Acquisition using an Arbitrary Number of Inputs

Participants: Valentin Deschaintre, Aittala Miika, Frédéric Durand, George Drettakis, Adrien Bousseau.

Single-image material acquisition methods try to solve the very ill-posed problem of appearance to parametric BRDF. We explore different acquisition configurations to solve the most important ambiguities while still focusing on convenience of acquisition. Our main exploration directions are multiple lights and view angles over multiple pictures. This is possible thanks to the use of deep learning and in-line input data rendering, allowing us to easily explore a wide variety of configurations simultaneously. We also specialize our network architecture to make the most of an arbitrary number of input, provided in any order.

6.2.3. Exploiting Repetitions for Image-Based Rendering of Facades

Participants: Simon Rodriguez, Adrien Bousseau, Frédéric Durand, George Drettakis.

Street-level imagery is now abundant but does not have sufficient capture density to be usable for Image-Based Rendering (IBR) of facades. We presented a method that exploits repetitive elements in facades – such as windows – to perform data augmentation, in turn improving camera calibration, reconstructed geometry and overall rendering quality for IBR. The main intuition behind our approach is that a few views of several

instances of an element provide similar information to many views of a single instance of that element. We first select similar instances of an element from 3-4 views of a facade and transform them into a common coordinate system (Fig. 7 (a)), creating a "platonic" element. We use this common space to refine the camera calibration of each view of each instance (Fig. 7 (b)) and to reconstruct a 3D mesh of the element with multi-view stereo, that we regularize to obtain a piecewise-planar mesh aligned with dominant image contours (Fig. 7 (c)). Observing the same element under multiple views also allows us to identify reflective areas – such as glass panels – (Fig. 7 (d)) which we use at rendering time to generate plausible reflections using an environment map. We also combine information from multiple viewpoints to augment our initial set of views of the elements (Fig. 7 (e)). Our detailed 3D mesh, augmented set of views, and reflection mask enable image-based rendering of much higher quality than results obtained using the input images directly(Fig. 7 (f)).

The work was published in Computer Graphics Forum, presented at the Eurographics Symposium on Rendering 2018 [16].

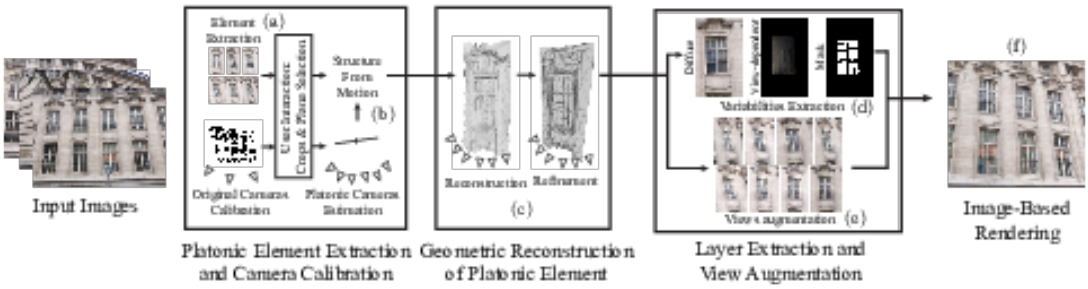


Figure 7. Overview of our technique for Image-Based Rendering of facades.

6.2.4. Plane-Based Multi-View Inpainting for Image-Based Rendering in Large Scenes

Participants: Julien Philip, George Drettakis.

Image-Based Rendering (IBR) allows high-fidelity free-viewpoint navigation using only a set of photographs and 3D reconstruction as input. It is often necessary or convenient to remove objects from the captured scenes, allowing a form of scene editing for IBR. This requires multi-view inpainting of the input images. Previous methods suffer from several major limitations: they lack true multi-view coherence, resulting in artifacts such as blur, they do not preserve perspective during inpainting, provide inaccurate depth completion and can only handle scenes with a few tens of images. Our approach addresses these limitations by introducing a new multi-view method that performs inpainting in intermediate, locally common planes. Use of these planes results in correct perspective and multi-view coherence of inpainting results. For efficient treatment of large scenes, we present a fast planar region extraction method operating on small image clusters. We adapt the resolution of inpainting to that required in each input image of the multi-view dataset, and carefully handle image resampling between the input images and rectified planes. We show results on large indoors and outdoors environments.

The work was presented at the ACM SIGGRAPH I3D Symposium on Interactive Computer Graphics and Games [19].

6.2.5. Deep Blending for Free-Viewpoint Image-Based Rendering

Participants: Julien Philip, George Drettakis.

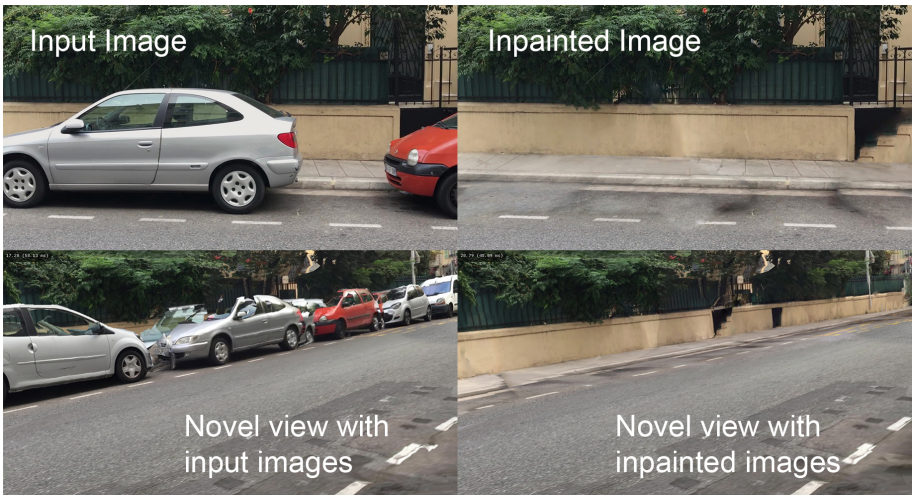


Figure 8. Our plane-based multi-view inpainting method allows us to remove cars in this large urban scene.

Free-viewpoint image-based rendering (IBR) is a standing challenge. IBR methods combine warped versions of input photos to synthesize a novel view. The image quality of this combination is directly affected by geometric inaccuracies of multi-view stereo (MVS) reconstruction and by view- and image-dependent effects that produce artifacts when contributions from different input views are blended. We present a new deep learning approach to blending for IBR, in which we use held-out real image data to learn blending weights to combine input photo contributions. Our Deep Blending method requires us to address several challenges to achieve our goal of interactive free-viewpoint IBR navigation. We first need to provide sufficiently accurate geometry so the Convolutional Neural Network (CNN) can succeed in finding correct blending weights. We do this by combining two different MVS reconstructions with complementary accuracy vs. completeness tradeoffs. To tightly integrate learning in an interactive IBR system, we need to adapt our rendering algorithm to produce a fixed number of input layers that can then be blended by the CNN. We generate training data with a variety of captured scenes, using each input photo as ground truth in a held-out approach. We also design the network architecture and the training loss to provide high quality novel view synthesis, while reducing temporal flickering artifacts. Our results demonstrate free-viewpoint IBR in a wide variety of scenes, clearly surpassing previous methods in visual quality, especially when moving far from the input cameras.

This work is a collaboration with Peter Hedman and Gabriel Brostow from University College London and True Price and Jan-Michael Frahm from University of North Carolina at Chapel Hill. It was published in ACM Transactions on Graphics and presented at SIGGRAPH Asia 2018 [14].

6.2.6. Thin Structures in Image Based Rendering

Participants: Theo Thonat, Abdelaziz Djelouah, Frédéric Durand, George Drettakis.

This work proposes a novel method to handle thin structures in Image-Based Rendering (IBR), and specifically structures supported by simple geometric shapes such as planes, cylinders, etc. These structures, e.g. railings, fences, oven grills etc, are present in many man-made environments and are extremely challenging for multi-view 3D reconstruction, representing a major limitation of existing IBR methods. Our key insight is to exploit multi-view information to compute multi-layer alpha mattes to extract the thin structures. We use two multi-view terms in a graph-cut segmentation, the first based on multi-view foreground color prediction and the second ensuring multi-view consistency of labels. Occlusion of the background can challenge reprojection error calculation and we use multi-view median images and variance, with multiple layers of thin structures.

Figure 9. Deep Blending for Free-Viewpoint Image-Based Rendering

Our end-to-end solution uses the multi-layer segmentation to create per-view mattes and the median colors and variance to extract a clean background. We introduce a new multi-pass IBR algorithm based on depth-peeling to allow free-viewpoint navigation of multi-layer semi-transparent thin structures. Our results show significant improvement in rendering quality for thin structures compared to previous image-based rendering solutions.

The work was published in the journal Computer Graphics Forum, and was presented at the Eurographics Symposium on Rendering (EGSR) 2018 [17].

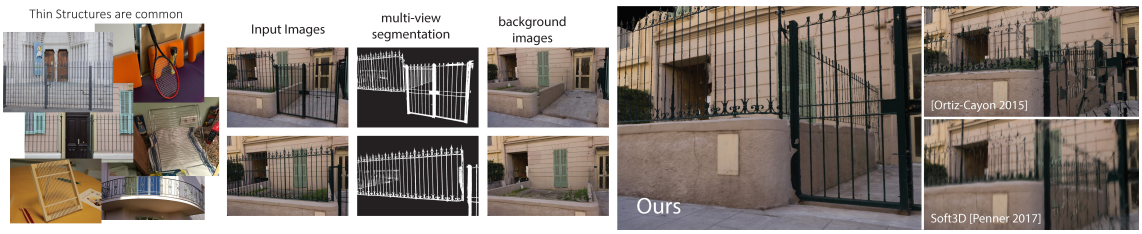


Figure 10. Thin structures are present in many environments, both indoors and outdoors (far left). Our solution extracts multi-view mattes together with clean background images and geometry (center). These elements are used by our multi-layer rendering algorithm that allows free-viewpoint navigation, with significantly improved quality compared to previous solutions (right).

6.2.7. Multi-Scale Simulation of Nonlinear Thin-Shell Sound with Wave Turbulence

Participants: Gabriel Cirio, George Drettakis.

Thin shells – solids that are thin in one dimension compared to the other two – often emit rich nonlinear sounds when struck. Strong excitations can even cause chaotic thin-shell vibrations, producing sounds whose energy spectrum diffuses from low to high frequencies over time – a phenomenon known as wave turbulence. It is all these nonlinearities that grant shells such as cymbals and gongs their characteristic “glinting” sound. Yet, simulation models that efficiently capture these sound effects remain elusive. In this project, we proposed a

physically based, multi-scale reduced simulation method to synthesize nonlinear thin-shell sounds. We first split nonlinear vibrations into two scales, with a small low-frequency part simulated in a fully nonlinear way, and a high-frequency part containing many more modes approximated through time-varying linearization. This allows us to capture interesting nonlinearities in the shells' deformation, tens of times faster than previous approaches. Furthermore, we propose a method that enriches simulated sounds with wave turbulent sound details through a phenomenological diffusion model in the frequency domain, and thereby sidestep the expensive simulation of chaotic high-frequency dynamics. We show several examples of our simulations, illustrating the efficiency and realism of our model, see Fig. 11 .

This work is a collaboration with Ante Qu from Stanford, Eitan Grinspun and Changzi Zheng from Columbia. This work was published at ACM Transactions on Graphics, and presented at SIGGRAPH 2018 [11].

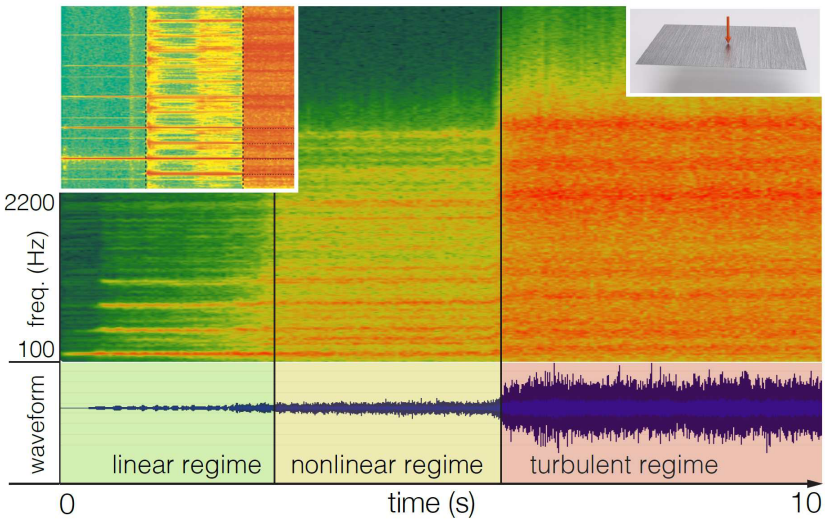


Figure 11. Thin-shell bifurcation. We excite a thin plate with increasing forces (the red arrow in the top-right inset) and simulate its dynamical responses). As the force increases, its vibration bifurcates, changing from linear vibration (left) to nonlinear (middle), and finally moving into a turbulent regime (right). This spectrogram is generated without any wave turbulence enrichment, indicating that model is able to capture chaos, albeit in low frequencies. We note that this spectrogram is qualitatively close to spectrograms from physical experiments, shown in the top-left inset (Image courtesy of Cyril Touzé).

6.2.8. Learning to Relight Multi-View Photographs from Synthetic Data

Participants: Julien Philip, George Drettakis.

We introduce an image relighting method that allows users to alter the lighting in their photos given multiple views of the same scene. Our method uses a deep convolutional network trained on synthetic photorealistic images. The use of a 3D reconstruction of the surroundings allows to guide the relighting process.

This ongoing project is a collaboration with Tinghui Zhou and Alexei A. Efros from UC Berkeley, and Michael Gharbi from Adobe research.

6.2.9. Exploiting Semantic Information for Street-level Image-Based Rendering

Participants: Simon Rodriguez, George Drettakis.

Following our work on facade rendering (Sec. 6.2.3), this ongoing project explores the use of semantic segmentation to inform Image-Based Rendering algorithms. In particular, we plan to devise algorithms that adapt to different types of objects in the scene (cars, buildings, trees).

6.2.10. Casual Video Based Rendering of Stochastic Phenomena

Participants: Theo Thonat, Miika Aittala, Frédéric Durand, 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 or DSLR camera, and a tripod. We focus on stochastic time-dependent textures such as leaves in the wind, water or fire, to cope with the challenge of using unsynchronized videos.

This ongoing work is a collaboration with Sylvain Paris from Adobe Research.

6.2.11. Cutting-Edge VR/AR Display Technologies

Participant: Koulieris Georgios.

Near-eye (VR/AR) displays suffer from technical, interaction as well as visual quality issues which hinder their commercial potential. We presented a tutorial that delivered an overview of cutting-edge VR/AR display technologies, focusing on technical, interaction and perceptual issues which, if solved, will drive the next generation of display technologies. The most recent advancements in near-eye displays were presented providing (i) correct accommodation cues, (ii) near-eye varifocal AR, (iii) high dynamic range rendition, (iv) gaze-aware capabilities, either predictive or based on eye-tracking as well as (v) motion-awareness (Fig. 12). Future avenues for academic and industrial research related to the next generation of AR/VR display technologies were analyzed.

This work is a collaboration with Kaan Akşit (NVIDIA), Christian Richardt (University of Bath), Rafal Mantiuk (University of Cambridge) and Katerina Mania (Technical University of Crete). The work was presented at IEEE VR 2018, 18-22 March, Reutlingen, Germany [18].

Figure 12. We presented novel display technologies, including but not limited to (left-to-right) varifocal augmented reality displays, body-tracking displays and focus-tunable displays.

HYBRID Project-Team

7. New Results

7.1. Virtual Reality Tools and Usages

7.1.1. Virtual Embodiment

Studying the Sense of Embodiment in VR Shared Experiences
Participants: Rebecca Fribourg, Ferran Argelaguet, Anatole Lécuyer

In [35], we explored the influence of sharing a virtual environment with another user on the sense of embodiment in virtual reality. For this aim, we conducted an experiment where users were immersed in a virtual environment while being embodied in an anthropomorphic virtual representation of themselves. To evaluate the influence of the presence of another user, two situations were studied: either users were immersed alone, or in the company of another user (see Figure 3). During the experiment, participants performed a virtual version of the well-known whac-a-mole game, therefore interacting with the virtual environment, while sitting at a virtual table. Our results show that users were significantly more "efficient" (i.e., faster reaction times), and accordingly more engaged, in performing the task when sharing the virtual environment, in particular for the more competitive tasks. Also, users experienced comparable levels of embodiment both when immersed alone or with another user. These results are supported by subjective questionnaires but also through behavioural responses, e.g. users reacting to the introduction of a threat towards their virtual body. Taken together, our results show that competition and shared experiences involving an avatar do not influence the sense of embodiment, but can increase user engagement. Such insights can be used by designers of virtual environments and virtual reality applications to develop more engaging applications.

This work was done with collaboration with Mimetic Inria team.

Figure 3. Studying the sense of embodiment in VR shared experiences: Setup of the experiment. Each user was able to interact in the virtual environment with his own avatar, while the physical setup provided both a reference frame and passive haptic feedback. From left to right: experimental conditions Alone, Mirror and Shared; Physical setup of the experiment.

Towards Novel Approaches to Characterise, Manipulate and Measure the Sense of Agency in Virtual Environments
Participants: Camille Jeunet, Ferran Argelaguet, Anatole Lécuyer

While the Sense of Agency (SoA) has so far been predominantly characterised in VR as a component of the Sense of Embodiment, other communities (e.g., in psychology or neurosciences) have investigated the SoA from a different perspective proposing complementary theories. Yet, despite the acknowledged potential benefits of catching up with these theories a gap remains. In [18], we first aimed to contribute to fill this gap by introducing a theory according to which the SoA can be divided into two components, the feeling and the judgment of agency, and relies on three principles, namely the principles of priority, exclusivity and consistency. We argued that this theory could provide insights on the factors influencing the SoA in VR systems. Second, we proposed novel approaches to manipulate the SoA in controlled VR experiments (based on these three principles) as well as to measure the SoA, and more specifically its two components based on neurophysiological markers, using ElectroEncephaloGraphy (EEG). We claim that these approaches would enable us to deepen our understanding of the SoA in VR contexts. Finally, we validated these approaches in an experiment (see Figure 4). Our results (N=24) suggest that our approach was successful in manipulating the SoA as the modulation of each of the three principles induced significant decreases of the SoA (measured using questionnaires). In addition, we recorded participants' EEG signals during the VR experiment, and neurophysiological markers of the SoA, potentially reflecting the feeling and judgment of agency specifically, were revealed. Our results also suggest that users' profile, more precisely their Locus of Control (LoC), influences their level of immersion and SoA.

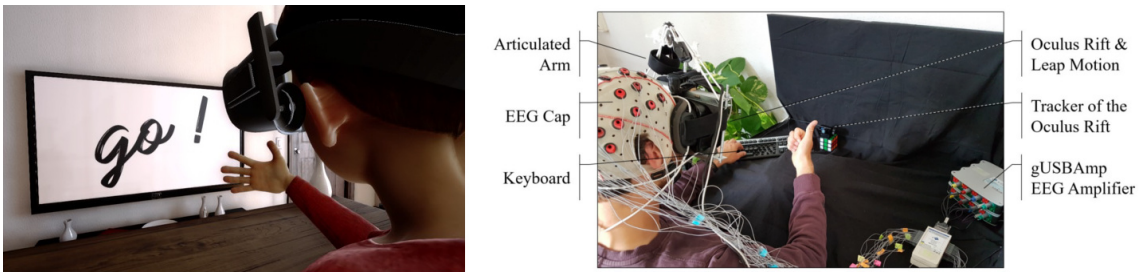


Figure 4. Studying the sense of agency in VR. Left: Third-person perspective: the participant receives a go signal and starts to perform the movement. Right: experimental set-up. The participant is equipped with an EEG cap, plugged to g.USBAmp amplifiers. In addition, he is immersed in the virtual environment using an Oculus Rift attached to his head and supported by an articulated arm (to avoid any pressure on the EEG cap and reduce the risk of muscular fatigue). Finally, his head is tracked by the Oculus tracker and his right hand is tracked by a Leap Motion fixed in front of the Oculus Rift.

Virtual Shadows for Real Humans in a CAVE: Influence on Virtual Embodiment and 3D Interaction
Participants: Guillaume Cortes, Ferran Argelaguet, Anatole Lécuyer

In immersive projection systems (IPS), the presence of the user's real body limits the possibility to elicit a virtual body ownership illusion. But, is it still possible to embody someone else in an IPS even though the users are aware of their real body? In order to study this question, we proposed to consider using a virtual shadow in the IPS, which can be similar or different from the real user's morphology [29]. We conducted an experiment (N=27) to study the users' sense of embodiment whenever a virtual shadow was or was not present (see Figure 5). Participants had to perform a 3D positioning task in which accuracy was the main requirement. The results showed that users widely accepted their virtual shadow (agency and ownership) and felt more comfortable when interacting with it (compare to no virtual shadow). Yet, due to the awareness of their real body, the users have less acceptance of the virtual shadow whenever the shadow gender differs from their own. Furthermore, the results showed that virtual shadows increase the users' spatial perception of the virtual environment by decreasing the inter-penetrations between the user and the virtual objects. Taken

together, our results promote the use of dynamic and realistic virtual shadows in IPS and pave the way for further studies on "virtual shadow ownership" illusion.

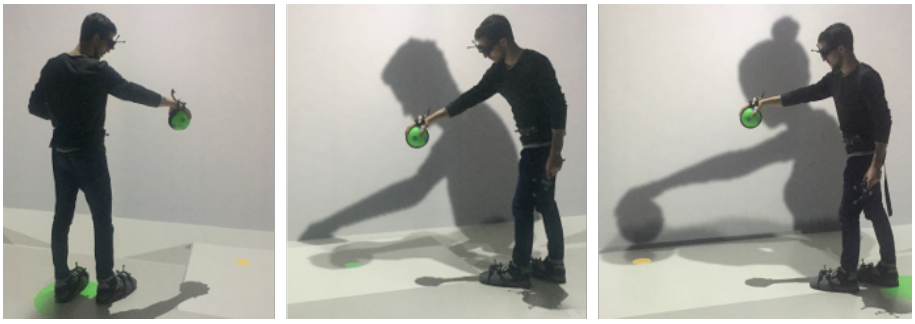


Figure 5. Various virtual shadows conditions. The participants performed a positioning task with 3 different virtual shadow conditions: None (N) (left), Male (M) (middle), Female (F) (right). The real shadow of the user is visible on the floor but does not match the natural behavior of a shadow in the virtual environment and is not taken into consideration.

This work was done with collaboration with Rainbow Inria team.

Influence of Being Embodied in an Obese Virtual Body on Shopping Behavior and Products Perception in VR

Participants: Jean-Marie Normand, Guillaume Moreau

In [26], we studied the changes an obese virtual body has on products perception (e.g., taste, etc.) and purchase behavior (e.g., number purchased) in an immersive virtual retail store. Participants (of a normal BMI on average) were embodied in a normal (N) or an obese (OB) virtual body and were asked to buy and evaluate food products in the immersive virtual store (see Figure 6 . Based on stereotypes that are classically associated with obese people, we expected that the group embodied in obese avatars would show a more unhealthy diet, (i.e., buy more food products and also buy more products with high energy intake, or saturated fat) and would rate unhealthy food as being tastier and healthier than participants embodied in "normal weight" avatars. Our participants also rated the perception of their virtual body: the OB group perceived their virtual body as significantly heavier and older. Stereotype activation failed for our participants embodied in obese avatars, who did not exhibit a shopping behavior following the (negative) stereotypes related to obese people. Participants might have rejected their virtual bodies when performing the shopping task, while the embodiment and presence ratings did not show significant differences, and purchased products based on their real (non-obese) bodies. This could mean that stereotype activation is more complex that previously thought.

7.1.2. VR and Building Information Modeling

OpenBIM-based Ontology for Interactive Virtual Environments

Participants: Anne-Solène Dris, François Lehericey, Valérie Gouranton, Bruno Arnaldi

We proposed an ontology improving the use of Building Information Modelling (BIM) models as an Interactive Virtual Environment (IVE) generator [33]. Our results enable to create a bidirectional link between the informed 3D database and the virtual reality application, and to automatically generate object-specific functions and capabilities according to their taxonomy. We presented an illustration of our results based on a Risk-Hunting training application. In such contexts, the notions of objects handling and scheduling of the construction are essential for the immersion of the future trainee as well as for the success of the training.

Risk-Hunting Training in Interactive Virtual Environments

Participants: Anne-Solène Dris, François Lehericey, Valérie Gouranton, Bruno Analdi

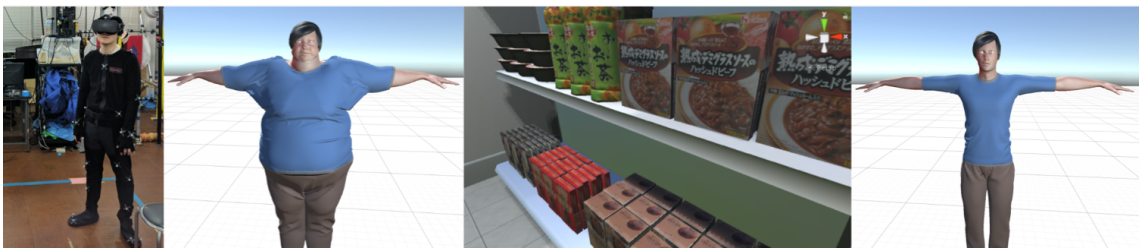


Figure 6. Being embodied in an obese virtual body. From left to right: A participant in a motion capture suit; The obese male avatar; A close-up on some products of our virtual store; The male avatar with a "normal" Body Mass Index.

Safety is an everlasting concern in construction environments. In such applications, when an accident happens it is rarely harmless. To raise awareness and train workers to safety procedures, training centers propose risk-hunting courses in which real-life equipment is set up in an incorrect way. Trainees can safely observe these environments and are supposed to point at risk situations. In [34], we proposed a risk-hunting course in Virtual Reality. With VR, we can put the trainee in a full construction environment with potentially dangerous hazards without engaging his safety. Contrary to others risk-hunting courses, we have designed a virtual environment with interactions to emphasize the importance of learning to correct the errors. First, instead of only having to spot the errors, the trainee had to fix them. Then, a second way to exploit VR interaction capabilities consisted in introducing consequences of not fixing an error. For example, not fixing an error in a scaffolding would make it collapse later. This implies to rely on script-writing the virtual environment to add causality on specific actions. Our goal was here to educate the trainee about the dramatic consequences that could arise when errors are not corrected.

7.1.3. Augmented Reality Methods and Applications

MoSART: Mobile Spatial Augmented Reality for 3D Interaction With Tangible Objects
Participants: Guillaume Cortes, Anatole Lécuyer

In [11] we introduced MoSART: a novel approach for Mobile Spatial Augmented Reality on Tangible objects. MoSART is dedicated to mobile interaction with tangible objects in single or collaborative situations. It is based on a novel "all-in-one" Head-Mounted Display (AMD) including a projector (for the SAR display) and cameras (for the scene registration). Equipped with the HMD the user is able to move freely around tangible objects and manipulate them at will. The system tracks the position and orientation of the tangible 3D objects and projects virtual content over them. The tracking is a feature-based stereo optical tracking providing high accuracy and low latency. A projection mapping technique is used for the projection on the tangible objects which can have a complex 3D geometry. Several interaction tools have also been designed to interact with the tangible and augmented content, such as a control panel and a pointer metaphor, which can benefit as well from the MoSART projection mapping and tracking features. The possibilities offered by our novel approach are illustrated in several use cases, in single or collaborative situations, such as for virtual prototyping, training or medical visualization.

This work was done with collaboration with Rainbow Inria team.
Evaluation of 2D and 3D Ultrasound Tracking Algorithms
Participants: Maud Marchal

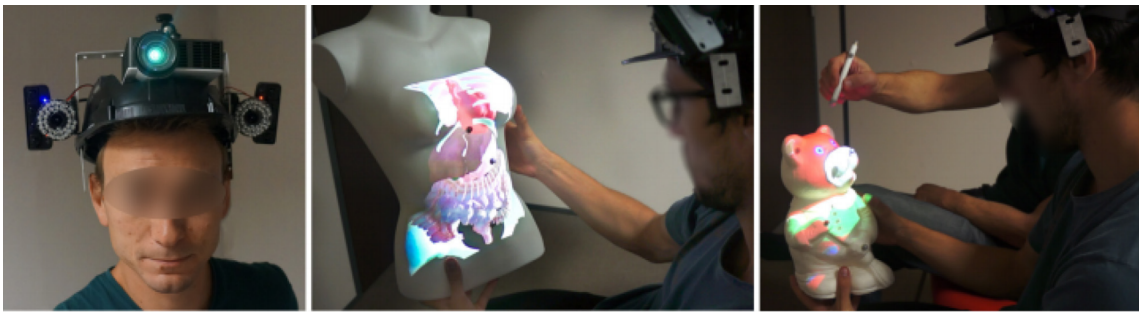


Figure 7. The MoSART headset for wearable augmented reality on tangible objects. Our novel system relies on an "all-in-one" Head-Mounted-Display (Left) which embeds a pico-projector for projection mapping and two cameras for feature-based stereo optical tracking of 3D tangible objects. The user can freely walk around and manipulate tangible objects superimposed with the projected images, such as for medical visualization purposes (Center). Tangible tools can also be used to interact with the virtual content such as for annotating or painting the objects in single or collaborative scenarios (Right).

Compensation for respiratory motion is important during abdominal cancer treatments. In [12], the results of the 2015 MICCAI Challenge on Liver Ultrasound Tracking are reported. These results extend the 2D results to relate them to clinical relevance in form of reducing treatment margins and hence sparing healthy tissues, while maintaining full duty cycle. The different methodologies of the MICCAI challenge are described for estimating and temporally predicting respiratory liver motion from continuous ultrasound imaging, used during ultrasound-guided radiation therapy. Furthermore, the trade-off between tracking accuracy and runtime in combination with temporal prediction strategies and their impact on treatment margins is also investigated. The paper follows the work of the PhD of Lucas Royer defended in 2016 and his methodology that was ranked first in the MICCAI challenge.

Evaluation of AR Inconsistencies on AR Placement Tasks: A VR Simulation Study
Participants: Romain Terrier, Jean-Marie Normand, Ferran Argelaguet

One of the major challenges of Augmented Reality (AR) is the registration of virtual and real contents. When errors occur during the registration process, inconsistencies between real and virtual contents arise and can alter user interaction. In this work, we assessed the impact of registration errors on the user performance and behaviour during an AR pick-and-place task in a Virtual Reality (VR) simulation [41]. The VR simulation ensured the repeatability and control over experimental conditions. The paper describes the VR simulation framework used and three experiments studying how registration errors (e.g., rotational errors, positional errors, shaking) and visualization modalities (e.g., transparency, occlusion) modify the user behaviour while performing a pick-and-place task. Our results show that users kept a constant behavior during the task, i.e., the interaction was driven either by the VR or the AR content, except if the registration errors did not enable to efficiently perform the task. Furthermore, users showed preference towards an half-transparent AR in which correct depth sorting is provided between AR and VR contents. Taken together, our results open perspectives for the design and evaluation of AR applications through VR simulation frameworks.

7.1.4. The 3DUI Contest 2018

Every year, the international IEEE Virtual Reality Conference organizes an annual 3D User Interfaces contest. This year, Hybrid submitted two different proposals.

Toward Intuitive 3D User Interfaces for Climbing, Flying and Stacking
Participants: Antonin Bernardin, Guillaume Cortes, Rebecca Fribourg, Tiffany Luong, Florian Nouviale, Hakim Si-Mohammed

Figure 8. First solution proposed to the 2018 3DUI Contest. First-Person Drone Flying: The 3D User Interface used to control the drone (left). Ladder Climbing: First person point of view of the ladder climbing (center). Object Stacking: Physical object manipulation with frame recording as indicated by the red round on the controller and time control (right).

In this first solution, we proposed 3D user interfaces that are adapted to specific Virtual Reality tasks: climbing a ladder using a puppet metaphor, piloting a drone thanks to a 3D virtual compass and stacking 3D objects with physics-based manipulation and time control [28]. These metaphors have been designed to provide the user with an intuitive, playful and efficient way to perform each task (see Figure 8).

Climb, Fly, Stack: Design of Tangible and Gesture-based Interfaces for Natural and Efficient Interaction
Participants: Alexandre Audinot, Emeric Goga, Vincent Goupil, Carl-Johan Jorgensen, Adrien Reuzeau, Ferran Argelaguet

In this second solution we proposed three different 3D interaction metaphors conceived to fulfill the three tasks proposed in the IEEE VR 3DUI Contest. We proposed the VladdeR, a tangible interface for Virtual laddeR climbing, the FPDrone, a First Person Drone control flying interface, and the Dice Cup, a tangible interface for virtual object stacking [27]. All three metaphors take advantage of body proprioception and previous knowledge of real life interactions without the need of complex interaction mechanics (see Figure 9): climbing a tangible ladder through arm and leg motions, control a drone like a child flies an imaginary plane by extending your arms or stacking objects as you will grab and stack dice with a dice cup.

7.2. Physically-Based Simulation and Haptic Feedback

7.2.1. Haptic Methods and Rendering

KinesTouch: 3D Force-Feedback Rendering for Tactile Surfaces
Participants: Antoine Costes, Ferran Argelaguet, Anatole Lécuyer

Haptic enhancement of touchscreens has been mostly addressed through the use of various types of vibrations, altering the physics of the finger sliding on the screen, in order to provide friction forces and even small relief sensations. However, such approaches do not allow for displaying other haptic properties such as stiffness or large-scale shapes. In [30], we introduced the "KinesTouch", a novel approach for touchscreen enhancement providing four types of haptic feedback with a single force-feedback device: compliance, friction, fine roughness, and shape. Regarding friction in particular, we proposed a novel effect based on large lateral motion that increases or diminishes the sliding velocity between the finger and the screen. Our results show that this effect is able to produce distinct sliding sensations. Our general approach is also illustrated through a set of interactive use cases of 2D/3D content manipulation in various contexts.

This work was done in collaboration with Technicolor.

Haptic Material: a Holistic Approach for Haptic Texture Mapping
Participants: Antoine Costes, Ferran Argelaguet, Anatole Lécuyer

Figure 9. Second solution proposed to the 2018 3DUI Contest. Left, flying interface. Center, climbing interface. Left, stacking interface.

The development of 3D scanning technologies made common the digitizing of objects in realistic virtual copies, but still at the cost of most of their haptic properties. Besides, while haptic devices and setups spread widely, little attention is paid to the reuse and compatibility of haptic data, which are most of the time context- or hardware-specific. In [31], we proposed a new format for haptic texture mapping which is not dependent on the haptic rendering setup hardware. Our "haptic material" format encodes ten elementary haptic features in dedicated maps, similarly to "materials" used in computer graphics. These ten different features enable the expression of compliance, surface geometry and friction attributes through vibratory, cutaneous and kinesthetic cues, as well as thermal rendering. The diversity of haptic data allows various hardware to share this single format, each of them selecting which features to render depending on its capabilities.

This work was done in collaboration with Technicolor.

Combining Tangible Objects and Wearable Haptics

Participants: Xavier de Tinguy, Maud Marchal, Anatole Lécuyer

In [32], we studied the combination of tangible objects and wearable haptics for improving the display of stiffness sensations in virtual environments. Tangible objects enable to feel the general shape of objects, but they are often passive or unable to simulate several varying mechanical properties. Wearable haptic devices are portable and unobtrusive interfaces able to generate varying tactile sensations, but they often fail at providing convincing stiff contacts and distributed shape sensations. We propose to combine these two approaches in virtual and augmented reality (VR/AR), becoming able of arbitrarily augmenting the perceived stiffness of real/tangible objects by providing timely tactile stimuli at the fingers. We developed a proof-of-concept enabling to simulate varying elasticity/stiffness sensations when interacting with tangible objects by using wearable tactile modules at the fingertips. We carried out a user study showing that wearable haptic stimulation can well alter the perceived stiffness of real objects, even when the tactile stimuli is not delivered at the contact point. We illustrated our approach both in VR and AR, within several use cases and different tangible settings, such as when touching surfaces, pressing buttons and pistons, or holding an object (see Figure 12). Taken together, our results pave the way for novel haptic sensations in VR/AR by better exploiting the multiple ways of providing simple, unobtrusive, and low-cost haptic displays.

This work was done in collaboration with Rainbow Inria team.

7.2.2. Haptic Applications

A Survey on the Use of Haptic and Tactile Information in the Car to Improve Driving Safety

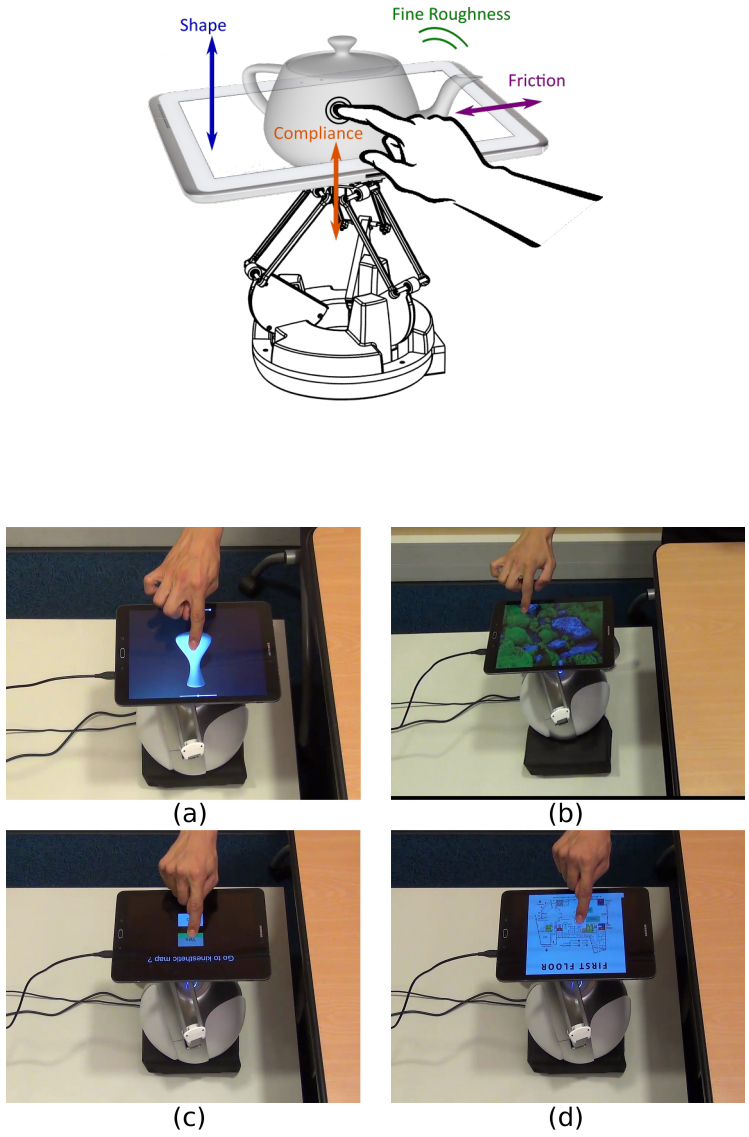


Figure 10. The KinesTouch approach. Top: concept of KinesTouch to provide four different types of haptic feedback to a touchscreen. Bottom: Use cases illustrating our approach: (a) Interaction with a 3D object, (b) Texture of a 2D image, (c) GUI and haptic buttons, (d) Interactive map.

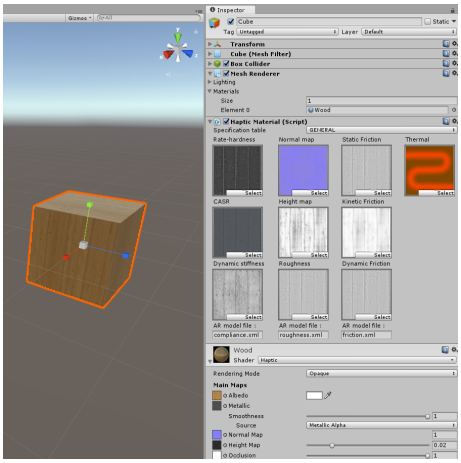


Figure 11. Implementation of our haptic material format in Unity3D.

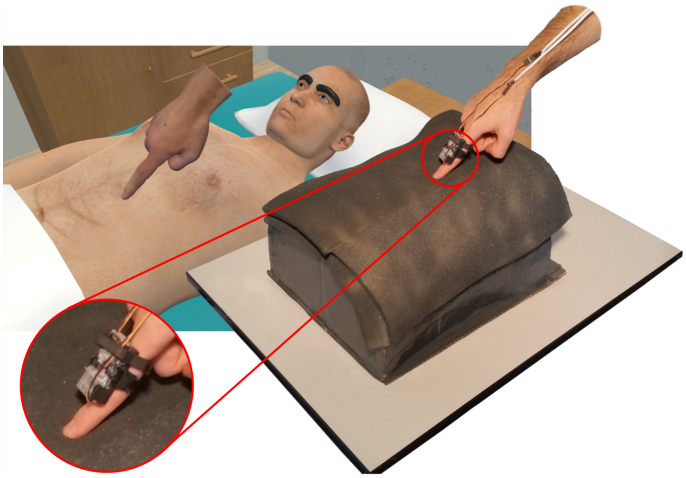


Figure 12. Combining tangible objects and wearable haptics in a VR medical palpation simulator. Passive tangible objects (a tangible chest here) provide haptic information about the global shape/percept of the virtual objects, while wearable haptic devices provide haptic information about dynamically changing mechanical properties (local elasticity here).

Participants: Yoren Gaffary, Anatole Lécuyer

In [15], we presented an overview of haptic technologies deployed in cars and their uses to enhance drivers' safety during manual driving. These technologies enable to deliver haptic (tactile or kinesthetic) feedback at various areas of the car, such as the steering wheel or the pedal. The paper explores two main uses of the haptic modality to fulfill the safety objective: providing driving assistance and warning. Driving assistance concerns the transmission of information usually conveyed with other modalities for controlling the cars' functions, maneuvering support, and guidance. Warning concerns the prevention of accidents using emergency warnings, increasing the awareness of surroundings, and preventing collisions, lane departures, and speeding. This paper discusses how haptic feedback has been introduced so far for these purposes and provides perspectives regarding the present and future of haptic cars meant to increase drivers' safety.

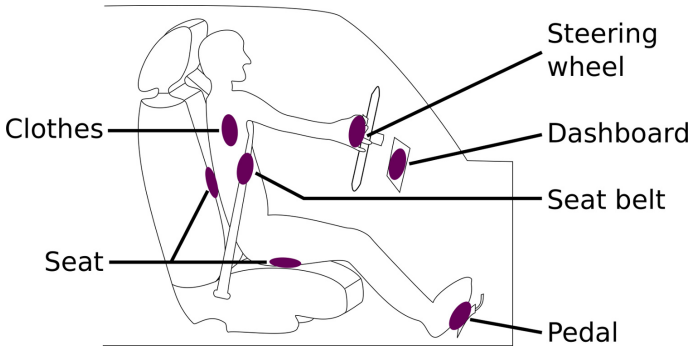


Figure 13. Using haptics in the car to improve driving safety: Different areas covered with haptic stimulations.

Toward Haptic Communication and Tactile Alphabets
Participants: Yoren Gaffary, Maud Marchal, Fernando Argelaguet Sanz, Anatole Lécuyer

In [14], we studied the possibility to convey information using tactile stimulation on fingertips. We designed and evaluated three tactile alphabets which are rendered by stretching the skin of the index's fingertip: (1) a Morse-like alphabet, (2) a symbolic alphabet using two successive dashes, and (3) a display of Roman letters based on the Unistrokes alphabet. All three alphabets (26 letters each) were evaluated through a user study in terms of recognition rate, intuitiveness and learning. Participants were able to perceive and recognize the letters with very good results (80%-97% recognition rates). Tactile alphabets with representations closer to Roman alphabet seem easier to learn. Taken together, our results pave the way to novel kinds of information communication using tactile modality.

This work was done in collaboration with CEA LIST.

7.3. Brain-Computer Interfaces

7.3.1. BCI Methods and Techniques

SimBCI: Novel Software Framework for Studying BCI Methods

Participants: Jussi Lindgren and Anatole Lécuyer

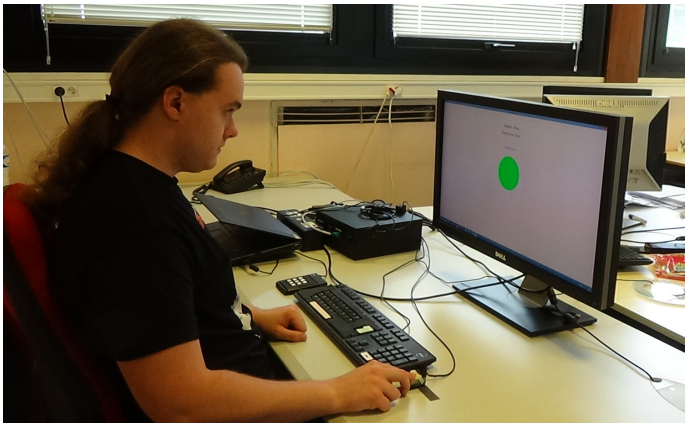


Figure 14. Toward tactile alphabets: A participant perceives a letter haptically stimulated using skin stretching at the level of his index.

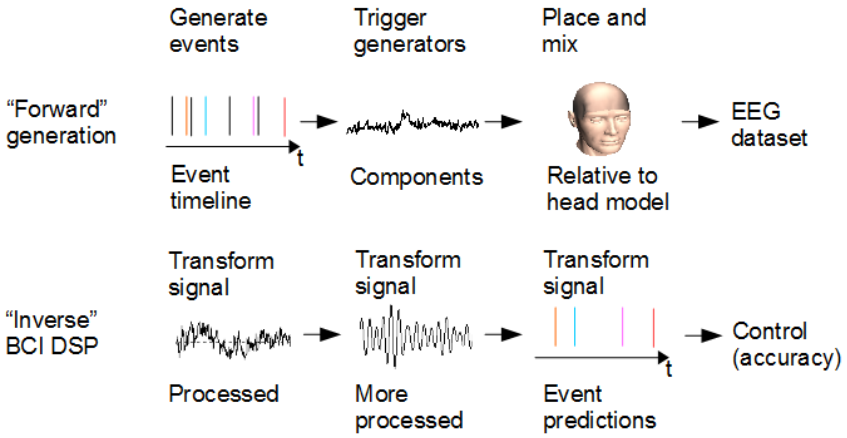


Figure 15. Simulated BCI data generation and testing in simBCI.

How to investigate the applicability of physiology-based source reconstruction for Brain-Computer Interfaces (BCIs)? The classic way is human experiments, but these unfortunately lack ground truth. The electrical activity inside the human brain is not fully described by external EEG measurements. In the CominLabs project SABRE, we have developed a BCI simulator framework called simBCI [22] to help in such studies. The framework allows modifying and changing generative models and their parameters inside a model brain, and studying what effects such changes have on the signal and subsequently the BCI signal processing. The modifiable parameters can include artifact properties, generative source locations, background activity characteristics and so on. We have released the framework as open source to the community (https://gitlab.inria.fr/sb/simbci/).

This work was done in collaboration with IMT Atlantique.

Novel Control Strategy for BCI Exploiting Visual Imagery and Attention

Participants: Jussi Lindgren and Anatole Lécuyer

Current paradigms for Brain-Computer Interfaces (BCIs) leave a lot to be desired in their accuracy and usability. We studied visual imagery as a potential new paradigm. In visual imagery, the user imagines objects or scenes visually, and the BCI is based on trying to classify the imagination type based on the EEG measurements. In [20], we studied to what extent can we distinguish the different mental processes of observing visual stimuli and imagining them based on the EEG. We found in a study of 26 users that we could somewhat differentiate (i) visual imagery vs. visual observation task (71% of classification accuracy), (ii) visual observation task towards different visual stimuli (classifying one observation cue versus another observation cue with an accuracy of 61%) and (iii) resting vs. observation/imagery (77% accuracy between imagery task versus resting state, and the accuracy of 75% between observation task versus resting state). All reported accuracies are averages over the users. Our results suggest that the presence of visual imagery and related alpha power changes may be useful to broaden the range of BCI control strategies.

Figure 16. Imagining or perceiving flowers and hammers. Can we tell from EEG which task the user is performing?

7.3.2. BCI Applications

BCI-based Interfaces for Augmented Reality: Feasibility, Design and Evaluation

Participants: Hakim Si-Mohammed, Camille Jeunet, Ferran Argelaguet and Anatole Lécuyer

In [25], we have studied the combination of BCI and Augmented Reality (AR). We first tested the feasibility of using BCI in AR settings based on Optical See-Through Head-Mounted Displays (OST-HMDs). Experimental results showed that a BCI and an OST-HMD equipment (EEG headset and Hololens in our case) are well compatible and that small movements of the head can be tolerated when using the BCI. Second, we introduced a design space for command display strategies based on BCI in AR, when exploiting a famous brain pattern called Steady-State Visually Evoked Potential (SSVEP). Our design space relies on five dimensions concerning the visual layout of the BCI menu ; namely: orientation, frame-of-reference, anchorage, size and explicitness. We implemented various BCI-based display strategies and tested them within the context of mobile robot control in AR. Our findings were finally integrated within an operational prototype based on a real mobile robot that is controlled in AR using a BCI and a HoloLens headset. Taken together our results (from four user studies) and our methodology could pave the way to future interaction schemes in Augmented Reality exploiting 3D User Interfaces based on brain activity and BCIs.

This work was done in collaboration with Loki Inria team.

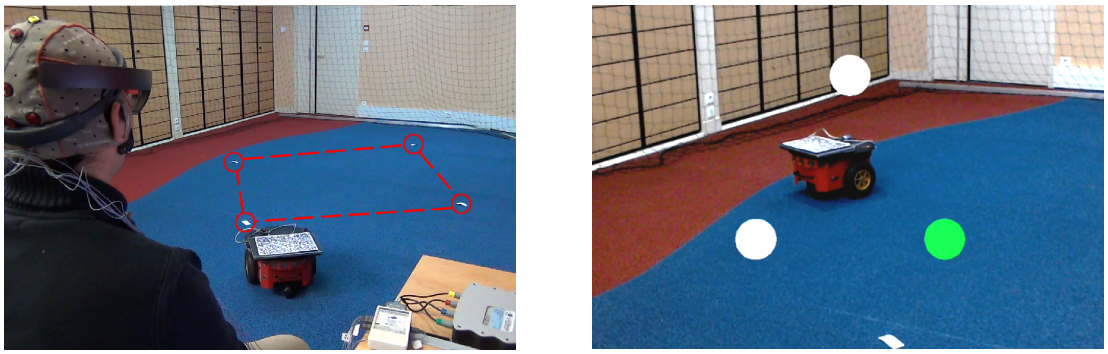


Figure 17. BCI-based interfaces in augmented reality: Illustration of our final prototype in use. (Left) general overview of the setup with the user equipped with EEG, sitting and facing the real mobile robot. (Right) First-person view, as seen from the HoloLens. The dashed line represents the path that the robot moved through during testing sessions.

Neurofeedback for Stroke Rehabilitation: A Case Report

Participants: Giulia Lioi, Mathis Fleury and Anatole Lécuyer

Neurofeedback (NF) consists in training self-regulation of brain activity by providing real-time information about the participant brain function. Few works have shown the potential of NF for stroke rehabilitation however its effectiveness has not been investigated yet. NF approaches are usually based on real-time monitoring of brain activity using a single imaging technique. Recent studies have revealed the potential of combining EEG and fMRI to achieve a more efficient and specific self-regulation. In this case report [49], we tested the feasibility of applying bimodal EEG-MRI NF on two stroke patients affected by left hemiplegia participated. The protocol included a calibration step (motor imagery of hemiplegic hand) and two NF sessions (5 minutes each). The experiment was run using a NF platform performing real-time EEG-fMRI processing and NF presentation. Both patients were found able to self-regulate their brain activity during the NF sessions. The EEG activity was harder to modulate than the BOLD activity. The patients were highly motivated to engage and satisfied with the NF animation, as assessed with a qualitative questionnaire. These results showed the feasibility and the potential of applying EEG-fMRI NF for stroke rehabilitation.

This work was done in collaboration with Visages Inria team.

Using EEG in Sport Performance Analysis

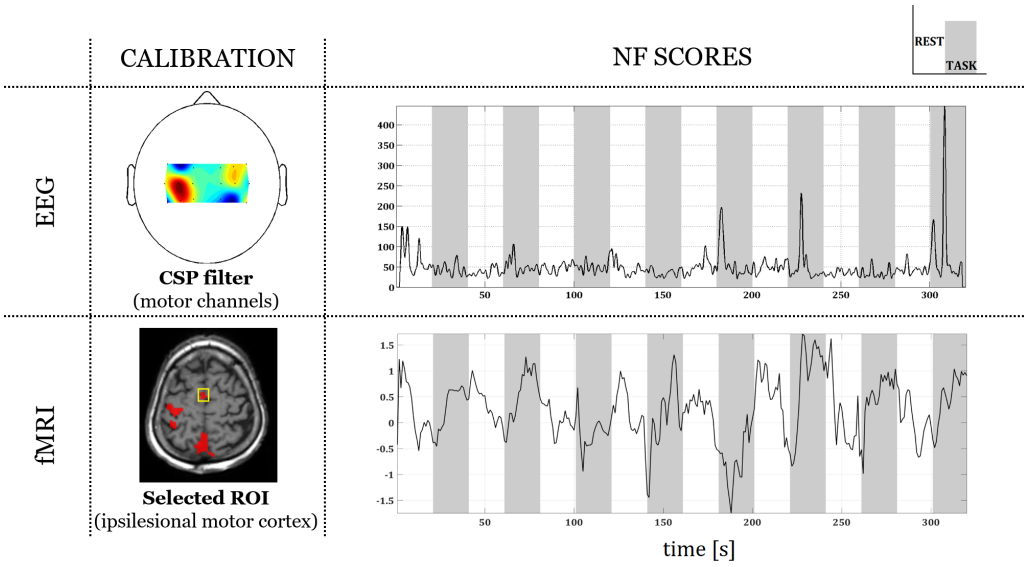


Figure 18. Neurofeedback for stroke patients: Examples of EEG and fMRI neurofeedback (NF) scores for a single session (patient 1). The left column represent the regions of interest selected to compute the NF signal during calibration. Resting blocks (20s) are indicated in white, NF training blocks (20s) in gray.

Participants: Ferran Argelaguet and Anatole Lécuyer

Competition changes the environment for athletes. The difficulty of training for such stressful events can lead to the well-known effect of "choking" under pressure, which prevents athletes from performing at their best level. To study the effect of competition on the human brain we recorded [24] pilot electroencephalography (EEG) data while novice shooters were immersed in a realistic virtual environment representing a shooting range. We found a differential between-subject effect of competition on mu (8–12 Hz) oscillatory activity during aiming; compared to training, the more the subject was able to desynchronize his mu rhythm during competition, the better was his shooting performance. Because this differential effect could not be explained by differences in simple measures of the kinematics and muscular activity, nor by the effect of competition or shooting performance per se, we interpret our results as evidence that mu desynchronization has a positive effect on performance during competition. It remains to show whether this effect can be generalized to expert shooters. Our findings could be relevant in sports training to help athletes avoid choking under pressure during competition. Confirmation through further experimental validation is however needed.

This work was done in collaboration with EPFL.

7.4. Cultural Heritage

Through several collaborations with Cultural Heritage partners such as archaeologists, historians, or curators, the Hybrid team has developed a methodology to propose new practices and tools in this domain. This methodology combines different technologies of digitization, such as CT scan, photogrammetry, or lidar, 3D production, such as 3D modelling or 3D printing, and 3D interactions in VR and AR.

7.4.1. 3D Printing and AR Applications

Lift the Veil of the Block Samples from the Warcq Chariot Burial

Participants: Ronan Gaugne and Valérie Gouranton

Figure 19. Experimental setup used in our sport performance study. (Left) Subjects were standing in our immersive projection system and were able to interact with the system using an ART Flystick. (Right) Subjects were wearing a high-density 64 channels EEG cap.

Cultural Heritage (CH) professionals such as archaeologists and conservators regularly experience the problem of working on concealed artifacts and face the potential destruction of source material without real understanding of the internal structure or state of decay or modification of the initial context by the micro-excavation process. Medical images-based digitization, such as MRI or CT scan, are increasingly used in CH as they provide information on the internal structure of archaeological material. Likewise, additive technologies are used more and more in the Cultural Heritage process, for example, in order to reproduce, complete, study or exhibit artifacts. 3D copies are based on digitization techniques such as laser scan or photogrammetry. In this case, the 3d copy remains limited to the external surface of objects. Different previous works illustrated the interest of combining 3D printing and Computed Tomography (CT) scans in order to extract concealed artifacts from larger archaeological material. The method was based on 3D segmentation techniques within volume data obtained by CT scans to isolate nested objects. This approach was useful to perform a digital extraction, but in some case it is also interesting to observe the internal spatial organization of an intricate object in order to understand its production process. Then, we proposed a method for the representation of a complex internal structure based on a combination of CT scan and emerging 3D printing techniques mixing colored and transparent parts of an aggregate of objects (see Figure 20), with very small pieces, from an exceptional aristocratic Gallic grave in the context of a preventive archaeological investigation [39].

This project was done in collaboration with UMR Trajectoires, Inrap and Image ET/BCRX.

Digital Introspection of a Mummy Cat

Participants: Ronan Gaugne and Valérie Gouranton

In the last decade, thanks to the dissemination of novel medical imaging technologies, research on the study of animal mummies of Ancient Egypt has became more and more important, leading to a better understanding of the history and culture of this civilization. Modern 3D technologies such as virtual reality, augmented reality and 3D printing enable to enrich the research process and open innovative possibilities for scenography in scientific mediation. In [36] we focused on one particular mummy cat and proposed to combine CT scan, 3D printing and augmented reality in a global process to accompany and support at the same time a scientific study of the object and a preparation of a mediation action in a Museum (see Figure 21 and Figure 22).

Figure 20. Transparent 3D printings from CT scan of archaeological materials.

This project was done in collaboration with Inrap, UMR Trajectoires, HISoMA and Musée des Beaux-Arts, Rennes.

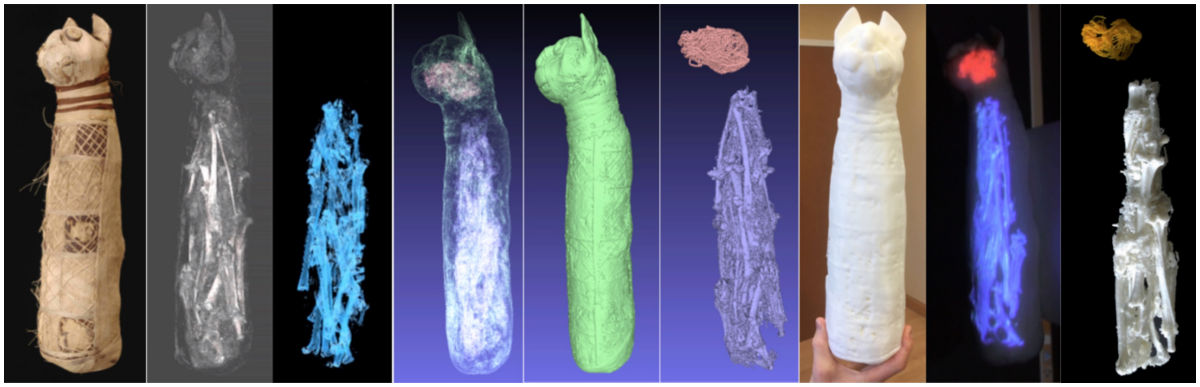


Figure 21. Digital introspection of a mummy cat. From left to right: the original mummy, CT scan of the mummy, volume rendering of the bones, mesh generation from CT scan, mesh of the external shape, meshes of internal parts, 3D printing of the external shape, projective AR of internal parts, 3D printing of internal parts.

7.4.2. VR Applications

EvoluSon: Walking through an Interactive History of Music

Participants: Ronan Gaugne, Florian Nouviale and Valérie Gouranton

The EvoluSon project [16] proposes an immersive experience where the spectator explores an interactive visual and musical representation of the main periods of the history of Western music (see Figure 23). The musical content is constituted of original musical compositions based on the theme of Bach's Art of Fugue to illustrate

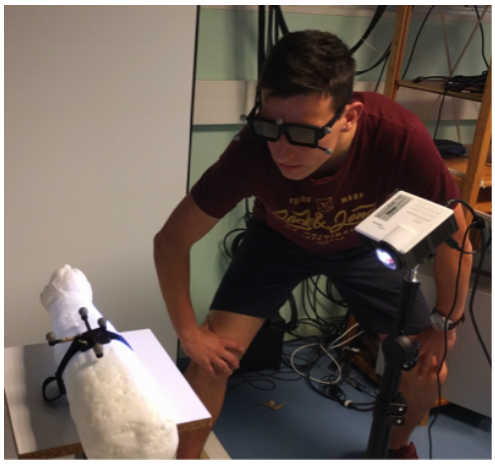


Figure 22. Projective AR system used for the visualization of the internal content of a mummy cat.

the eight main musical eras from Antiquity to the contemporary epoch. The EvoluSon project contributes at the same time to the usage of VR for intangible culture representation and to interactive digital art that puts the user at the centre of the experience. The EvoluSon project focuses on music through a presentation of the history of Western music, and uses virtual reality to valorise the different pieces through the ages. The user is immersed in a coherent visual and sound environment and can interact with both modalities. This project is the result of collaboration between a computer science research laboratory and a research laboratory on art and music. It was first presented to a public event on science and music organised by the computer science research laboratory.

This project was done in collaboration with the Research Laboratory on Art and Music of University Rennes 2.

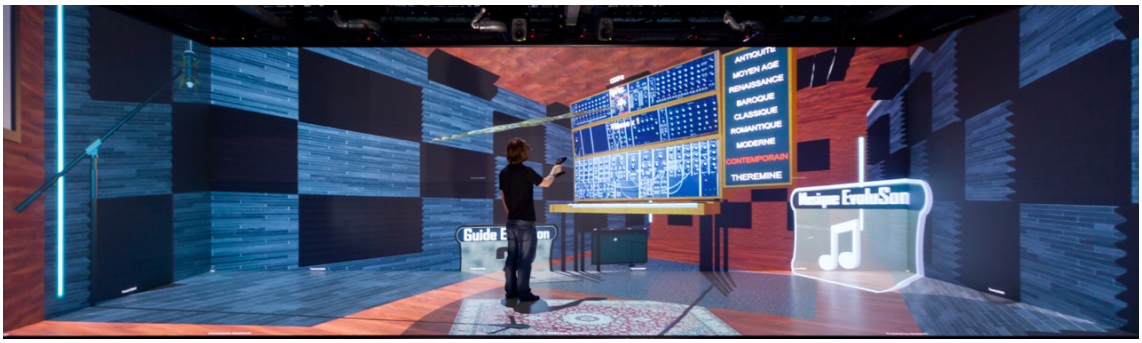


Figure 23. Interacting with the music through the ages inside Evoluson.

INSIDE Interactive and Non-destructive Solution for Introspection in Digital Environments

Participants: Flavien Lécuyer, Valérie Gouranton, Ronan Gaugne and Bruno Arnaldi

The development of scanning technologies allowed to limit the destructiveness induced by the excavation. However, it is not enough, as the rendering is not enough to study a scanned artifact. We proposed to use virtual reality as a legitimate tool for the inspection of artifacts modelled in 3D: INSIDE [38], with tools to lead a complete virtual excavation (see Figure 24). This tool opens a new way of practicing archaeology, more efficient and safer for the content being excavated.

This project was done in collaboration with the Research Laboratory on Archeology and History, UMR CReAAH, UMR Trajectoires, and Inrap.

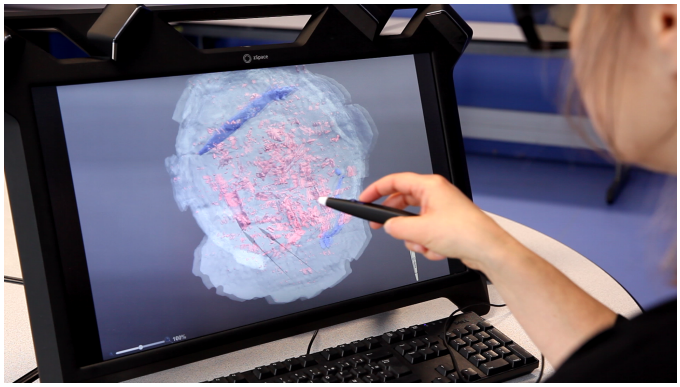


Figure 24. The INSIDE system used by an archaeologist within a workbench system.

VR Interactions with Multiple Interpretations of Archaeological Artefacts

Participants: Ronan Gaugne and Valérie Gouranton

The incorporation of 3D printed artefacts into Virtual Reality and Augmented Reality experiences is gaining strong interest from Cultural Heritage professionals. Indeed, in most cases, virtual environments cannot convey information such as the physical properties of artefacts. In [37], we presented a methodology for the development of VR experiences which incorporate 3D replicas of artefacts as user interfaces. The methodology is applied on the development of an experience to present various interpretations of an urn which was found at the edge of a cliff on the south east coastal area of the United Kingdom in 1910. In order to support the understanding of the multiple interpretations of this artefact, the system deploys a virtual environment and a physical replica to allow users to interact with the artefacts and the environment (see Figure 25). Feedback from heritage users suggests VR technologies along with digitally fabricated replicas can meaningfully engage audiences with multiple interpretations of cultural heritage artefacts.

This project was done in collaboration with University of Brighton (UK), Inrap, CNRS and UMR CReAAH.

Figure 25. Interaction in VR using the physical replica of a funeral urn.

ILDA Project-Team

6. New Results

6.1. Gestures and Tangibles

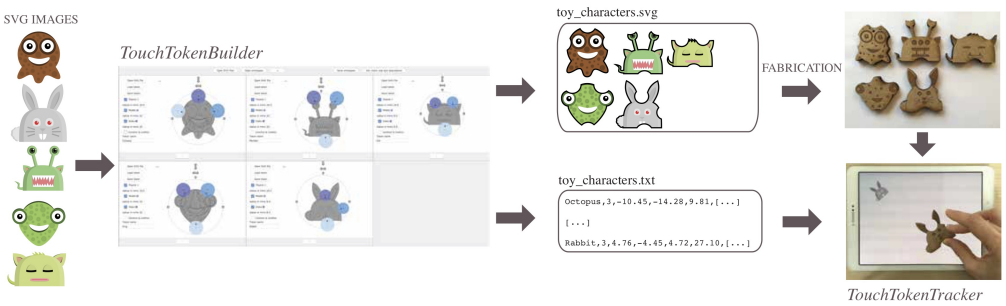


Figure 4. TouchToken-Builder (left) assists users in placing grasping notches on arbitrarily-shaped tokens, warning them about spatial configurations that could generate recognition conflicts or that might be uncomfortable to manipulate. It outputs both a vector and a numerical description of the tokens' geometry (middle). Those are used respectively to build the tokens (top-right), and to track them on any touchscreen using TouchToklen-Tracker (bottom-right).

6.1.1. Custom-made Tangible Interfaces with TouchTokens

One of our main results in this area is the design, development and evaluation of TouchTokens, a new way of prototyping and implementing low-cost tangible interfaces [6]. The approach requires only passive tokens and a regular multi-touch surface. The tokens constrain users' grasp, and thus, the relative spatial configuration of fingers on the surface, theoretically making it possible to design algorithms that can recognize the resulting touch patterns. Our latest project on TouchTokens [17] has been about tailoring tokens, going beyond the limited set of geometrical shapes studied in [6], as illustrated in Figure 4 .

6.1.2. Designing Coherent Gesture Sets for Multi-scale Navigation on Tabletops

We designed a framework for the study of multi-scale navigation (Figure 5) and conducted a controlled experiment of multi-scale navigation on tabletops [25]. We first conducted a guessability study in which we elicited user-defined gestures for triggering a coherent set of navigation actions, and then proposed two interface designs that combine the now-ubiquitous slide, pinch and turn gestures with either two-hand variations on these gestures, or with widgets. In a comparative study, we observed that users can easily learn both designs, and that the gesture-based, visually-minimalist design is a viable option, that saves display space for other controls.

6.1.3. Command Memorization, Gestures and other Triggering Methods

In collaboration with Telecom ParisTech, we studied the impact of semantic aids on command memorization when using either on-body interaction or directional gestures [21]. Previous studies had shown that spatial memory and semantic aids can help users learn and remember gestural commands. Using the body as a support to combine both dimensions had therefore been proposed, but no formal evaluations had been reported. We compared, with or without semantic aids, a new on-body interaction technique (BodyLoci) to mid-air Marking menus in a virtual reality context, considering three levels of semantic aids: no aid, story-making, and story-making with background images.

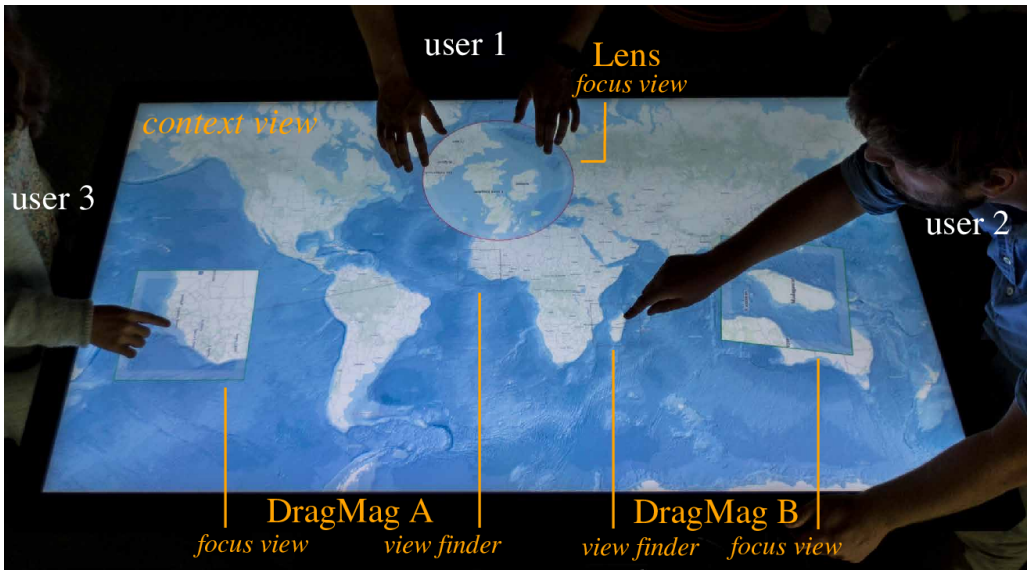


Figure 5. Our framework for the study of multi-scale navigation on tabletops enables users to both pan & zoom the context view and to create independent focus views, either DragMags or lenses.

As part of the same collaboration, we also studied how memorizing positions or directions affects gesture learning for command selection. Many selection techniques either rely on directional gestures (e.g. Marking menus) or pointing gestures using a spatially-stable arrangement of items (e.g. FastTap). Both types of techniques are known to leverage memorization, but not necessarily for the same reasons. We investigated whether using directions or positions affects gesture learning [20].

6.2. Interacting with the Semantic Web of Linked Data

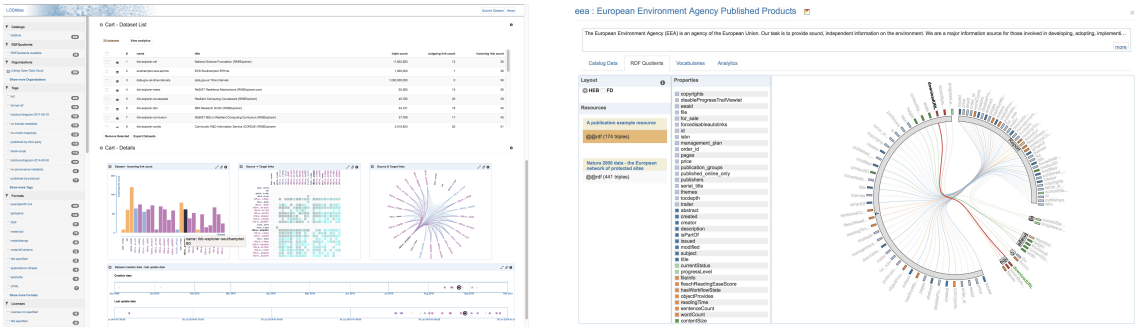


Figure 6. Browsing Linked Data Catalogs with LODAtlas [22]. Left: Visualization of the characteristics of, and links between, datasets selected by the user. Right: RDFQuotients-derived visual summary of a dataset. The summary shows how properties relate instances of the different classes.

The Web of Data is growing fast, as exemplified by the evolution of the Linked Open Data (LOD) cloud over the last ten years. One of the consequences of this growth is that it is becoming increasingly difficult for application developers and end-users to find the datasets that would be relevant to them. Semantic Web search engines, open data catalogs, datasets and frameworks such as LODStats and LOD Laundromat, are all useful but only give partial, even if complementary, views on what datasets are available on the Web. We started working on a platform called LODAtlas in 2016. LODAtlas [22] is a portal that enables users to find datasets of interest (see Figure 6) . Users can make different types of queries about both the datasets' metadata and contents, aggregated from multiple sources. They can then quickly evaluate the matching datasets' relevance, thanks to summary visualizations of their general metadata, connections and contents. The latter has been developed in collaboration with project-team CEDAR, based on their recent work on RDF Quotients.

Linked Data is structured as a directed labeled graph, or more precisely as a multitude of such graphs, that can be interlinked and distributed over the World Wide Web. Graph structures play an essential role at different scales in the Web of Data, and while it is now clear that basic approaches based on node-link diagram representations are only useful for small datasets, such visualizations remain meaningful for the representation of subsets of these multi-variate data. As part of a larger effort that started in the summer of 2016 to investigate novel interactive visual exploration techniques for multi-variate graphs, we introduced a design space and Web-based framework for generating what we call *animated edge textures*. Network edge data attributes are usually encoded using color, opacity, stroke thickness and stroke pattern, or some combination thereof. But in addition to these static variables, it is also possible to animate dynamic particles flowing along the edges. These can be seen as animated edge textures, that offer additional visual encodings that have potential not only in terms of visual mapping capacity but also playfulness and aesthetics. While such particle-based visual encodings have been featured in several commercial and design-oriented visualizations, this has to our knowledge almost always been done in a relatively ad hoc manner. Beyond the design space and Web framework, we also conducted an initial evaluation of particle properties – particle speed, pattern and frequency – in terms of visual perception. This work [24] was performed in collaboration with Nathalie Henry-Riche from Microsoft Research and Benjamin Bach from Edimburgh University.

6.3. Visualization

A significant part of our activity in this axis has been dedicated to geovisualization for various surfaces, including desktop workstations, tabletops and wall displays, in the context of ANR project MapMuxing. We investigated the representation of time in geovisualizations, more particularly how to convey changes in satellite images. Before-and-after images show how entities in a given region have evolved over a specific period of time. These images are used both for data analysis and to illustrate the resulting findings to diverse audiences. We introduced Baia [4], a framework to create advanced animated transitions, called animation plans, between before-and-after images. Baia relies on a pixel-based transition model that gives authors much expressive power, while keeping animations for common types of changes easy to create thanks to predefined animation primitives (Figures 7 and 2).

Still in the area of geovisualization, in the context of ADT project Seawall, conducted in collaboration with project-team Lemon at Inria SAM / Montpellier and with Inria Chile, we have participated to the 2018 SciVis contest, which this year was about the visualization of data related to tsunamis generated by the impact of asteroids in deep water [31]. We used the WILDER ultra-high-resolution wall display to make it easier for analysts to visually compare and contrast different simulations from a deep water asteroid impact ensemble dataset. See Section 5.7.1 and Figure 3 .

In the area of scientific data analysis, we have been collaborating with neuroscientists that explore large quantities of EEG data at different temporal scales. As a first step, we explored if automated algorithmic processes, that aid in the search for similar patterns in large datasets, actually match human intuition. We studied if we perceive as similar the results of these automatic measures, using three time-series visualizations: line charts, horizon graphs and colorfields. Our findings [15], [30] indicate that the notion of similarity is visualization-dependent, and that the best visual encoding varies depending on the automatic similarity measure considered.

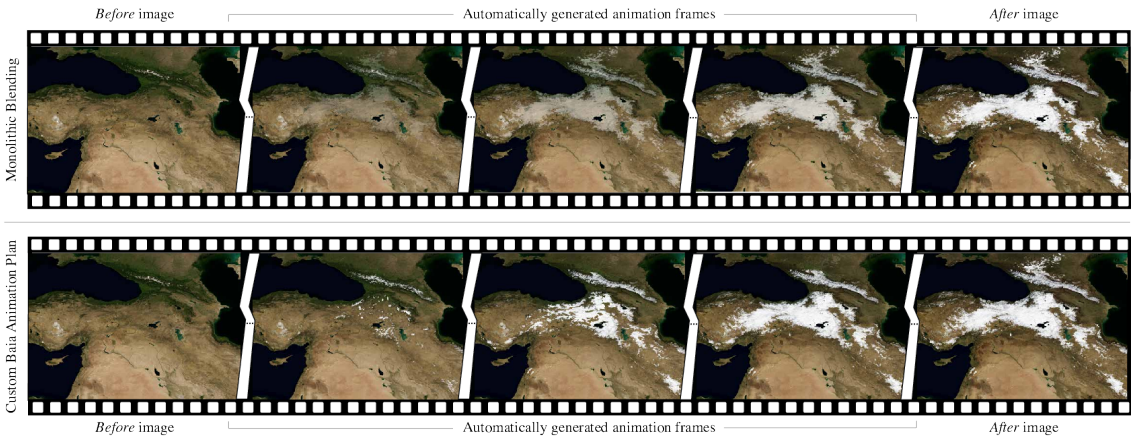


Figure 7. Animated transitions [4] based on one single before-and-after image pair showing seasonal snow cover over northern Middle East. The top row shows keyframes generated using basic monolithic blending. Snow fades in gradually but uniformly, regardless of altitude. The bottom row shows keyframes generated using a Baia animation plan derived from a Digital Elevation Model. Snow fades in gradually, but this time spreading from high-altitude to low-altitude areas.

Anstasia Bezerianos co-advised the PhD work of Evanthia Dimara in project-team Aviz together with P. Dragicevic. Last year, they had already confirmed that the cognitive bias known as the *attraction effect* does exist in visualizations [49]. This was followed-up this year by an exploration of different ways to mitigate this bias [12] (in collaboration with Northwestern University and Sorbonne Université). It was observed that the approach that consists of deleting all unwanted alternatives interactively removed the bias, a result that previous research has shown to be extremely hard to achieve. They also explored how different interactive visualizations of multidimensional datasets can affect decision making [13], and created a task-based taxonomy of cognitive biases for information visualization [14].

Our collaboration with INRA researchers has focused on mixed-initiative systems that combine human learning, machine learning and evolution. Results in this area for this year include an interactive evolutionary algorithm to learn from user interactions and steer the exploration of multidimensional datasets towards two-dimensional projections that are interesting to the analyst, and guidelines on how to evaluate such mixed initiative systems [29].

6.4. Collaboration, Multi-display environments, Large and Small Displays

We studied awareness techniques to aid transitions between personal and shared workspaces in multi-display environments, that include large shared displays and desktops (Figure 8). In such contexts, including crisis management and control rooms, users can engage in both close collaboration and parallel or personal work. Transitioning between different displays can be challenging. To provide workspace awareness and to facilitate these transitions, we designed and implemented three interactions techniques that display users' activities. We explored how and where to display this activity: briefly on the shared display, or more persistently on a peripheral floor display. In a user study motivated by the context of a crisis room where multiple operators with different roles need to cooperate, we tested the usability of the techniques and provided insights on such transitions in systems running on MDEs [23]. We also contributed on a book chapter discussing how to best support collaboration in immersive environments that can range from MDE to mixed reality ones [28].

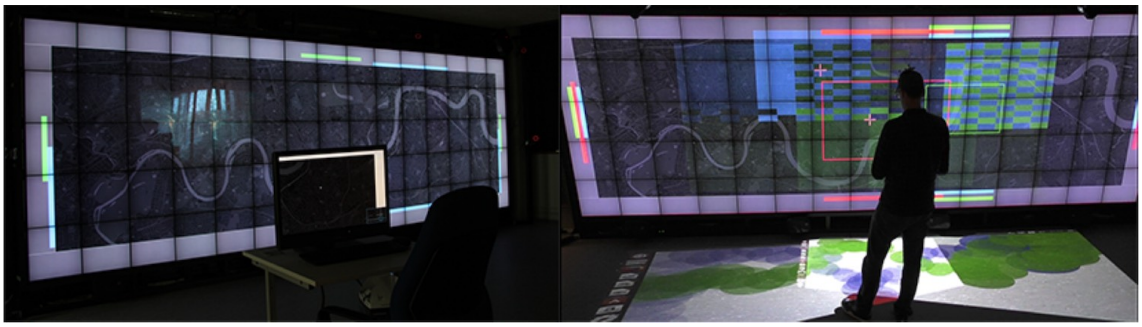


Figure 8. (left) Multi-Display Environment composed of a wall display and two workstations (one visible in the photo). (right) Three workspace awareness techniques: Awareness Bars at the edges of the wall, Focus Map on the wall display, and Step Map projected on the ground.

We collaborated with members from Inria project-team Aviz on the topic of small-scale visualization. This year, new results include a study about the perception of visualizations on smartwatches, performed together with Microsoft Research [11], [26]. The study was designed to assess how quickly people can perform a simple data comparison task for small-scale visualizations on a smartwatch. The goal was to extend our understanding of design constraints for smartwatch visualizations. We tested three chart types common on smartwatches: bar charts, donut charts, and radial bar charts with three different data sizes: 7, 12, and 24 data values. Results show that bar and donut charts should be preferred on smartwatch displays when quick data comparisons are necessary.

IMAGINE Project-Team

7. New Results

7.1. Sculpting Mountains: Interactive Terrain Modeling Based on Subsurface Geology

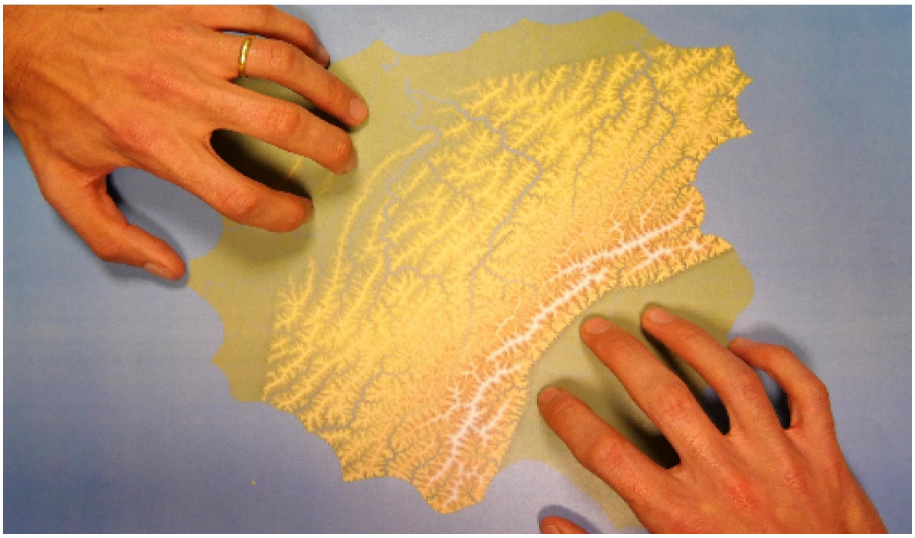


Figure 2. Sculpting Mountains: Interactive Terrain Modeling Based on Subsurface Geology

Most mountain ranges are formed by the compression and folding of colliding tectonic plates. Subduction of one plate causes large-scale asymmetry while their layered composition (or stratigraphy) explains the multi-scale folded strata observed on real terrains. As part of Guillaume Cordonnier's PhD thesis, we introduced a novel interactive modeling technique to generate visually plausible, large scale terrains that capture these phenomena (illustrated in Fig. 2). Our method draws on both geological knowledge for consistency and on sculpting systems for user interaction. The user is provided hands-on control on the shape and motion of tectonic plates, represented using a new geologically-inspired model for the Earth crust. The model captures their volume preserving and complex folding behaviors under collision, causing mountains to grow. It generates a volumetric uplift map representing the growth rate of subsurface layers. Erosion and uplift movement are jointly simulated to generate the terrain. The stratigraphy allows us to render folded strata on eroded cliffs. We validated the usability of our sculpting interface through a user study, and compare the visual consistency of the earth crust model with geological simulation results and real terrains.

7.2. Exploratory design of mechanical devices with motion constraints

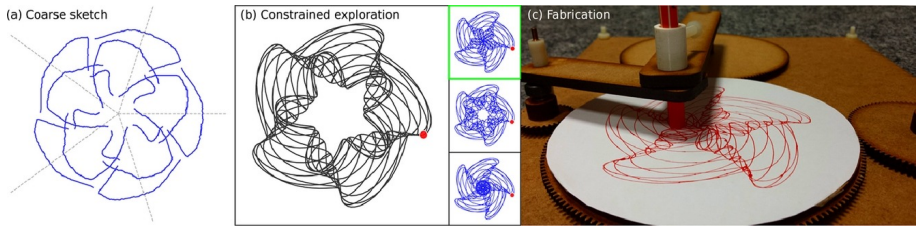


Figure 3. Exploratory design of mechanical devices with motion constraints. (a) The user first selects a mechanically feasible drawing by providing a rough sketch. (b) The user is then able to interactively explore local alternatives (b) by defining visual constraints directly on the pattern (here, the cusp position). (c) The resulting machine is automatically exported to laser cutter profiles for fabrication.

Mechanical devices are ubiquitous in our daily lives, and the motion they are able to transmit is often a critical part of their function. While digital fabrication devices facilitate their realization, motion-driven mechanism design remains a challenging task. We take drawing machines as a case study in exploratory design. Devices such as the Spirograph can generate intricate patterns from an assembly of simple mechanical elements. Trying to control and customize these patterns, however, is particularly hard, especially when the number of parts increases. We propose a novel constrained exploration method that enables a user to easily explore feasible drawings by directly indicating pattern preferences at different levels of control. This is (illustrated in Fig. 3). The user starts by selecting a target pattern with the help of construction lines and rough sketching, and then fine-tunes it by prescribing geometric features of interest directly on the drawing. The designed pattern can then be directly realized with an easy-to-fabricate drawing machine. The key technical challenge is to facilitate the exploration of the high dimensional configuration space of such fabricable machines. To this end, we propose a novel method that dynamically reparameterizes the local configuration space and allows the user to move continuously between pattern variations, while preserving user-specified feature constraints. We tested our framework on several examples, conducted a user study, and fabricated a sample of the designed examples.

7.3. Automatic Generation of Geological Stories from a Single Sketch

Describing the history of a terrain from a vertical geological cross-section is an important problem in geology, called geological restoration. Designing the sequential evolution of the geometry is usually done manually, involving many trials and errors. In this work, we recast this problem as a storyboarding problem, where the diifferent stages in the restoration are automatically generated as storyboard panels and displayed as geological stories. Our system allows geologists to interactively explore multiple scenarios by selecting plausible geological event sequences and backward simulating them at interactive rate, causing the terrain layers to be progressively un-deposited, un-eroded, un-compacted, unfolded and un-faulted. Storyboard sketches are generated along the way. When a restoration is complete, the storyboard panels can be used for automatically generating a forward animation of the terrain history, enabling quick visualization and validation of hypotheses. As a proof-of-concept, we describe how our system was used by geologists to restore and animate cross-sections in real examples at various spatial and temporal scales and with diffierent levels of complexity, including the Chartreuse region in the French Alps.

7.4. 3D Shape Decomposition and Sub-parts Classification

This paper (illustrated in Fig. 5) introduces a measure of significance on a curve skeleton of a 3D piecewise linear shape mesh, allowing the computation of both the shape's parts and their saliency. We begin by reformulating three existing pruning measures into a non-linear PCA along the skeleton. From this PCA,

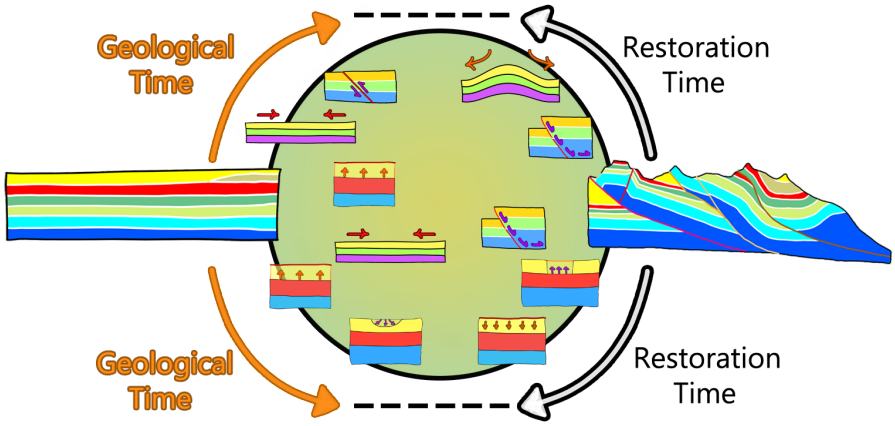


Figure 4. Automatic Generation of Geological Stories from a Single Sketch. From left to right, the original terrain from several million years ago undergoes events that will transform it to its current state. From right to left, the current terrain is restored and undergoes undo events that will transform it back to its original state.

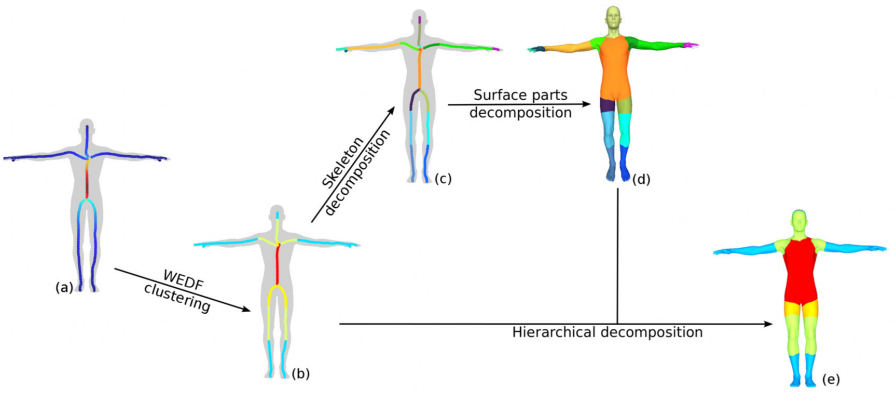


Figure 5. 3D Shape Decomposition and Sub-parts Classification. Starting from a 3D shape and its curve skeleton, we compute a new measure called WEDF on the curve skeleton (a) and, by clustering WEDF values, we decompose the skeleton into hierarchical parts (b). To each connected part on the skeleton –shown with a different color (c)– a connected region of the surface mesh is assigned (d). Then, a salience value according to the hierarchy is assigned to each corresponding surface part (e) –parts of same importance get a similar color.

we then derive a volume-based salience measure, the 3D WEDF, that determines the relative importance to the global shape of the shape part associated to a point of the skeleton. First, we provide robust algorithms for computing the 3D WEDF on a curve skeleton, independent on the number of skeleton branches. Then, we cluster the WEDF values to partition the curve skeleton, and coherently map the decomposition to the associated surface mesh. Thus, we develop an unsupervised hierarchical decomposition of the mesh faces into visually meaningful shape regions that are ordered according to their degree of perceptual salience. The shape analysis tools introduced in this paper are important for many applications including shape comparison, editing, and compression.

7.5. Interactive Generation of Time-evolving, Snow-Covered Landscapes with Avalanches

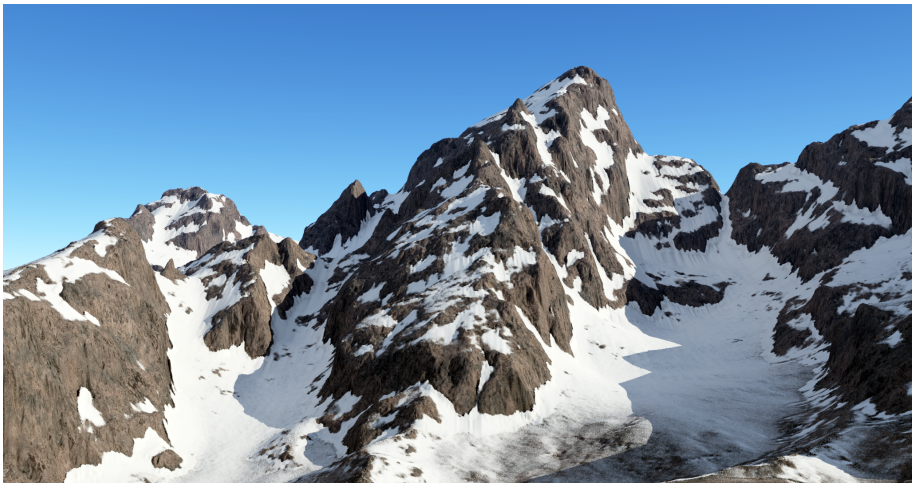


Figure 6. Interactive Generation of Time-evolving, Snow-Covered Landscapes with Avalanches

As part of Guillaume Cordonnier's PhD thesis, we also introduced a novel method for interactive generation of visually consistent, snow-covered landscapes, which provides control of their dynamic evolution over time. Our main contribution (illustrated in Fig. 6) was the real-time phenomenological simulation of avalanches and other user-guided events, such as tracks left by Nordic skiing, which can be applied to interactively sculpt the landscape. The terrain is modeled as a height field with additional layers for stable, compacted, unstable, and powdery snow, which behave in combination as a semi-viscous fluid. We incorporate the impact of several phenomena, including sunlight, temperature, prevailing wind direction, and skiing activities. The snow evolution includes snow-melt and snowdrift, which affect stability of the snow mass and the probability of avalanches. A user can shape landscapes and their evolution either with a variety of interactive brushes, or by prescribing events along a winter season time-line. Our optimized GPU-implementation allows interactive updates of snow type and depth across a large (10×10 km) terrain, including real-time avalanches, making this suitable for visual assets in computer games. We evaluated our method through perceptual comparison against exiting methods and real snow-depth data.

LOKI 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. Although our strategy is to investigate issues and address challenges for all of the three levels of dynamics, our immediate priority is to focus on micro-dynamics since it concerns very fundamental knowledge about interaction and relates to very low-level parts of interactive systems, which is likely to influence our future research and developments at other levels.

7.2. Micro-dynamics

Participants: Axel Antoine, Géry Casiez [correspondent], 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 introduced a new prediction algorithm for direct touch devices based on (i) a state-of-the-art finite-time derivative estimator, (ii) a smoothing mechanism based on input speed, and (iii) a post-filtering of the prediction in two steps (see Figure 2 left). Using both a preexisting dataset of touch input as benchmark, and subjective data from a new user study, we showed that this new predictor outperforms those currently available in the literature and industry, based on metrics that model user-defined negative side-effects caused by input prediction. In particular, our predictor can predict up to 2 or 3 times further than existing techniques with minimal negative side-effects [23].

We also proposed a hybrid hardware and software input prediction technique specifically designed for partially compensating end-to-end latency in indirect pointing (see Figure 2 right). We combined a computer mouse with a high frequency accelerometer to predict the future location of the pointer using Euler based equations. Our prediction method results in more accurate prediction than previously introduced prediction algorithms for direct touch. A controlled experiment also revealed that it can improve target acquisition time in pointing tasks [15], [28].

Finally, on the topic of transfer functions we performed some preliminary analysis of the kinematics of a pointing task with varying linear velocity based transfer functions to assess how we use vision and haptics to plan and control our movement [25].

7.2.2. Understanding touch interaction

Atomic interactions in touch interfaces, like tap, drag, and flick, are well understood in terms of interaction design, but less is known about their physical performance characteristics. We conducted a study to gather baseline data about finger pitch and roll orientation during atomic touch input actions [21]. Our results showed differences in orientation and range for different fingers, hands, and actions: for a given hand, the little, ring and middle fingers are used in a similar manner, whereas the thumb uses different range of orientations. Additional analyses about how changing the angle of the tablet affects people's finger orientations suggest that

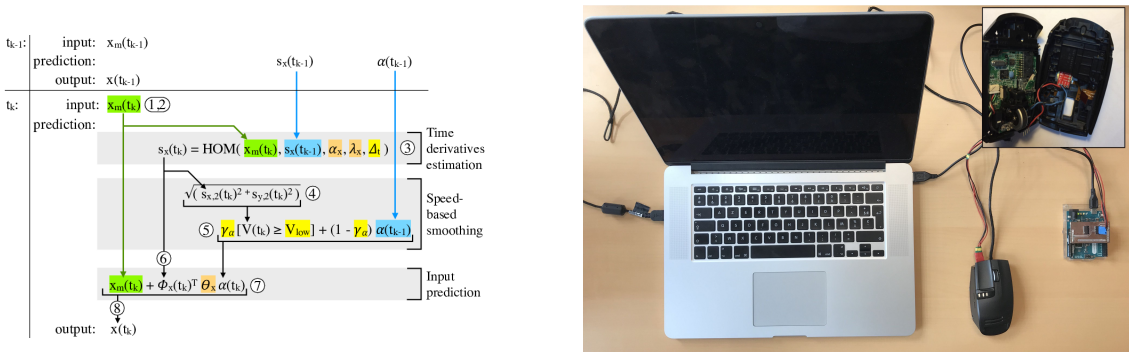


Figure 2. (left) General description of our real-time input prediction method, with step numbers. Input in green, previously computed variables in blue, general parameters in yellow, optimized parameters in orange. (right) Our hybrid setup for input prediction comprises a Logitech G9 Laser Mouse connected via USB to the host computer with the MPU-9250 chip embedded inside, which is itself connected to an Arduino board.

ranges of orientation tighten as the tablet pitch increases. This data provides designers and researchers with better understanding of what kind of interactions are possible in different settings (e. g., using the left or right hand), to design novel interaction techniques that use orientation as input (e. g., using finger tilt as an implicit mode), and to anticipate the feasibility of new sensing techniques (e. g., using fingerprints for identifying specific finger touches).

7.3. Meso-dynamics

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

7.3.1. Improving interaction bandwidth and expressiveness

Despite the ubiquity of touch-based input and the availability of increasingly computationally powerful touchscreen devices, there has been comparatively little work on enhancing basic canonical gestures such as swipe-to-pan and pinch-to-zoom. We introduced transient pan and zoom, i. e., pan and zoom manipulation gestures that temporarily alter the view and can be rapidly undone [16]. Leveraging typical touchscreen support for additional contact points, we designed our transient gestures so that they co-exist with traditional pan and zoom interaction. In addition to reducing repetition in multi-level navigation, our transient pan-and-zoom also facilitates rapid movement between document states.

Image editing software feature various pixel selection tools based on geometrical (rectangle, ellipses, polygons) or semantical (magic wand, selection brushes) data from the image. They are efficient in many situations, but are limited when selecting bitmap representations of handwritten text for e. g., interpreting scanned historical documents that cannot be reliably analyzed by automatic OCR methods: strokes are thin, with many overlaps and brightness variations. We have designed a new selection tool dedicated to this purpose [27]: a cursor based brush selection tool with two additional degrees of freedom: brush size and brightness threshold. The brush cursor displays feedforward clues that indicates the user which pixels will be selected upon pressing the mouse button. This brush provides a fine grain control to the user over the selection.

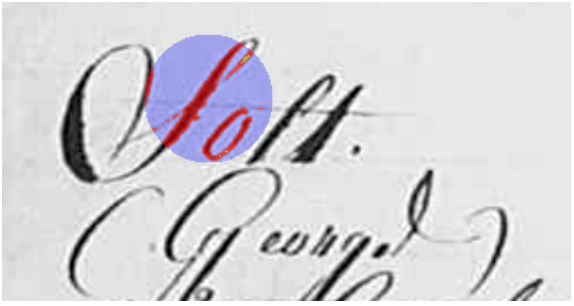


Figure 3. A four-dimensional selection brush for digitized handwritten documents. Red pixels will be selected, blue pixels will not.

7.3.2. Interacting with specific setups (Large-Displays, Virtual & Augmented Reality)

Large displays are becoming commonplace at work, at home, or in public areas. Handheld devices such as smartphones and smartwatches are ubiquitous, but little is known on regarding how these devices could be used to point at remote large displays. We conducted a survey on possession and use of smart devices, as well as a controlled experiment comparing seven distal pointing techniques on phone or watch, one- and two-handed, and using different input channels and mappings [26]. Our results favor using a smartphone as a trackpad, but also explore performance tradeoffs that can inform the choice and design of distal pointing techniques for different contexts of use.

In virtual reality environments, raycasting is the most common target pointing technique. However, performance on small and distant targets is impacted by the accuracy of the pointing device and the user's motor skills. Existing pointing facilitation techniques are currently only applied in the context of a virtual hand, i. e., for targets within reach. We studied how a user-controlled cursor could be added on the ray in order to enable target proximity-based pointing techniques –such as the Bubble Cursor– to be used for targets that are out of reach [17]. We conducted a study comparing several visual feedbacks for this technique (see Figure 4). Our results showed that simply highlighting the nearest target reduces the selection time by 14.8% and the error rate by 82.6% compared to standard Raycasting. For small targets, the selection time is reduced by 25.7% and the error rate by 90.8%.

Brain-Computer Interfaces (BCIs) enable users to interact with computers without any dedicated movement, bringing new hands-free interaction paradigms that could be beneficial in an Augmented Reality (AR) setup. We first tested the feasibility of using BCI in AR settings based on Optical See-Through Head-Mounted Displays (OST-HMDs) [12]. Experimental results showed that a BCI and an OST-HMD equipment (EEG headset and Hololens in our case) are well compatible and that small movements of the head can be tolerated when using the BCI. Then, we introduced a design space for command display strategies based on BCI in AR, when exploiting a famous brain pattern called Steady-State Visually Evoked Potential (SSVEP). Our design space relies on five dimensions concerning the visual layout of the BCI menu: orientation, frame-of-reference, anchorage, size and explicitness. We implemented various BCI-based display strategies and tested them within the context of mobile robot control in AR. Our findings were finally integrated within an operational prototype based on a real mobile robot that is controlled in AR using a BCI and a HoloLens headset. Taken together, our results (4 user studies) and our methodology could pave the way to future interaction schemes in Augmented Reality exploiting 3D User Interfaces based on brain activity and BCIs.

More generally, we also contributed to a reflexion on the complexity and scientific challenges associated to virtual and augmented realities [29] and the challenges to make virtual environments more closely related to the real world [30].

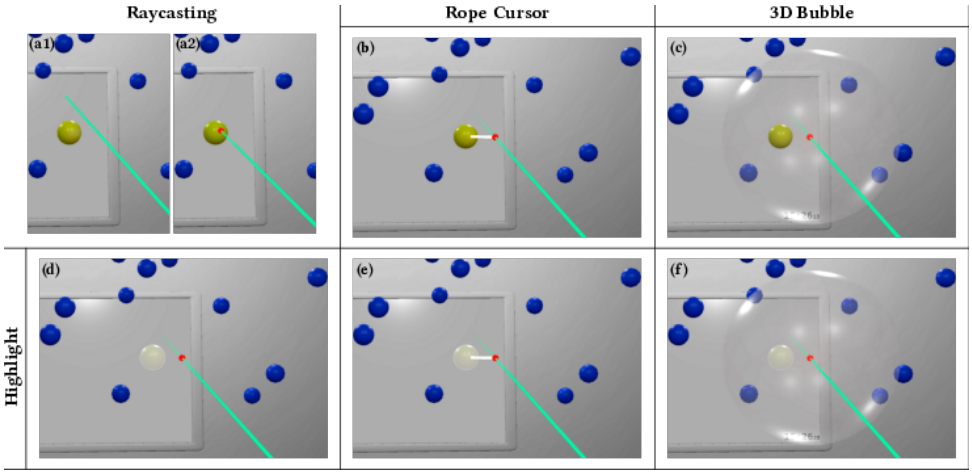


Figure 4. Visual feedback for RayCursor: (a1,a2) classical Raycasting; (b) Rope Cursor: a stroke between the closest target and the cursor; (c) 3D Bubble: a bubble centered on the cursor which contains the nearest target; (d) Highlighting on the nearest target; (e,f), highlight + rope and 3D Bubble.

7.3.3. Tools for prototyping and programming interaction

Touch interactions are now ubiquitous, but few tools are available to help designers quickly prototype touch interfaces and predict their performance. On one hand, for rapid prototyping, most applications only support visual design. On the other hand, for predictive modeling, tools such as CogTool generate performance predictions but do not represent touch actions natively and do not allow exploration of different usage contexts. To combine the benefits of rapid visual design tools with underlying predictive models, we developed the *Storyboard Empirical Modeling (StEM)* tool [20], [19] for exploring and predicting user performance with touch interfaces (see Figure 5). StEM provides performance models for mainstream touch actions, based on a large corpus of realistic data. We evaluated StEM in an experiment and compared its predictions to empirical times for several scenarios. The study showed that our predictions are accurate (within 7% of empirical values on average), and that StEM correctly predicted differences between alternative designs. Our tool provides new capabilities for exploring and predicting touch performance, even in the early stages of design.

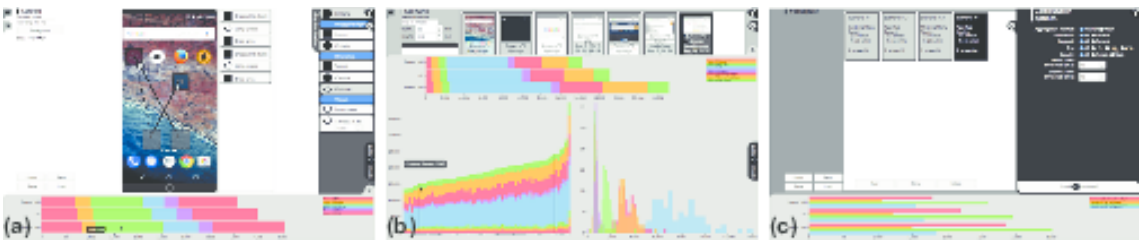


Figure 5. Storyboard Empirical Modeling (StEM): (a) users drag and drop actions onto a timeline to construct an interaction sequence; (b) users can visualize prediction times for a scenario composed of different screens; (c) users can compare scenarios, and filter the predictions according to contextual factors such as screen size or user's expertise.

Following our main objective of revisiting interactive system, we have also proposed two systems for defining and programming interactive behaviors and interactions.

Much progress has been made on interactive behavior development tools for expert programmers. However, less effort has been made in investigating how these tools support creative communities who typically struggle with technical development. This is the case, for instance, of media artists and composers working with interactive environments. To address this problem, we have introduces ZenStates [18], a new specification model for creative interactive environments that combines Hierarchical Finite-States Machines, expressions, off-the-shelf components called Tasks, and a global communication system called the Blackboard. We have implemented our model in a direct manipulation-based software interface and probed ZenStates' expressive power through 90 exploratory scenarios. we have also conducted a user study to investigate the understandability of ZenStates' model. Results support ZenStates viability, its expressiveness, and suggest that ZenStates is easier to understand –in terms of decision time and decision accuracy– compared to popular alternatives such as standard object-oriented programming and a data-flow visual language.

In a more general context, we have introduced a new GUI framework based on the *Entity-Component-System* model (ECS), where interactive elements (Entities) can acquire any data (Components) [24]. Behaviors are managed by continuously running processes (Systems) which select entities by the components they possess. This model facilitates the handling and reuse of behaviors. It allows to define the interaction modalities of an application globally, by formulating them as a set of Systems. We have implemented an experimental toolkit based on this approach, *Polyphony*, in order to demonstrate the use and benefits of this model.

7.4. Macro-dynamics

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

One conspicuous feature of the current evolution of interactive devices is the spread of touch-sensitive surfaces. Typically, modern smartphones are equipped with such touch-sensitive surfaces that also support normal force-based input capabilities, which can for instance be used to control the range of a text selection by varying the force applied to the touchscreen (on e. g., iOS devices). However, this interaction mechanism is difficult to discover and many users simply ignore it exists. To overcome this problem, we introduced ForceSelect (see Figure 6 , left), a force-based text selection techniques that relies on a simple mode gauge (see Figure 6 , right) that does not require additional screen real-estate and help users to discover and master the use of force input in text selection tasks [22]. We conducted two studies that suggest that this mode gauge successfully provides enhanced discoverability of the force-based input and combines support for novices and experts, whereas it was never worse than the standard iOS technique and was also preferred by participants.

7.5. Interaction Machine

Several of our new results this year contributed to our global objective of building an Interaction Machine, especially at the micro-dynamics level. Our work on prediction algorithms and our hybrid hardware-software latency compensation method 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 the characterization of the dimensions of touch interaction, especially angle of touch, highlighted the need for additional dimensions in input events that are not yet accessible in actual systems. All in all, this confirm our hypothesis that we have to redefine input management and input events propagation in order to better account for human factors in interactive systems, to extend the possibilities for designing more efficient and expressive interaction methods.

At the meso-dynamics level, our work on improving basic interaction methods in non-standard setups (e. g., VR, AR) highlighted the need for more open and flexible system architectures and tools that ease the design and prototyping of alternative interaction techniques based on mixed modalities. The new prototyping and programing tools that we proposed this year (StEM, ZenStates and Polyphony) are our first explorations toward such system-integrated frameworks dedicated to interaction.

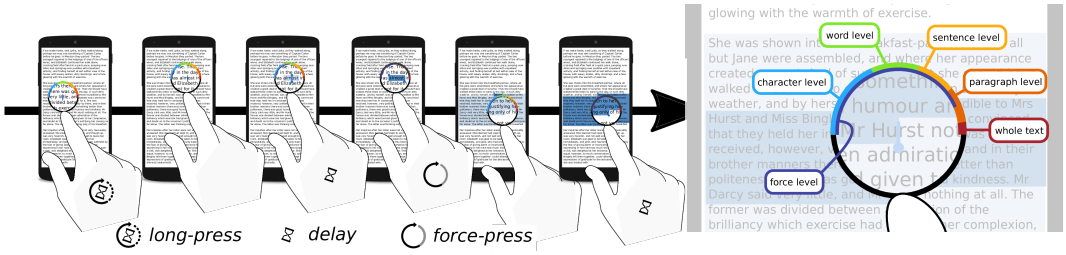


Figure 6. (left) Example of text selection using ForceSelect. The user performs a long-press that displays the callout magnifier. Keeping the force in the character level, the user adjusts its position by moving her finger. She then holds the force in the word level of the "mode gauge", locks the selection and enters the clutch mode. When force-pressing to the whole text level of the "mode gauge", she un-clutches the selection and updates it.; (right) Close-up of the "mode gauge". There are two types of text highlighting in the background: dark highlighting covers between both handles and light highlighting acts as a feedforward of which portion of text will be selected if the user released her finger (here the whole paragraph).

MANAO Project-Team

7. New Results

7.1. Analysis and Simulation

7.1.1. Visual Features in the Perception of Liquids

Perceptual constancy—identifying surfaces and objects across large image changes—remains an important challenge for visual neuroscience. Liquids are particularly challenging because they respond to external forces in complex, highly variable ways, presenting an enormous range of images to the visual system. To achieve constancy, the brain must perform a causal inference that disentangles the liquid's viscosity from external factors—like gravity and object interactions—that also affect the liquid's behavior. Here, we tested whether the visual system estimates viscosity using "midlevel" features that respond more to viscosity than other factors. Our findings demonstrate that the visual system achieves constancy by representing stimuli in a multidimensional feature space—based on complementary, midlevel features—which successfully cluster very different stimuli together and tease similar stimuli apart, so that viscosity can be read out easily.

7.1.2. Teaching Spatial Augmented Reality: a Practical Assignment for Large Audiences

We conceived a new methodology to teach spatial augmented reality in a practical assignment to large audiences. Our approach does not require specific equipment such as video projectors while teaching the principal topics and difficulties involved in spatial augmented reality applications, and especially calibration and tracking. The key idea is to set up a scene graph consisting of a 3D scene with a simulated projector that "projects" content onto a virtual representation of the real-world object. For illustrating the calibration, we simplify the intrinsic parameters to using the field of view, both for the camera and the projector. For illustrating the tracking, instead of relying on specific hardware or software, we exploit the relative transformations in the scene graph.

7.2. From Acquisition to Display

7.2.1. Comparison of Plenoptic Imaging Systems

Plenoptic cameras provide single-shot 3D imaging capabilities, based on the acquisition of the Light-Field, which corresponds to a spatial and directional sampling of all the rays of a scene reaching a detector. Specific algorithms applied on raw Light-Field data allow for the reconstruction of an object at different depths of the scene. Two different plenoptic imaging geometries have been reported, associated with two reconstruction algorithms: the traditional or unfocused plenoptic camera, also known as plenoptic camera 1.0, and the focused plenoptic camera, also called plenoptic camera 2.0. Both systems use the same optical elements, but placed at different locations: a main lens, a microlens array and a detector. These plenoptic systems have been presented as independent. We have demonstrated the continuity between them, by simply moving the position of an object. We have also compared the two reconstruction methods. We have finally theoretically shown that the two algorithms are intrinsically based on the same principle and could be applied to any Light-Field data. However, the resulting images resolution and quality depend on the chosen algorithm.

7.2.2. Capturing Illumination for Augmented Reality using RGB-D Images

RGB-D sensors is becoming more and more available. We have proposed an automatic framework to recover the illumination (from light sources both in and out of the camera's view) of indoor scenes based on a single RGB-D image. Unlike previous works, our method can recover spatially varying illumination without using any lighting capturing devices or HDR information. The recovered illumination can produce realistic rendering results. Using the estimated light sources and geometry model, environment maps at different points in the scene are generated that can model the spatial variance of illumination. The experimental results have demonstrated the validity of our approach and the possibilities offered to Augemented Reality by the use of more dedicated hardware.

7.2.3. Diffraction Removal in an Image-based BRDF Measurement Setup

Material appearance is traditionally represented through its Bidirectional Reflectance Distribution Function (BRDF), quantifying how incident light is scattered from a surface over the hemisphere. To speed up the measurement process of the BRDF for a given material, which can necessitate millions of measurement directions, image-based setups are often used for their ability to parallelize the acquisition process: each pixel of the camera gives one unique configuration of measurement. With highly specular materials, the High Dynamic Range (HDR) imaging techniques are used to acquire the whole BRDF dynamic range, which can reach more than 10 orders of magnitude. Unfortunately, HDR can introduce star-burst patterns around highlights arising from the diffraction by the camera aperture. Therefore, while trying to keep track on uncertainties throughout the measurement process, one has to be careful to include this underlying diffraction convolution kernel. A purposely developed algorithm is used to remove most part of the pixels polluted by diffraction, which increase the measurement quality of specular materials, at the cost of discarding an important amount of BRDF configurations (up to 90% with specular materials). Finally, our setup succeed to reach a 1.5 degree median accuracy (considering all the possible geometrical configurations), with a repeatability from 1.6% or the most diffuse materials to 5.5% for the most specular ones. Our new database, with their quantified uncertainties, will be helpful for comparing the quality and accuracy of the different experimental setups and for designing new image-based BRDF measurement devices.

7.3. Rendering, Visualization and Illustration

7.3.1. A View-Dependent Metric for Patch-Based LOD Generation & Selection

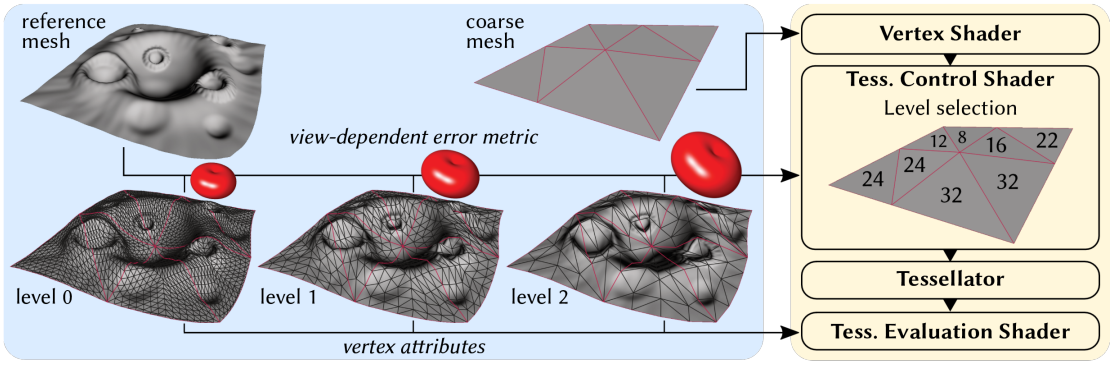


*Figure 8. Full processing pipeline — **(a)** as a pre-process, the LOD is generated from the reference mesh by decimation and, for each patch, at each level, its approximation error with respect to the reference surface is summarized into a compact view-dependent metric, **(b)** these are then used during hardware tessellation to select the most appropriate patch level according to the current viewing distance and direction.*

With hardware tessellation, highly detailed geometric models are decomposed into patches whose tessellation factor can be specified dynamically and independently at render time to control polygon resolution. Yet, to achieve maximum efficiency, an appropriate factor needs to be selected for each patch according to its content (geometry and appearance) and the current viewpoint distance and orientation. We proposed [4] a novel patch-based error metric that addresses this problem (Fig. 8). It summarizes both the geometrical error and the texture parametrization deviation of a simplified patch compared to the corresponding detailed surface. This metric is compact and can be efficiently evaluated on the GPU along any view direction. Furthermore, based

on this metric, we devise an easy-to-implement refitting optimization that further reduces the simplification error of any decimation algorithm, and propose a new placement strategy and cost function for edge-collapses to reach the best quality/performances trade-off.

7.3.2. MNPR: A Framework for Real-Time Expressive Non-Photorealistic Rendering of 3D Computer Graphics

Figure 9. A 3D scene rendered through MNPR in watercolor, oil and charcoal styles. Baba Yaga's hut model ©Inuciian.

We developed [12] a framework for expressive non-photorealistic rendering of 3D computer graphics: MNPR. Our work focuses on enabling stylization pipelines with a wide range of control, thereby covering the interaction spectrum with real-time feedback. In addition, we introduce control semantics that allow cross-stylistic art-direction, which is demonstrated through our implemented watercolor, oil and charcoal stylizations (Fig. 9). Our generalized control semantics and their style-specific mappings are designed to be extrapolated to other styles, by adhering to the same control scheme. We then share our implementation details by breaking down our framework and elaborating on its inner workings. Finally, we evaluate the usefulness of each level of control through a user study involving 20 experienced artists and engineers in the industry, who have collectively spent over 245 hours using our system. MNPR is implemented in Autodesk Maya and open-sourced through this publication, to facilitate adoption by artists and further development by the expressive research and development community.

7.4. Editing and Modeling

7.4.1. Interactive optimal transport solver

Optimal transport is a fundamental tool that appeared in various forms in numerous application domains. We developed a novel and extremely fast algorithm to compute continuous transport maps between 2D probability densities discretized on uniform grids. It follows the Monge-Ampère formulation, and it converges in a few cheap iterations thanks to the novel derivative-free non-linear solver we developed along this work. We achieve interactive performance in various applications such as blue noise sampling, feature sensitive remeshing, and caustic design (Fig. 10).

7.4.2. A Composite BRDF Model for Hazy Gloss

A new bidirectional reflectance distribution function (BRDF) model is introduced for the rendering of materials that exhibit hazy reflections, whereby the specular reflections appear to be flanked by a surrounding halo. The focus of this work is on artistic control and ease of implementation for real-time and off-line rendering. The material model is based on a pair of arbitrary BRDF models; however, instead of controlling their physical parameters, we expose perceptual parameters inspired by visual experiments. The

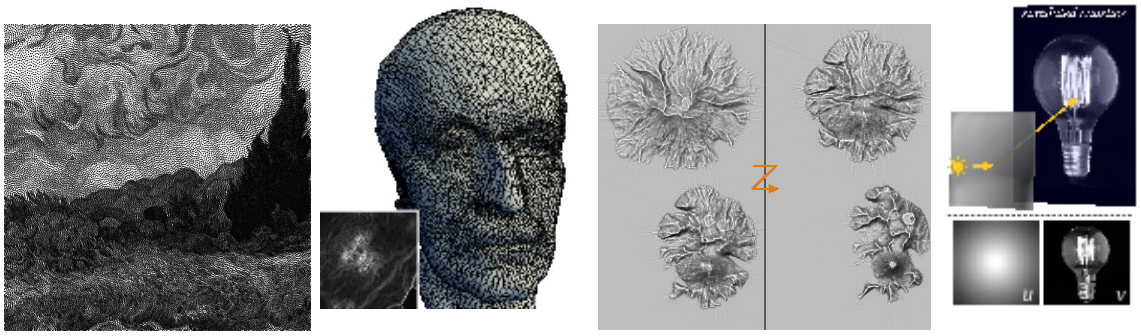


Figure 10. Our fast mass-transport solver enables many applications such as adaptive sampling, surface remeshing, heightfield morphing and caustic design with interactive performance.

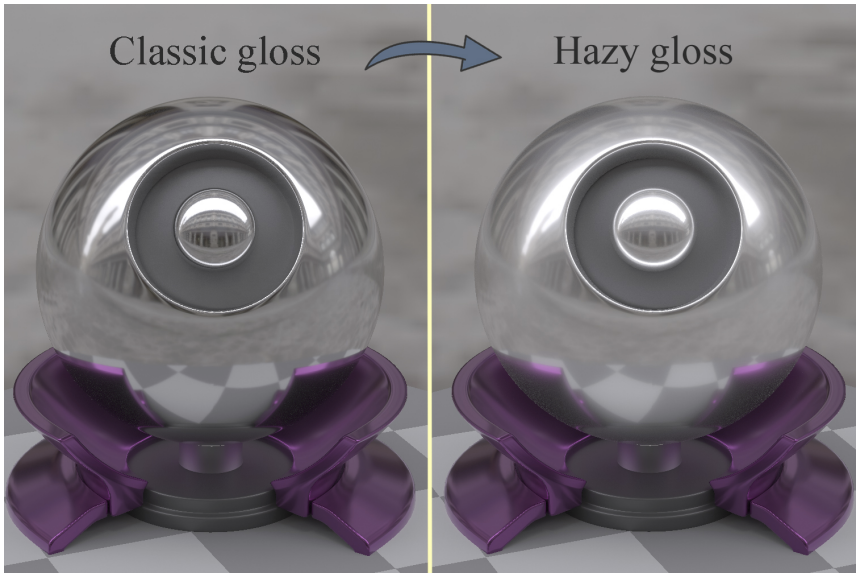


Figure 11. An object rendered with a classic glossy material (left), and with our hazy gloss material model (right), exhibiting specular reflections flanked by a halo.

main contribution then consists in a mapping from perceptual to physical parameters that ensures the resulting composite BRDF is valid in terms of reciprocity, positivity and energy conservation. The immediate benefit of this approach is to provide direct artistic control over both the intensity and extent of the haze effect (Fig. 11), which is not only necessary for editing purposes, but also essential to vary haziness spatially over an object surface.

MAVERICK Project-Team

7. New Results

7.1. Expressive Rendering

7.1.1. A workflow for designing stylized shading effects

Participants: Alexandre Bléron, Romain Vergne, Thomas Hurtut, Joëlle Thollot.

In this report [18], we describe a workflow for designing stylized shading effects on a 3D object, targeted at technical artists. Shading design, the process of making the illumination of an object in a 3D scene match an artist vision, is usually a time-consuming task because of the complex interactions between materials, geometry, and lighting environment. Physically based methods tend to provide an intuitive and coherent workflow for artists, but they are of limited use in the context of non-photorealistic shading styles. On the other hand, existing stylized shading techniques are either too specialized or require considerable hand-tuning of unintuitive parameters to give a satisfactory result. Our contribution is to separate the design process of individual shading effects in three independent stages: control of its global behavior on the object, addition of procedural details, and colorization. Inspired by the formulation of existing shading models, we expose different shading behaviors to the artist through parametrizations, which have a meaningful visual interpretation. Multiple shading effects can then be composited to obtain complex dynamic appearances. The proposed workflow is fully interactive, with real-time feedback, and allows the intuitive exploration of stylized shading effects, while keeping coherence under varying viewpoints and light configurations (see Fig. 2). Furthermore, our method makes use of the deferred shading technique, making it easily integrable in existing rendering pipelines.

7.1.2. MNPR: A framework for real-time expressive non-photorealistic rendering of 3D computer graphics

Participants: Santiago Montesdeoca, Hock Soon Seah, Amir Semmo, Pierre Bénard, Romain Vergne, Joëlle Thollot, Davide Benvenuti.

We propose a framework for expressive non-photorealistic rendering of 3D computer graphics: MNPR. Our work focuses on enabling stylization pipelines with a wide range of control, thereby covering the interaction spectrum with real-time feedback. In addition, we introduce control semantics that allow cross-stylistic art-direction, which is demonstrated through our implemented watercolor, oil and charcoal stylizations (see Fig. 3). Our generalized control semantics and their style-specific mappings are designed to be extrapolated to other styles, by adhering to the same control scheme. We then share our implementation details by breaking down our framework and elaborating on its inner workings. Finally, we evaluate the usefulness of each level of control through a user study involving 20 experienced artists and engineers in the industry, who have collectively spent over 245 hours using our system. MNPR is implemented in Autodesk Maya and open-sourced through this publication, to facilitate adoption by artists and further development by the expressive research and development community. This paper was presented at Expressive [13] and received the best paper award.

7.1.3. Motion-coherent stylization with screen-space image filters

Participants: Alexandre Bléron, Romain Vergne, Thomas Hurtut, Joëlle Thollot.

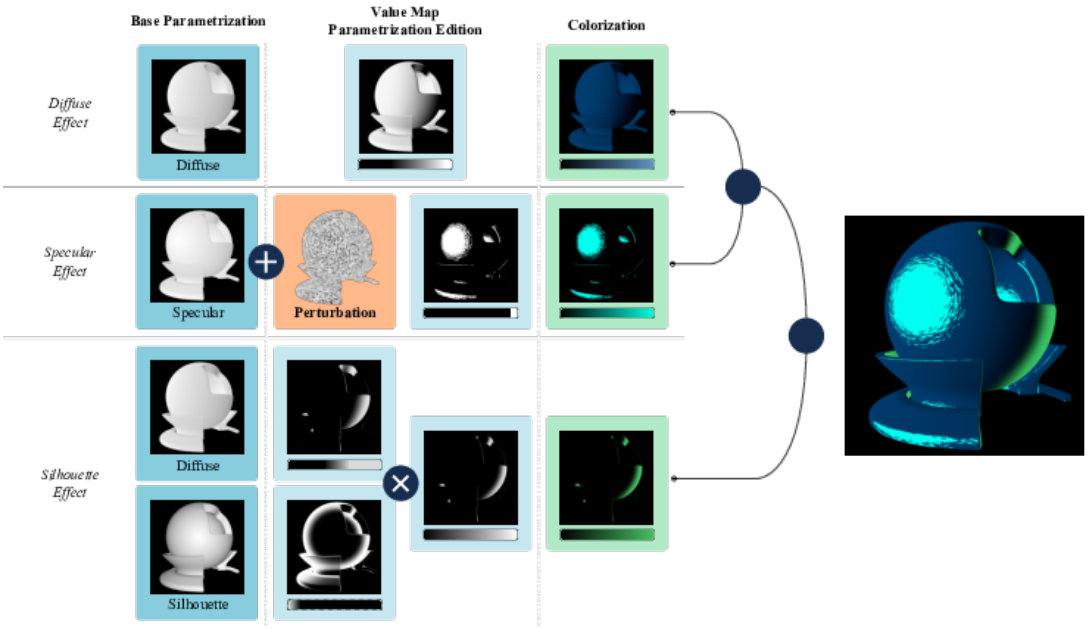


Figure 2. Illustration of our workflow showing an example with three appearance effects. A user can modify and combine base parametrizations to design the shading behavior (blue nodes) of an appearance effect, using value maps and combination operations. A color map (green nodes) is then applied on the designed behavior to colorize the effect. Output effects are then composited to obtain the final appearance. Perturbations (orange nodes) can be attached to every operation in order to add procedural details to an effect. The orientation of the perturbation can be controlled by the gradient of a shading behavior (as shown here), or by an external vector field, such as a tangent map.

Figure 3. A scene rendered through MNPR in different styles. Baba Yaga's hut model, © Inuciian.

One of the qualities sought in expressive rendering is the 2D impression of the resulting style, called flatness. In the context of 3D scenes, screen-space stylization techniques are good candidates for flatness as they operate in the 2D image plane, after the scene has been rendered into G-buffers. Various stylization filters can be applied in screen-space while making use of the geometrical information contained in G-buffers to ensure motion coherence. However, this means that filtering can only be done inside the rasterized surface of the object. This can be detrimental to some styles that require irregular silhouettes to be convincing. In this paper, we describe a post-processing pipeline that allows stylization filters to extend outside the rasterized footprint of the object by locally *inflating* the data contained in G-buffers (see Fig. 4). This pipeline is fully implemented on the GPU and can be evaluated at interactive rates. We show how common image filtering techniques, when integrated in our pipeline and in combination with G-buffer data, can be used to reproduce a wide range of *digitally-painted* appearances, such as directed brush strokes with irregular silhouettes, while keeping enough motion coherence. This paper was presented at Expressive [11].

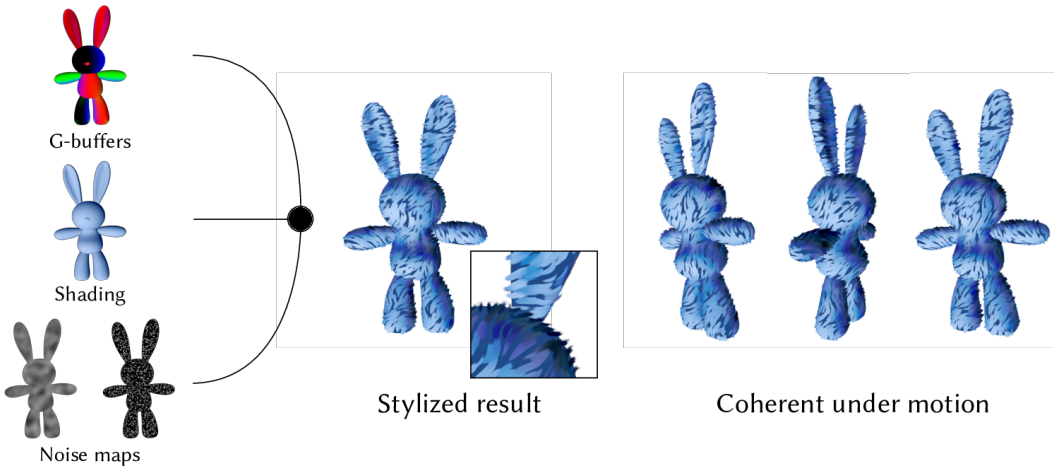


Figure 4. Using standard G-buffers and auxiliary buffers (noise, shading) as input, our pipeline can reproduce stylization effects that extend outside the original rasterized footprint of the object. Visual features produced by the filters stay coherent under motion or viewpoint changes.

7.2. Illumination simulation and materials

7.2.1. Rendering homogeneous participating media

Participants: Beibei Wang, Nicolas Holzschuch, Liangsheng Ge, Lu Wang.

Illumination effects in translucent materials are a combination of several physical phenomena: refraction at the surface, absorption and scattering inside the material. Because refraction can focus light deep inside the material, where it will be scattered, practical illumination simulation inside translucent materials is difficult. We have worked on a Point-Based Global Illumination method for light transport on homogeneous translucent materials with refractive boundaries. We start by placing light samples inside the translucent material and organizing them into a spatial hierarchy. At rendering, we gather light from these samples for each camera ray. We compute separately the sample contributions for single, double and multiple scattering, and add them. Multiple scattering effects are precomputed and stores in a table, accessed at runtime. An illustration of our approach is given in Fig 5 . We present two implementations of our algorithm: an offline version for high-quality rendering and an interactive GPU implementation. The offline version provides significant speed-ups and reduced memory footprints compared to state-of-the-art algorithms, with no visible impact on quality.

The GPU version yields interactive frame rates: 30 fps when moving the viewpoint, 25 fps when editing the light position or the material parameters. This work was published in IEEE Transactions on Visualization and Computer Graphics [9].

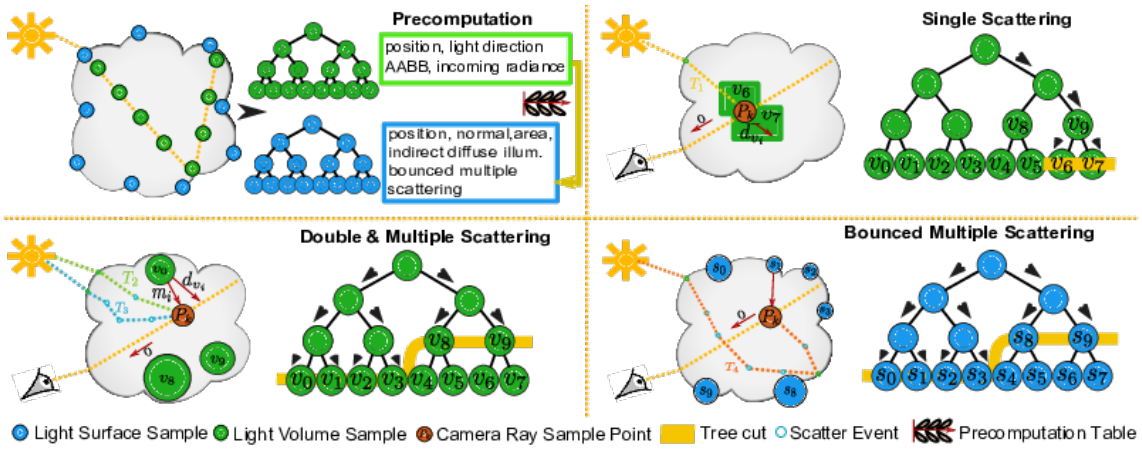


Figure 5. Our algorithm: we begin by computing incoming light at volume and surface samples. We then compute Single-, Double- and Multiple scattering effects for each camera ray using these volume and surface samples.

Storing the precomputed table for these multiple scattering effects is the largest memory cost for this algorithm. In a separate work, we used a neural network to encode these effects. We replaced the precomputed multiple scattering table with a trained neural network, with a cost of 6490 bytes (1623 floats). At runtime, the neural network is used to generate multiple scattering. The approach can be combined with many rendering algorithms, as illustrated in Fig. 6 . This work was published as a Siggraph Talk [12].

7.2.2. Fast global illumination with discrete stochastic microfacets using a filterable model

Participants: Beibei Wang, Lu Wang, Nicolas Holzschuch.

Many real-life materials have a sparkling appearance, whether by design or by nature. Examples include metallic paints, sparkling varnish but also snow. These sparkles correspond to small, isolated, shiny particles reflecting light in a specific direction, on the surface or embedded inside the material. The particles responsible for these sparkles are usually small and discontinuous. These characteristics make it difficult to integrate them efficiently in a standard rendering pipeline, especially for indirect illumination. Existing approaches use a 4-dimensional hierarchy, searching for light-reflecting particles simultaneously in space and direction. The approach is accurate, but still expensive. We have shown that this 4-dimensional search can be approximated using separate 2-dimensional steps. This approximation allows fast integration of glint contributions for large footprints, reducing the extra cost associated with glints be an order of magnitude, as illustrated in Fig. 7 . This work was published in Computer Graphics Forum and presented at the Pacific Graphics conference [10].

7.2.3. Handling fluorescence in a uni-directional spectral path tracer

Participants: Michal Mojzík, Alban Fichet, Alexander Wilkie

We present two separate improvements to the handling of fluorescence effects in modern uni-directional spectral rendering systems.

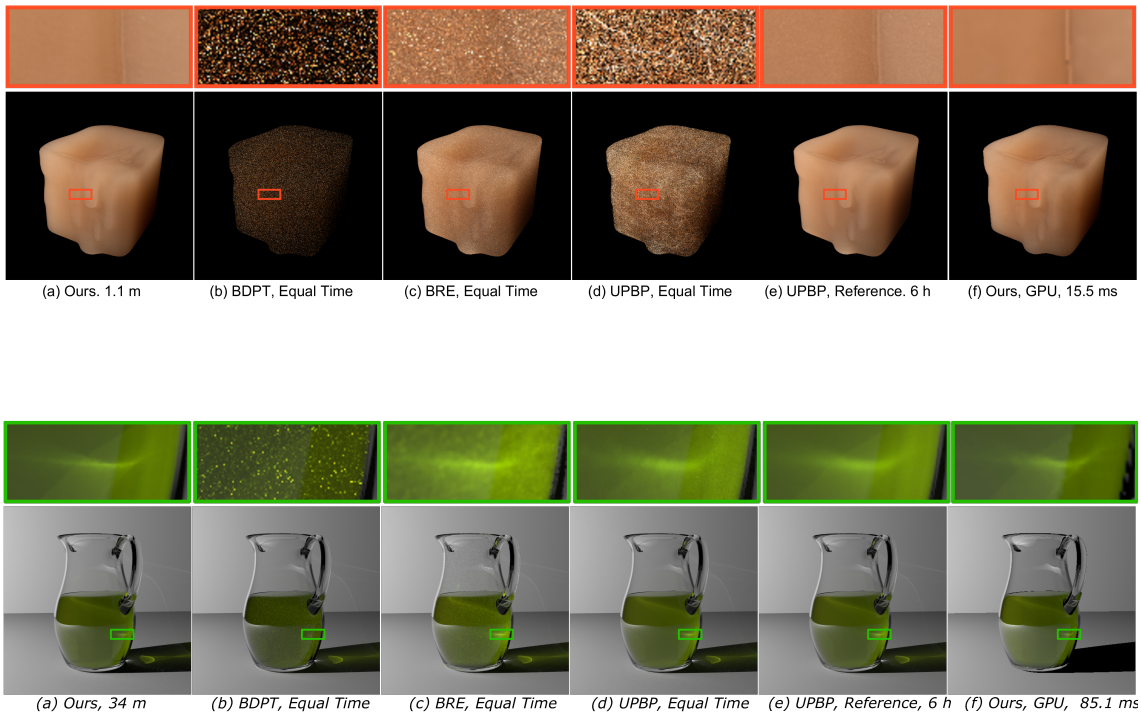


Figure 6. comparison between our algorithm, other algorithms with equal time or equal quality and reference images. Top row: wax. For this material, with a large albedo and a small mean free path, multiple scattering effects dominate. Bottom row: olive oil. For this material with low albedo and large mean-free-path, low-order scattering effects dominate.

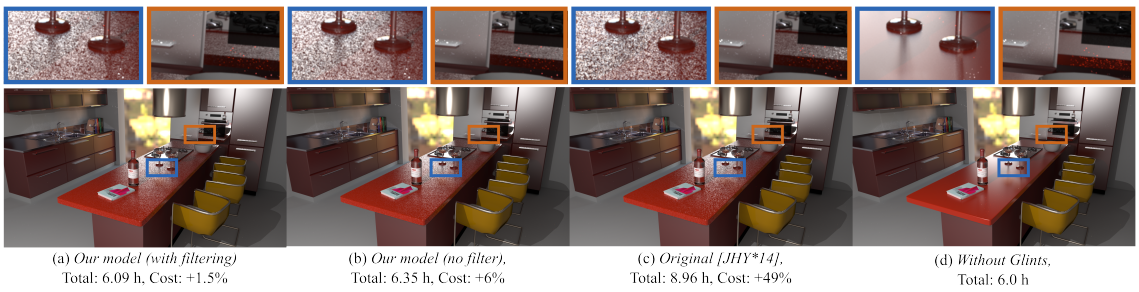
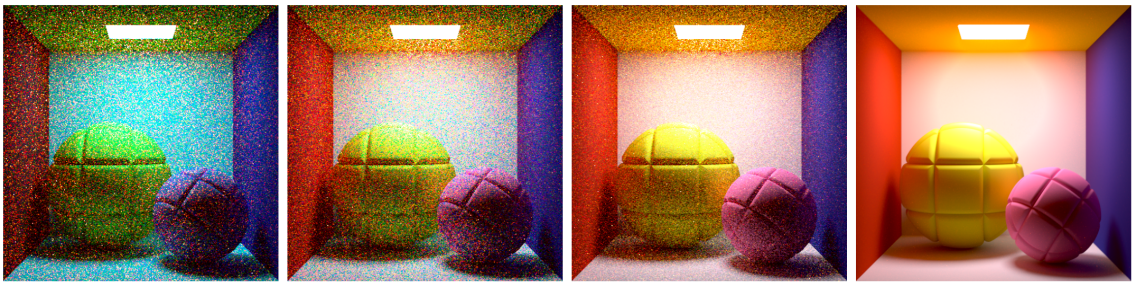


Figure 7. Our algorithm, compared to the original Discrete Stochastic Microfacets model (c). Converting the 4D search to a product of 2D searches (b) produces almost identical results. This is the basis for our filterable model (a), which allows fast global illumination with negligible cost..

The first is the formulation of a new distance tracking scheme for fluorescent volume materials which exhibit a pronounced wavelength asymmetry. Such volumetric materials are an important and not uncommon corner case of wavelength-shifting media behaviour, and have not been addressed so far in rendering literature. This new tracking scheme (figure 8 (b)) converges faster than a simple modification that can be added to the traditional exponential tracking (figure 8 (a)).

The second one is that we introduce an extension of Hero wavelength sampling which can handle fluorescence events, both on surfaces, and in volumes. Both improvements are useful by themselves, and can be used separately: when used together, they enable the robust inclusion of arbitrary fluorescence effects in modern uni-directional spectral MIS path tracers (figure 8 (c)). Our extension of Hero wavelength sampling is generally useful, while our proposed technique for distance tracking in strongly asymmetric media is admittedly not very efficient. However, it makes the most of a rather difficult situation, and at least allows the inclusion of such media in uni-directional path tracers, albeit at comparatively high cost. Which is still an improvement since up to now, their inclusion was not really possible at all, due to the inability of conventional tracking schemes to generate sampling points in such volume materials. This work was published in the journal Computer Graphics Forum [6].



(a) $\max(\sigma_t, \sigma_s)$, 16 spp, mono (b) Fluorescent-aware, 16 spp, mono (c) Fluo-aware, 16spp, HWSS (d) Fluo-aware, 65000 spp, HWSS

Figure 8. Comparison of proposed techniques to improve rendering of fluorescence.

7.2.4. A versatile parameterization for measured material manifolds

Participants: Cyril Soler, Kartic Subr, Derek Nowrouzezahrai.

A popular approach for computing photorealistic images of virtual objects requires applying reflectance profiles measured from real surfaces, introducing several challenges: the memory needed to faithfully capture realistic material reflectance is large, the choice of materials is limited to the set of measurements, and image synthesis using the measured data is costly. Typically, this data is either compressed by projecting it onto a subset of its linear principal components or by applying non-linear methods. The former requires many components to faithfully represent the input reflectance, whereas the latter necessitates costly extrapolation algorithms. We learn an underlying, low-dimensional non-linear reflectance manifold amenable to rapid exploration and rendering of real-world materials. We can express interpolated materials as linear combinations of the measured data, despite them lying on an inherently non-linear manifold. This allows us to efficiently interpolate and extrapolate measured BRDFs, and to render directly from the manifold representation. We exploit properties of Gaussian process latent variable models and use our representation for high-performance and offline rendering with interpolated real-world materials. This work has been published in the journal Computer Graphics Forum [7], and presented at Eurographics 2018.

7.3. Complex scenes

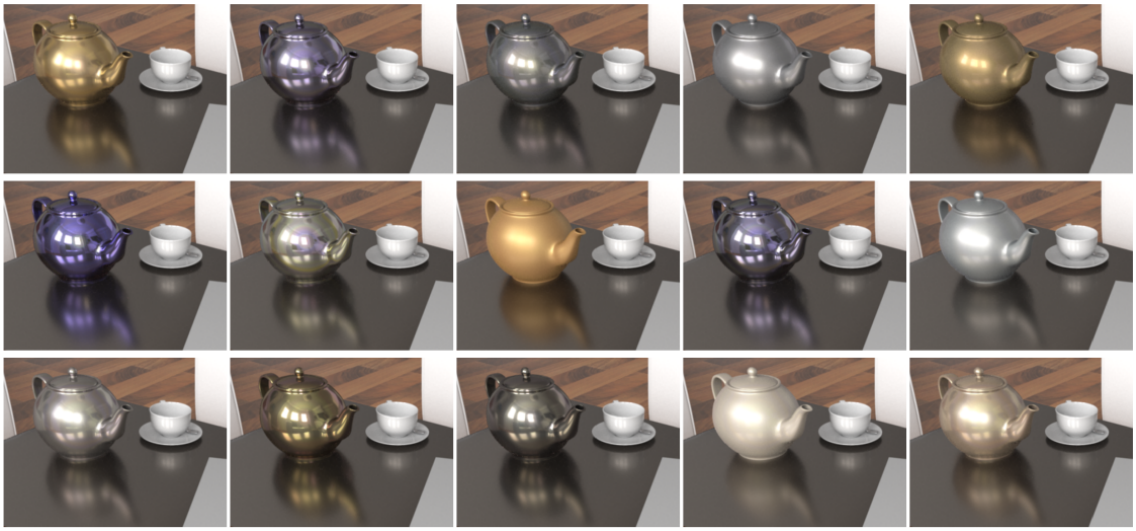


Figure 9. Four of the images above (Number 2, 4, 6 and 12 in reading order) are rendered with measured BRDFs from the MERL dataset, the remaining 11 being rendered with BRDFs randomly picked from our parameterization of the non-linear manifold containing MERL materials. We explore this manifold interactively to produce high-quality BRDFs which retain the physical properties and perceptual aspect of real materials.

7.3.1. A new microflake model with microscopic self-shadowing for accurate volume downsampling

Participants: Guillaume Loubet, Fabrice Neyret.

In this work, we addressed the problem of representing the effect of internal self-shadowing in elements about to be filtered out at a given LOD, in the scope of volume of voxels containing density and phase-function (represented by a microflakes).

Naïve linear methods for downsampling high resolution microflake volumes often produce inaccurate results, especially when input voxels are very opaque. Preserving correct appearance at all resolutions requires taking into account inter- and intravoxel self-shadowing effects (see Figure 10). We introduce a new microflake model whose parameters characterize self-shadowing effects at the microscopic scale. We provide an anisotropic self-shadowing function and a microflake distribution for which scattering coefficients and phase functions of our model have closed-form expressions. We use this model in a new downsampling approach in which scattering parameters are computed from local estimations of self-shadowing in the input volume. Unlike previous work, our method handles datasets with spatially varying scattering parameters, semi-transparent volumes and datasets with intricate silhouettes. We show that our method generates LoDs with correct transparency and consistent appearance through scales for a wide range of challenging datasets, allowing for huge memory savings and efficient distant rendering without loss of quality. This work received the Best Paper Award at Eurographics 2018 and was published in the journal Computer Graphics Forum [5].

7.4. Texture synthesis

7.4.1. Gabor noise revisited

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

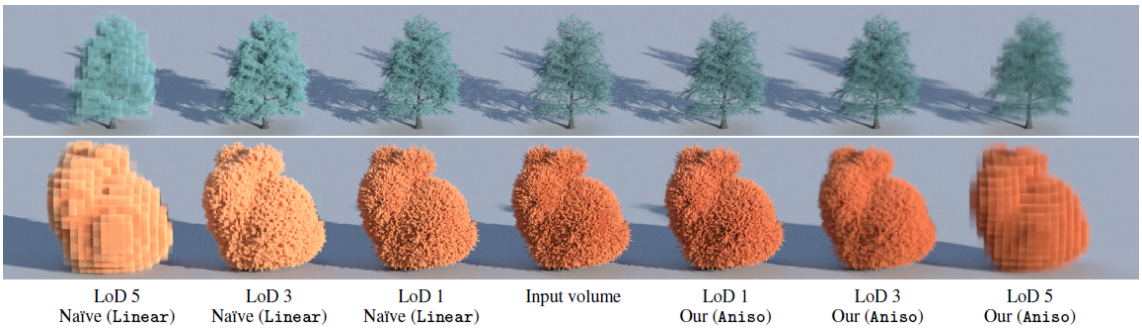
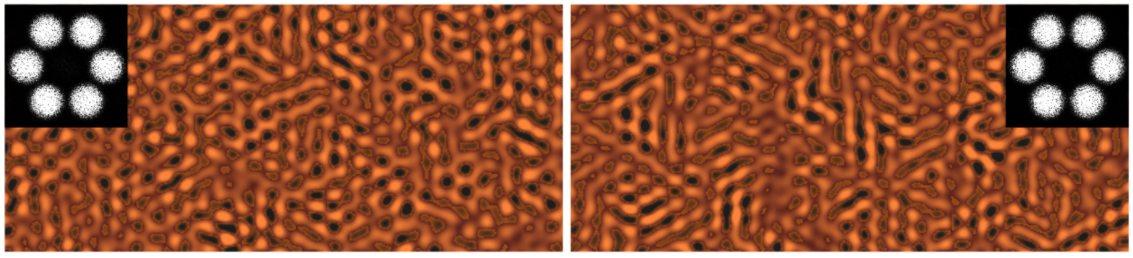


Figure 10. Comparison between naïve downsampling of microflake volumes and our method ("Aniso"). Naïve dowsampling of volumes with dense voxels often lead to inaccurate results due to the loss of inter- and intra-voxel self-shadowing effects. Our method is based on a new participating medium model and on local estimations of self-shadowing. It generates LoDs with correct transparency and consistent appearance through scales. Rendered with volume path tracing in Mitsuba (http://www.mitsuba-renderer.org/): the trunk of the cedar is a mesh.

Gabor Noise is a powerful procedural texture synthesis technique, but 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 $17-24\times$ speed-up with same or better quality (see Fig. 11).

This paper was presented at AFIG [17] and received the best paper award. An article has been submitted to Eurographics-short 2019.



(a) Seminal Gabor, $N = 45$ (b) Bernoulli+strat.+sin, $N = 3$

Figure 11. Real case with complex power spectrum (3 kernels, cf. inset) and non-linear post-treatment. Our optimized set of ingredients achieves the same visual quality in $1/17^{th}$ of the time required by the seminal method.

7.4.2. High-performance by-example noise using a histogram-preserving blending operator

Participants: Eric Heitz, Fabrice Neyret.

We propose a new by-example noise algorithm that takes as input a small example of a stochastic texture and synthesizes an infinite output with the same appearance. It works on any kind of random-phase inputs as well as on many non-random-phase inputs that are stochastic and non-periodic, typically natural textures such as moss, granite, sand, bark, etc. Our algorithm achieves high-quality results comparable to state-of-the-art procedural-noise techniques but is more than 20 times faster. Our approach is conceptually simple: we partition the output texture space on a triangle grid and associate each vertex with a random patch from the input such that the evaluation inside a triangle is done by blending 3 patches. The key to this approach is the blending operation that usually produces visual artifacts such as ghosting, softened discontinuities and reduced contrast, or introduces new colors not present in the input. We analyze these problems by showing how linear blending impacts the histogram and show that a blending operator that preserves the histogram prevents these problems. The main requirement for a rendering application is to implement such an operator in a fragment shader without further post-processing, i.e. we need a histogram-preserving blending operator that operates only at the pixel level. Our insight for the design of this operator is that, with Gaussian inputs, histogram-preserving blending boils down to mean and variance preservation, which is simple to obtain analytically. We extend this idea to non-Gaussian inputs by "Gaussianizing" them with a histogram transformation and "de-Gaussianizing" them with the inverse transformation after the blending operation. We show how to precompute and store these histogram transformations such that our algorithm can be implemented in a fragment shader, as illustrated in Fig. 12 . This work received the Best Paper Award at High Performance Graphics 2018 [4].

7.5. Visualization

7.5.1. A "What if" approach for eco-feedback

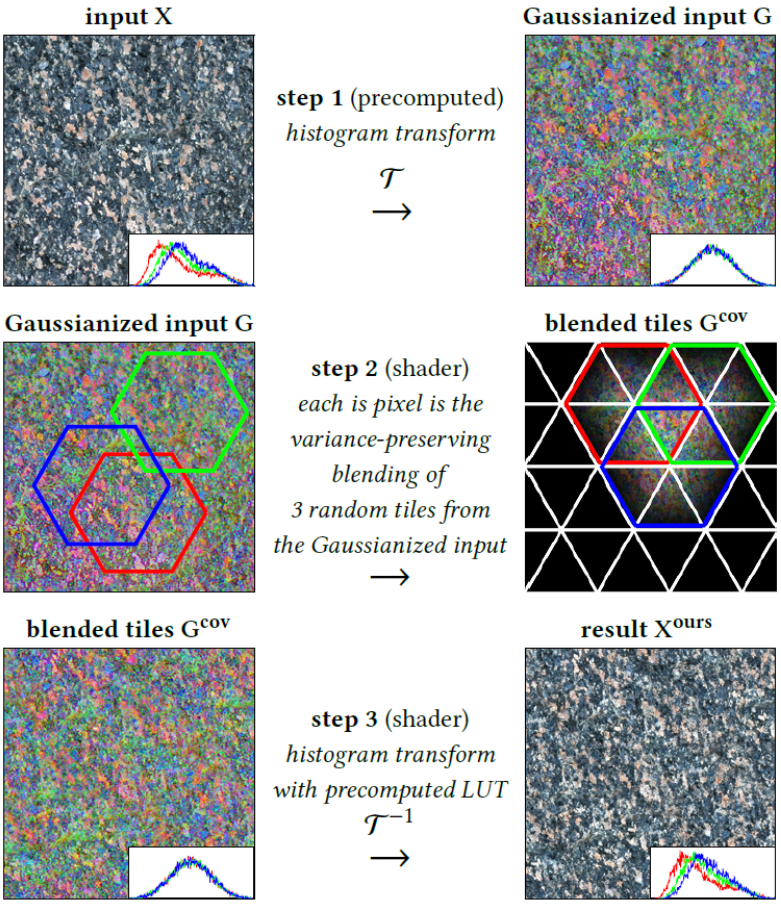
Participants: Jérémy Wambecke, Georges-Pierre Bonneau, Romain Vergne, Renaud Blanch.

Many households share the objective of reducing electricity consumption for either economic or ecological motivations. Eco-feedback technologies support this objective by providing users with a visualization of their consumption. However as pointed out by several studies, users encounter difficulties in finding concrete actions to reduce their consumption. To overcome this limitation, we introduce and evaluate Activelec, a system based on the visualization and interaction with user's behavior rather than raw consumption data. The user's behavior is modeled as the set of actions modifying the state of appliances over time. A key novelty of our solution is its focus on the What if approach applied to eco-feedback. Users can analyze and experiment scenarios by selecting and modifying their usage of electrical appliances over time and visualize the impact on the consumption, as illustrated in Fig. 13 . In [16] we conducted two laboratory user studies that evaluate the usability of Activelec and the relevance of the What if approach for electricity consumption. Our results show that users understand the interaction paradigm and can easily find relevant modifications in their usage of appliances. Moreover participants judge these changes of behavior would require little effort to be adopted. In [15] we conducted an in-situ evaluation of Activelec, confirming these results in a real setting.

7.5.2. Morphorider: a new way for Structural Monitoring via Shape Acquisition

Participants: Tibor Stanko, Laurent Jouanet, Nathalie Saguin-Sprynski, Georges-Pierre Bonneau, Stefanie Hahmann.

In collaboration with CEA-Leti we introduce a new kind of monitoring device, illustrated in Fig. 14 , allowing the shape acquisition of a structure via a single mobile node of inertial sensors and an odometer. Previous approaches used devices placed along a network with fixed connectivity between the sensor nodes (lines, grid). When placed onto a shape, this sensor network provides local surface orientations along a curve network on the shape, but its absolute position in the world space is unknown. The new mobile device provides a novel way of structures monitoring: the shape can be scanned regularly, and following the shape or some specific parameters along time may afford the detection of early signs of failure. Here, we present a complete framework for 3D shape reconstruction. To compute the shape, our main insight is to formulate the reconstruction as a set of optimization problems. Using discrete representations, these optimization problems are resolved efficiently and at interactive time rates. We present two main contributions. First, we introduce a novel method for creating well-connected networks with cell-complex topology using only orientation and distance measurements and



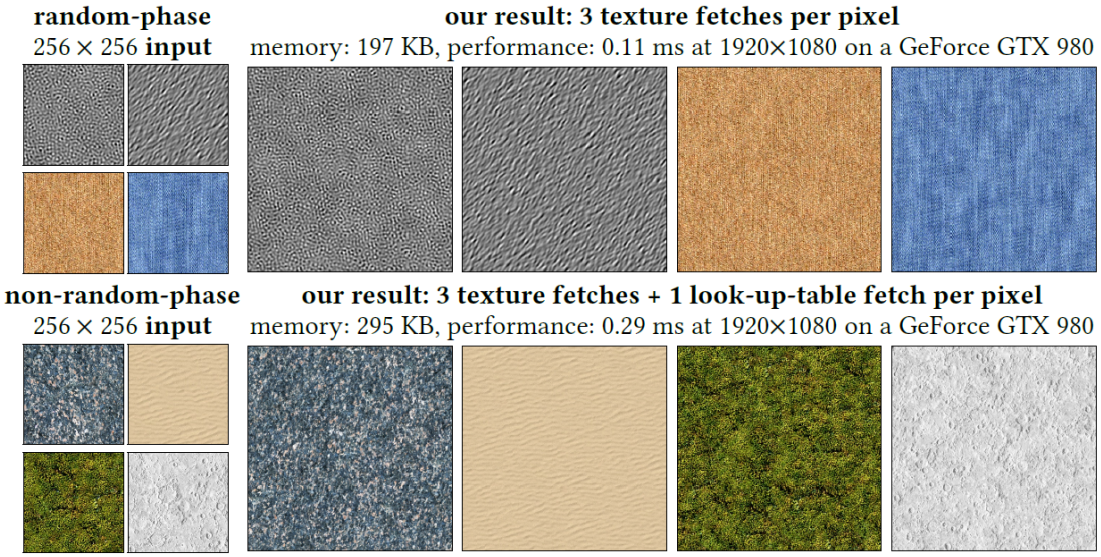


Figure 12. Top: method overview. Bottom: results and performances.

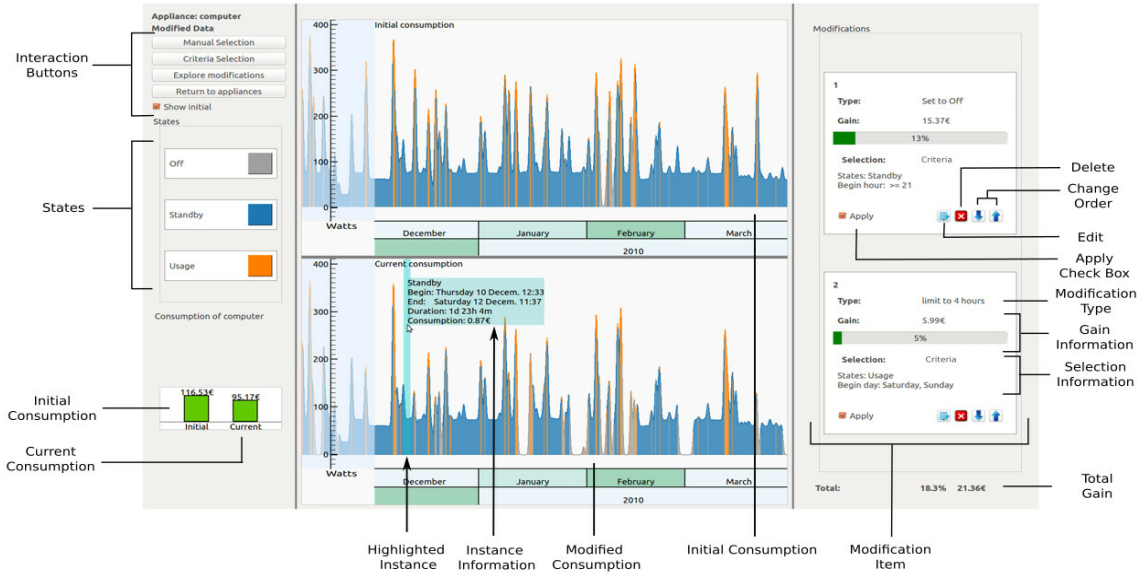


Figure 13. Interface of our system. A computer has been chosen by the user, whose states are Usage (orange) and Standby (blue). At the right, we can see that the user has applied two modifications, the first one to remove instances of Standby after 9 P.M, and the second one to limit the instances of On to 4 hours during the weekend. When the user is selecting instances, this panel displays information about the selection.

a set of user-defined constraints. Second, we address the problem of surfacing a closed 3D curve network with given surface normals. The normal input increases shape fidelity and allows to achieve globally smooth and visually pleasing shapes. The proposed framework was tested on experimental data sets acquired using our device. A quantitative evaluation was performed by computing the error of reconstruction for our own designed surfaces, thus with known ground truth. Even for complex shapes, the mean error remains around 1%. This work was published at the 9th European Workshop on Structural Health Monitoring [14].

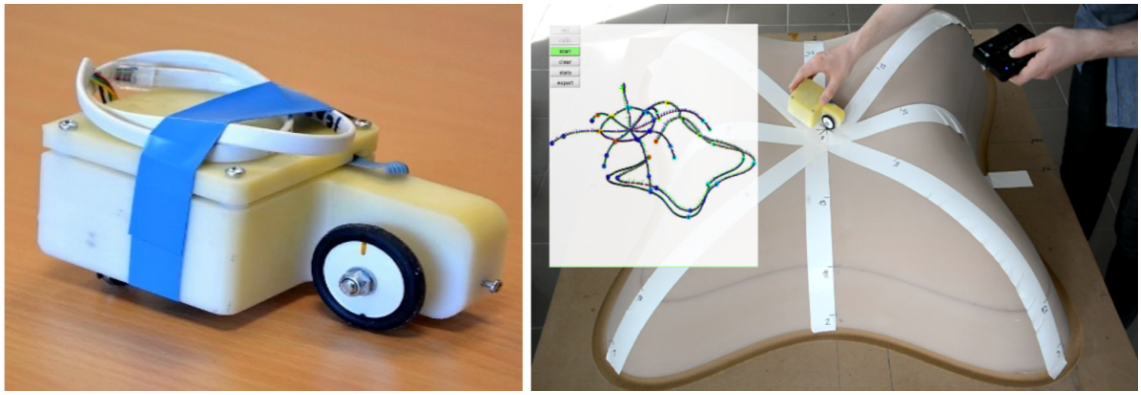


Figure 14. Morphorider: Structural Monitoring via Shape Acquisition (right) with a mobile device (left) equipped with an inertial node of sensors and an odometer.

MFX Team

6. New Results

6.1. Carving Large Cavities in Shapes for Fast Fused Deposition Modeling

Participants: Samuel Hornus, Sylvain Lefebvre.

FDM Fused Deposition Modeling: fabricating things by depositing fused material into layers.

In 2016, we developed a technique for modeling a tight shield that protects the part being manufactured during 3D-printing with multi-material. In particular, the shield catches oozing material before it reaches the part [19]. The technique was implemented on a voxel representation of the shape. We also demonstrated its use for the modeling of a large *self-supported* cavity inside the shape.

In this more recent work, we have extended the technique to iteratively carve large cavities in the shape in order to hollow a shape as much as possible while maintaining its ability to be fabricated without internal support. (see Figure 2) We developed a polygonal implementation of the technique that provides much higher quality results. The work was published at the 2018 Eurographics conference as a short paper [14]. An implementation is now available to the general public in our software IceSL.

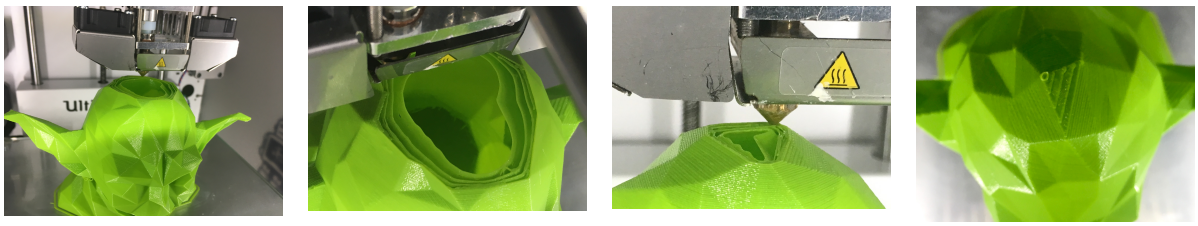


Figure 2. From Section 6.1 . Timelapse of the printing of a Yoda model. (Middle-left.) Note how the print is mostly empty and the nested cavity walls. Middle-right. Approaching the top of the head. Right. Closing the top of the head.

6.2. A Metamaterial for Fused Filament Fabrication

Participants: Jonàs Martínez Bayona, Samuel Hornus, Sylvain Lefebvre.

A critical advantage of additive manufacturing is its ability to fabricate complex small-scale structures. These microstructures can be understood as a *metamaterial*: they exist at a much smaller scale than the volume they fill, and are collectively responsible for an average elastic behavior different from that of the base printing material. For instance, this can make the fabricated object lighter and/or flexible along specific directions. In addition, the average behavior can be graded spatially by progressively modifying the microstructure geometry (see Figure 3).

The definition of a microstructure is a careful trade-off between the geometric requirements of manufacturing and the properties one seeks to obtain within a shape: in our case a wide range of elastic behaviors. Most existing microstructures are designed for stereolithography (SLA) and laser sintering (SLS) processes. The requirements are however different than those of continuous deposition systems such as fused filament fabrication, for which there was a lack of microstructures enabling graded elastic behaviors.

Figure 3. A 3D printed shoe sole. Left: Control fields used on the model, density (top), orthotropy strength (middle) and angle (bottom). Right: Printed shoe, top, side and bending. The shoe is printed without any skin to reveal the foam structure.

We introduced a novel type of metamaterial that *strictly enforces* all the requirements of Fused Filament Fabrication (FFF): continuity, self-support and overhang angles. This metamaterial offers a range of orthotropic elastic responses that can be graded spatially. This allows us to fabricate parts usually reserved to the most advanced technologies on widely available inexpensive printers that also benefit from a continuously expanding range of materials.

This work was presented at the SIGGRAPH conference and published in ACM Transactions on Graphics [12], and is integrated in the publicly available IceSL software. This was a joint work with Haichuan Song, then a post-doctoral researcher in ALICE.

6.3. Topology Optimization of Parametrized Stochastic Microstructures

Participants: Jonàs Martínez Bayona, Sylvain Lefebvre.

Different works have explored the topology optimization of parametrized periodic microstructures by the homogenization method. A promising venue of work lies in Additive Manufacturing technologies, that allow us to physically realize the intricate designs obtained with topology optimization. In order to fabricate the results, the parametrized microstructures must be projected at some finite scale taking into account the minimum printable size. However, for periodic microstructures it remains difficult to project and continuously grade the material properties since the boundary and transition between tiles has to be carefully handled.

We have an ongoing project in collaboration with Perle Geoffroy-Donders and Grégoire Allaire at École Polytechnique, to investigate the applicability of stochastic microstructures for topology optimization. This year we studied two different stochastic microstructures (isotropic and orthotropic) solely parametrized by an anisotropic metric and a Poisson point process. Both stochastic microstructures are amenable to efficient and scalable computation of their geometry. Unlike previous methods dealing with the projection of orthotropic microstructures the presented microstructures are able to easily follow a field of orthotropy orientation (see Figure 4).

6.4. Hash-based CSG Evaluation on GPU

Participants: Cédric Zanni, Sylvain Lefebvre.

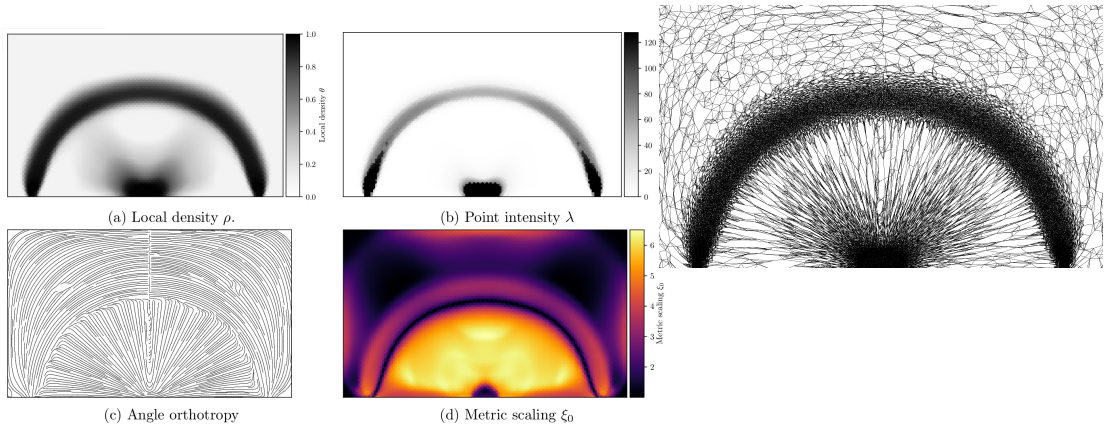


Figure 4. Optimization of a bridge problem with an orthotropic material, and our parametric stochastic microstructure. Left: Optimized parameters of the microstructure (density, angle of orthotropy, and degree of orthotropy). Right: Projection of the stochastic microstructure at a finite scale.

We have developed a new evaluation scheme for Constructive Solid Geometry (CSG) modeling that is well adapted to modern GPUs. The approach falls into the category of screen space techniques and can handle a large range of geometric representations. The proposed method relies on the idea of hashing in order to reduce the memory footprint for the processing of a given ray in the scene (*e.g.*, for discovering which part of the space is within or outside the object) while allowing the evaluation of the CSG in amortized constant time. This memory reduction in turn allows the space to be subdivided in order to apply progressively the rendering algorithm, ensuring that required data fit in the graphic memory. This improvement over previous approaches allows us to handle objects of higher complexity during both modeling and slicing for additive manufacturing.

The work was presented at the 2018 Symposium on Interactive 3D Graphics and Games conference and published in the ACM journal Computer Graphics and Interactive Techniques [15]. It was then integrated in the current version of our software IceSL.

6.5. Tile-based Pattern Design with Topology Control

Participant: Sylvain Lefebvre.

This project is a collaboration with Li-Yi Wei (HKU/Adobe) in the context of the PrePrint3D associated team. We consider the problem of producing tilings with boundary constraints, while enforcing global topology constraints. Tilings are composed by assembling a number of square tiles. Only tiles with compatible boundaries may be placed next to each others. In our context the tiles contain solid shapes connecting some of the borders together (corners, bars, crosses, etc.). Our algorithm is able to produce tilings that enforce border constraints as well as global topology constraints – in particular obtaining a connected network. This has applications in digital manufacturing, for instance to design decorative panels, but also in Computer Graphics, to synthesize large environments guaranteed to be navigable. These results were published in the ACM journal Computer Graphics and Interactive Techniques [10] and presented at the 2018 Symposium on Interactive 3D Graphics and Games conference.

We continue exploring tiling related problems, for instance to encode information within synthesized tilings [17].

6.6. Curved Deposition

Participants: Sylvain Lefebvre, Jimmy Etienne.

This project continues in collaboration with the ALICE team.

We are pursuing a line of research around curved deposition. The objective is to go beyond the flat-layers currently used. Indeed, some processes would allow for deposition along curved paths, however this capability is rarely used: proofs of concept exist, but no general algorithm can generate curved paths given an input geometry.

There are several key potential advantages to curved deposition: reducing the constraints in terms of geometries that can be manufactured, achieving better mechanical properties (*e.g.*, by aligning deposition with respect to a computed stress field), achieving better surface quality.

In this context, we achieved new results to reduce support material, in a joint project with Charlie C.L. Wang (TU Delft) [11]. The 3D printer is a 5-DOF robotic arm equipped with a standard FDM extruder. The algorithm we developed is based on a heuristic growth process within a discretized version of the model (voxels). The growth process attempts to place additional material where it is already supported from below, while avoiding cases where some unfinished parts of the model would become inaccessible due to collisions.

This led us to the first general algorithm for multi-axis 3D printing. It produces tool-paths that allow the robotic arm to fabricate most parts without any support, while avoiding collisions. Many challenges remain, both related to geometry and robotics, and we are pursuing this collaboration, jointly with Nicolas Ray (ALICE-Inria).

6.7. Colored 3D Printing

Participants: Sylvain Lefebvre, Jonàs Martínez Bayona, Noémie Vennin, Pierre Bedell.

In 2018 we kept developing our project regarding colored FDM printing. We have a paper accepted with minor revisions in ACM Transactions on Graphics. This was a joint work with Haichuan Song, then a post-doctoral researcher in ALICE. We worked on revising and refining our initial results throughout the year.

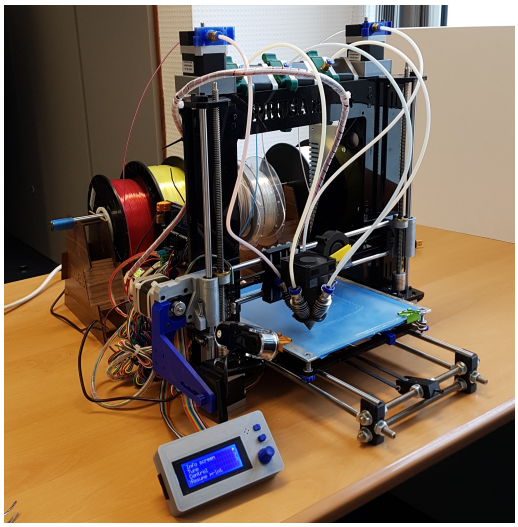


Figure 5. 3D printer Diamonds 5 filaments.

We proposed a novel algorithm for the problem of determining micro-layer mixtures to reproduce a subspace of material mixing ratios. We express the problem as fitting a simplex of minimal volume enclosing a set of points. The vertices of the simplex correspond to micro-layer mixtures, while the point set captures the desired mixtures within the model. This algorithm replaces the previous non-linear, gradient based optimizer. It achieves better results at a fraction of the previous computation time.

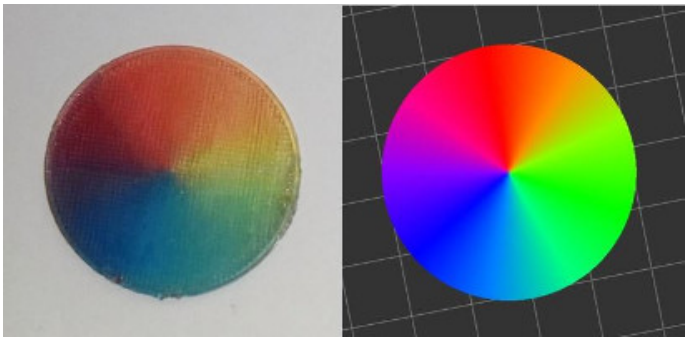


Figure 6. Disc showing all the gradation of color with 3 filaments (red, yellow, blue). Left: 3D printed disc. Right: Numerical view on the software IceSL.

We also developed, through the internship of Pierre Bedell, a 3D printer able to mix up to five filaments. We ran extensive testing and implemented additional improvements regarding flow control during deposition. Pierre Bedell joined the team as a research engineer and will keep participating in this project.

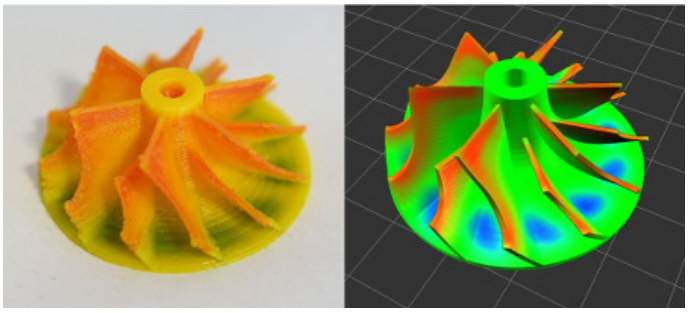


Figure 7. Turbine with a colored simulation of friction. Left: 3D printed turbine. Right: Numerical view on the software IceSL.

Noémie Vennin joined the team on a funding from *Université de Lorraine* to explore material aspects of the project, in a close collaboration with Sandrine Hoppe (LRGP). This part of the project receives support from the CPER Cyber-entreprise, thanks to which we acquired equipment to develop our own filament formulations, mixing pigments and additives to control color and transparency. Pierre Bedell and Noémie Vennin are developing a calibration process to determine the achievable color space given specific filaments, but also to tackle the inverse problem of designing filaments spanning a desired color space.

6.8. IceSL

Participants: Sylvain Lefebvre, Salim Perchy, Cédric Zanni, Samuel Hornus, Jonàs Martínez Bayona, Jimmy Etienne, Noémie Vennin, Pierre Bedell.

IceSL is the software developed within the team that serves as a research platform, a showcase of our research results, a test bed for comparisons and a vector of collaborations with both academic and industry partners. The software is freely available at https://icesl.loria.fr, both as a desktop and an online version.

In 2018, IceSL has been featured in news, exhibitions and fairs as a well-established tool for 3D printing. Additionally, since its inception, IceSL's community has grown significantly together with the number of new features included in it for slicing and modeling.

In February 2018, we organized the first event to introduce basic and advanced features which differentiate IceSL from other 3D printing tools. The event, targeted towards enthusiasts, allowed its participants to follow tutorials, interact with its developers and suggest additions and new directions for the software.

IceSL was also presented in May 2018 at the Strasbourg's Mini MakerFaire to a general audience that included high school students. The audience was introduced to IceSL's new features first hand and their applications to 3D printing. In addition to this, IceSL was shown to designers in November 2018 at Affinité Design (http://www.affinitedesign.com/) with part of the developing team demonstrating and answering questions on the use of IceSL as a modeling tool.

In October 2018, both the desktop and the online versions of IceSL were featured in a list of the 24 best 3D printing software tools. [0]

Regarding new features and additions to the software in 2018, IceSL has added several innovative methods for modeling and slicing. With respect to modeling, these include the ability to interactively paint values in a script (field tweaks), the option to export the shape generated with CSG to a mesh via dual contouring, texture synthesis on 3D objects, better font geometry creation as well as numerous improvements on its user interface and compatibility with hardware.

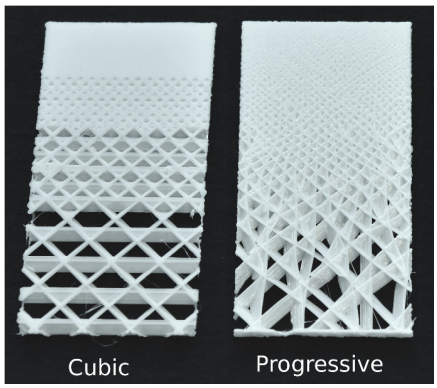


Figure 8. New infill patterns introduced in IceSL, the left one (cubic) has now been adopted by most other slicing software.

On the slicing front, IceSL has introduced new material infilling methods such as *polyfoam* [12], progressive and cubic structures (Figure 8) as well as putting a system in place allowing the user to specify an infill pattern through program image assets or shaders.

[0] https://all3dp.com/fr/1/meilleur-logiciel-imprimante-3d-gratuit-en-ligne/

IceSL also added a new method to compute supports called “wings,” a new framework for mixing colors into a 3D print [20] (presented in several exhibitions), curved printing covers, a faster slicing algorithm (in case of tessellated geometry), and a new geometry renderer [15].

The social community of IceSL is also growing accordingly. Its twitter account has around 200 followers and there are 150 users frequently interacting in its google forum. Downloads have increased around 30% after the first event in February 2018 to make a cumulative of 30k downloads since its initial release. [0] Youtube videos done by third persons on the usage of IceSL are also common (around a dozen in three different languages). And finally, in October 2018 IceSL launched its new website with a more professional look and additional resources (documentation, tutorials, videos, online version and new features).

6.9. Chill

Participants: Jimmy Etienne, Sylvain Lefebvre.

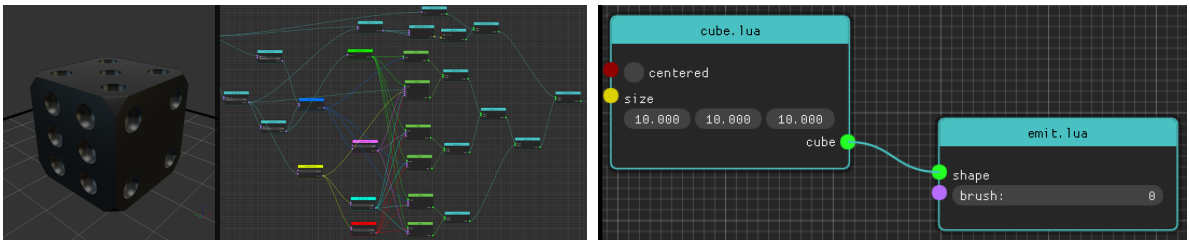


Figure 9. The Chill node-based GUI.

Chill is an open source GUI for IceSL, illustrated in Figure 9 . It features a node-based interface that hides the scripting language used to model shapes in IceSL. This enables a larger group of people to use IceSL in their projects, without having to type any code. The user of Chill creates 3D shapes by connecting various nodes arranged in a directed graph. The shape visualization is updated instantly as the graph is modified.

The source code is publicly available at https://github.com/shapeforge/Chill. We are planning to communicate broadly about the software in the first months of 2019.

[0] See https://gforge.inria.fr/top/toplist.php?type=downloads. Due to the removal of a file, the download counter on gforge.inria.fr is off by 6000 downloads.

MIMETIC Project-Team

7. New Results

7.1. Outline

In 2018, MimeTIC has maintained his activity in motion analysis, modelling and simulation. In motion analysis, we focused our efforts on two major points: 1) being able to simplify the calibration and simulation of customized musculoskeletal models of the subjects, 2) explore how visual perception act on collision avoidance in pedestrian locomotion with an extension to group behavior.

From a long time, MimeTIC has been promoting the idea of using Virtual Reality to train human performance. On the one hand, it leads to an efficient trade-off between high control and naturalness of the situation. On the other hand, it raises several fundamental questions about the automatic evaluation of the performance of the user, and the transfer of the skills trained in VR to real practice. In 2017, we explored these two questions by 1) developing new automatic methods for users' performance recognition and evaluation, especially online action detection, and 2) biofidelity of mass manipulation in VR using haptic interfaces.

In virtual cinematography, we applied the analysis/synthesis approach to extract and simulate film styles and narration. We also extended our previously defined Toric Space for camera placement to drone toric space to control a group of drones filming the action of an actor to ensure covering cinematographic distinct viewpoints. We also developed original VR-based staging and cinematography methods to make these processes be more interactive and immersive.

7.2. Motion analysis

7.2.1. Biomechanics for motion analysis-synthesis

Participants: Charles Pontonnier [contact], Georges Dumont, Franck Multon, Antoine Muller, Pierre Puchaud.

Based on a former PhD thesis (of Antoine Muller), we aim at democratizing the use of musculoskeletal analysis for a wide range of users. We proposed contributions enabling better performances of such analyses and preserving accuracy, as well as contributions enabling an easy subject-specific model calibration [47], [48]. In order to control the whole analysis process, we propose a global approach of all the analysis steps: kinematics, dynamics and muscle forces estimation. For all of these steps, quick analysis methods have been proposed. Particularly, a quick muscle force sharing problem resolution method [26] has been proposed, based on interpolated data and improvements have been proposed [25]. Moreover, the Music Toolbox is now proposed as an opensource software.

The determination of maximal torque envelopes method that we defined for the elbow torque analysis have been used for the shoulder [44]. It is important, in order to calibrate muscular models, to be able to identify force parameters in a musculoskeletal.

7.2.2. Interactions between walkers

Participants: Anne-Hélène Olivier [contact], Armel Crétual, Richard Kulpa, Sean Lynch.

Interaction between people, and especially local interaction between walkers, is a main research topic of MimeTIC. We propose experimental approaches using both real and virtual environments to study both perception and action aspects of the interaction. In the context of Sean Lynch's PhD, which was defended in October 2018 [12], we aimed at manipulating the nature of the visual information available to the participants to understand which information about the other walker are important to avoid a collision. We presented at IEEE VR 2018, our work on the influence of global and local information appearances [46] as well as on the influence of mutual gaze in the interaction [39].

In the context of transportation research, we developed a new collaboration with Ifsttar (LEPSIS, LESCOT) involving questions about interaction between pedestrians on a narrow sidewalk [50], [42].

We also provide lot of efforts to investigate, in collaboration with Julien Pettré from Inria Rainbow team, the process involved in the selection of interactions within our neighbourhood. Considering the complex case of multiple interactions, we first performed experiments in real conditions where a participant walked across a room whilst either one (i.e., pairwise) or two (i.e., group) participants crossed the room perpendicularly. By comparing these pairwise and group interactions, we assessed whether a participant avoids two upcoming collisions simultaneously, or as sequential pairwise interactions. Results showed that pedestrians are able to interact with two other walkers simultaneously, rather than treating each interaction in sequence. These results are currently in press in Frontiers in Psychology [22]. Second, we performed experiments involving 40 people to understand how collective behaviour emerges [31]. Third, in virtual conditions, we also coupled the analysis of gaze behaviour and the trajectory and showed that human gaze, during navigation, is attracted by other walkers presenting the higher risk of future collision [21], [32].

Finally, we continue working on the applications of studying human behaviour for application in human-moving robot interactions. The development of Robotics accelerated these recent years, it is clear that robots and humans will share the same environment in a near future. In this context, understanding local interactions between humans and robots during locomotion tasks is important to steer robots among humans in a safe manner. In collaboration with Philippe Souères and Christian Vassallo (LAAS, Toulouse), our work analyzed the behavior of human walkers crossing the trajectory of a mobile robot that was programmed to reproduce this human avoidance strategy. In contrast with a previous study, which showed that humans mostly prefer to give the way to a non-reactive robot, we observed similar behaviors between human-human avoidance and human-robot avoidance when the robot replicates the human interaction rules. This result highlight the importance of controlling robots in a human-like way in order to ease their cohabitation with humans [28]. In collaboration with Jose Grimaldo da Silva and Thierry Fraichard (Inria Grenoble), we designed a shared-effort model during interaction between a moving robot and a human relying on walker-walker collision avoidance data [34].

7.2.3. Biomechanical analysis of tennis serve

Participants: Richard Kulpa [contact], Benoit Bideau, Pierre Touzard.

In the context of the exclusive collaboration with the FFT (French Tennis Federation), we made new experiments on top-level young French players (between 12 up to 18 years old) to quantify the relation between motor technical errors and their impact on the risk of injury. We thus concurrently captured the kinematics of their tennis serve and the muscular activities of the upper-body. We recently validated that the Waiter's serve implies higher risk of injuries [27]. It is a movement that was know by the coaches as not productive and risky but it was never validated. Moreover, we evaluated the strategies of pacing use during five-set matches in the top tennis tournaments [20].

7.3. Virtual human simulation

7.3.1. Novel Distance Geometry based approaches for Human Motion Retargeting

Participants: Franck Multon [contact], Ludovic Hoyet, Antonio Mucherino, Zhiguang Liu.

Since September 2016, Antonio Mucherino has a half-time Inria detachment in the MimeTIC team (ended Sept 2018), in order to collaborate on exploring distance geometry-based problems in representing and editing human motion.

In this context, an extension of a distance geometry approach to dynamical problems was proposed in [24], and we co-supervised Antonin Bernardin for his Master thesis in 2017, which focused on applying such extended approach for retargeting human motions. In character animation, it is often the case that motions created or captured on a specific morphology need to be reused on characters having a different morphology. However, specific relationships such as body contacts or spatial relationships between body parts are often lost during this process, and existing approaches typically try to determine automatically which body part relationships should be preserved in such animation. Instead, we proposed a novel frame-based approach to

motion retargeting which relies on a normalized representation of all the body joints distances to encompass all the relationships existing in a given motion. In particular, we proposed to abstract postures by computing all the inter-joint distances of each animation frame and to represent them by Euclidean Distance Matrices (EDMs). Such EDMs present the benefits of capturing all the subtle relationships between body parts, while being adaptable through a normalization process to create a morphology independent distance-based representation. Finally, they can also be used to efficiently compute retargeted joint positions best satisfying newly imposed distances. We demonstrated that normalized EDMs can be efficiently applied to a different skeletal morphology by using a dynamical distance geometry approach, and presented results on a selection of motions and skeletal morphologies.

Concurrently, we proposed a pose transfer algorithm from a source character to a target character, without using skeleton information. Previous work mainly focused on retargeting skeleton animations whereas the contextual meaning of the motion is mainly linked to the relation- ship between body surfaces, such as the contact of the palm with the belly. In the context of the Inria PRE program, we propose a new context-aware motion retargeting framework [38], based on deforming a target character to mimic a source character poses using harmonic mapping. We also introduce the idea of Context Graph: modeling local interactions between surfaces of the source character, to be preserved in the target character, in order to ensure fidelity of the pose. In this approach, no rigging is required as we directly manipulate the surfaces, which makes the process totally automatic. Our results demonstrate the relevance of this automatic rigging-less approach on motions with complex contacts and interactions between the character's surface.

7.3.2. Investigating the Impact of Training for Example-Based Facial Blendshape Creation.

Participant: Ludovic Hoyet [contact].

In collaboration with Technicolor and Trinity College Dublin, we explored how certain training poses can influence the Example-Based Facial Rigging (EBFR) method [33]. We analysed the output of EBFR given a set of training poses to see how well the results reproduced our ground truth actor scans compared to a pure Deformation Transfer approach (Figure 5). We found that, while the EBFR results better matched the ground truth overall, there were certain cases that didn't see any improvement. While some of these results may be explained by lack of sufficient training poses for the area of the face in question, we found that certain lip poses weren't improved by training despite a large number of mouth training poses supplied. Our initial goal for this project was to identify what facial expressions are important to use as training when using Example-Based Facial Rigging to create facial rigs. This preliminary work has indicated certain parts of the face that might require more attention when automatically creating blendshapes, which still require to be further investigated, e.g., to identify a subset of facial expressions that would be considered the "ideal" subset to use for training the EBFR algorithm.

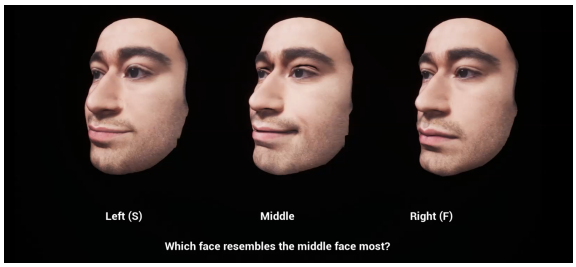


Figure 5. An example of stimuli shown to participants comparing the facial rigs created without training, and those with training.

7.4. Human motion in VR

7.4.1. Motion recognition and classification

Participants: Franck Multon, Richard Kulpa [contact], Yacine Boulahia.

Action recognition based on human skeleton structure represents nowadays a prospering research field. This is mainly due to the recent advances in terms of capture technologies and skeleton extraction algorithms. In this context, we observed that 3D skeleton-based actions share several properties with handwritten symbols since they both result from a human performance. We accordingly hypothesize that the action recognition problem can take advantage of trial and error approaches already carried out on handwritten patterns. Therefore, inspired by one of the most efficient and compact handwriting feature-set, we proposed a skeleton descriptor referred to as Handwriting-Inspired Features. First of all, joint trajectories are preprocessed in order to handle the variability among actor's morphologies. Then we extract the HIF3D features from the processed joint locations according to a time partitioning scheme so as to additionally encode the temporal information over the sequence. Finally, we used Support Vector Machine (SVM) for classification. Evaluations conducted on two challenging datasets, namely HDM05 and UTKinect, testify the soundness of our approach as the obtained results outperform the state-of-the-art algorithms that rely on skeleton data.

Being able to interactively detect and recognize actions based on skeleton data, in unsegmented streams, has become an important computer vision topic. It raises three scientific problems in relation with variability. The first one is the temporal variability that occurs when subjects perform gestures with different speeds. The second one is the inter-class spatial variability, which refers to disparities between the displacement amounts induced by different classes (i.e. long vs. short movements). The last one is the intra-class spatial variability caused by differences in style and gesture amplitude. Hence, we designed an original approach that better considers these three issues [15]. To address temporal variability we introduce the notion of curvilinear segmentation. It consists in extracting features, not on temporally-based sliding windows, but on segments in which the accumulated curvilinear displacement of skeleton joints equals a specific amount. Second, to tackle inter-class spatial variability, we define several competing classifiers with their dedicated curvilinear windows. Last, we address intraclass spatial variability by designing a fusion system that takes the decisions and confidence scores of every competing classifier into account. Extensive experiments on four challenging skeleton-based datasets demonstrate the relevance and efficiency of the proposed approach.

This work has been carried-out in collaboration with the IRISA Intuidoc team, with Yacine Boulahia who is a co-supervised PhD student with Eric Anquetil.

7.4.2. Automatic evaluation of sports gesture

Participant: Richard Kulpa [contact].

Automatically evaluating and quantifying the performance of a player is a complex task since the important motion features to analyze depend on the type of performed action. But above all, this complexity is due to the variability of morphologies and styles of both the experts who perform the reference motions and the novices. Only based on a database of expert's motions and no additional knowledge, we propose an innovative 2-level DTW (Dynamic Time Warping) approach to temporally and spatially align the motions and extract the imperfections of the novice's performance for each joints [23]. We applied our method on tennis serve and karate katas.

7.4.3. Studying the Sense of Embodiment in VR Shared Experiences

Participants: Rebecca Fribourg, Ludovic Hoyet [contact].

To explore how the sense of embodiment is influenced by the fact of sharing a virtual environment with another user, we conducted an experiment where users were immersed in a virtual environment while being embodied in an anthropomorphic virtual representation of themselves [36], in collaboration with Hybrid Inria team. In particular, two situations were studied: either users were immersed alone, or in the company of another user (see Figure 6). During the experiment, participants performed a virtual version of the well-known whac-a-mole game, therefore interacting with the virtual environment, while sitting at a virtual table. Our results show

that users were significantly more “efficient” (i.e., faster reaction times), and accordingly more engaged, in performing the task when sharing the virtual environment, in particular for the more competitive tasks. Also, users experienced comparable levels of embodiment both when immersed alone or with another user. These results are supported by subjective questionnaires but also through behavioural responses, e.g. users reacting to the introduction of a threat towards their virtual body. Taken together, our results show that competition and shared experiences involving an avatar do not influence the sense of embodiment, but can increase user engagement. Such insights can be used by designers of virtual environments and virtual reality applications to develop more engaging applications.

Figure 6. Setup of the experiment: each user was able to interact in the virtual environment with his own avatar, while the physical setup provided both a reference frame and passive haptic feedback. From left to right: experimental conditions Alone, Mirror and Shared; Physical setup of the experiment.

7.4.4. Biofidelity in VR

Participants: Simon Hilt, Charles Pontonnier, Georges Dumont [contact].

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user’s body parts. Whatever the system is, one of the main challenge is to be able to automatically recognize and analyze the user’s performance according to poor and noisy signals. Hence, recognizing and measuring human performance are important scientific challenges especially when using low-cost and noisy motion capture systems. MimeTIC has addressed the above problems in two main application domains. In this section, we detail the ergonomics application of such an approach. Firstly, in ergonomics, we explored the impact of uncertainties on friction coefficients on haptic feedback. The coefficients are tuned thanks to an experimental protocol enabling a subjective comparison between real and virtual manipulations of a low mass object. The compensation of friction on the first and second axes of the haptic interface showed significant improvement of both realism and perceived load. This year, we conducted experiments aiming at comparing gesture, recorded by an optoelectronic setup, and muscular activities, recorded by EMG, between real and virtual (with haptic feedback) manipulation.

7.5. Digital storytelling

7.5.1. Film Editing Patterns: Thinking like a Director

Participant: Marc Christie [contact].

We have introduced *Film Editing Patterns (FEP)*, a language to formalize film editing practices and stylistic choices found in movies. FEP constructs are constraints expressed over one or more shots from a movie sequence [29] that characterize changes in cinematographic visual properties such as shot size, region, angle of on-screen actors.

We have designed the elements of the FEP language, then introduced its usage in annotated film data, and described how it can support users in the creative design of film sequences in 3D. More specifically: (i) we proposed the design of a tool to craft edited filmic sequences from 3D animated scenes that uses FEPs to support the user in selecting camera framings and editing choices that follow certain best practices used in cinema; (ii) we conducted an evaluation of the application with professional and non-professional filmmakers. The evaluation suggested that users generally appreciate the idea of FEP, and that it can effectively help novice and medium experienced users in crafting film sequences with little training and satisfying results.

7.5.2. Directing Cinematographic Drones

Participants: Marc Christie [contact], Quentin Galvane.

We have designed a set of high-level tools for filming dynamic targets with quadrotor drones. To this end, we proposed a specific camera parameter space (the Drone Toric space) together with interactive on-screen viewpoint manipulators compatible with the physical constraints of a drone. We then designed a real-time path planning approach in dynamic environments which ensures both cinematographic properties in viewpoints along the path and ensures the feasibility of the path by a quadrotor drone. We finally have demonstrated how the Drone Toric Space can be combined with our path planning technique to coordinate positions and motions of multiple drones around dynamic targets to ensure the coverture of cinematographic distinct viewpoints. The proposed research prototypes have been evaluation by an experienced drone pilot and filmmaker, as well as by non-experts users. Not only does the tool demonstrate it's benefit in rehearsing complex camera moves for the film and documentary industries, but it demonstrates it's usability for everyday recording of aesthetic camera motions. The work was published in the Transactions on Graphics journal and was accepted for presentation at SIGGRAPH [18].

In addition we have focused on full automated and non-reactive path-planning for cinematographic drones. Most existing tools typically require the user to specify and edit the camera path, for example by providing a complete and ordered sequence of key viewpoints. In our contribution, we propose a higher level tool designed to enable even novice users to easily capture compelling aerial videos of large-scale outdoor scenes. Using a coarse 2.5D model of a scene, the user is only expected to specify starting and ending viewpoints and designate a set of landmarks, with or without a particular order. Our system automatically generates a diverse set of candidate local camera moves for observing each landmark, which are collision-free, smooth, and adapted to the shape of the landmark. These moves are guided by a landmark-centric view quality field, which combines visual interest and frame composition. An optimal global camera trajectory is then constructed that chains together a sequence of local camera moves, by choosing one move for each landmark and connecting them with suitable transition trajectories. This task is formulated and solved as an instance of the Set Traveling Salesman Problem. The work was published and presented at SIGGRAPH [30].

7.5.3. Automated Virtual Staging

Participants: Marc Christie [contact], Quentin Galvane, Fabrice Lamarche, Amaury Louarn.

While the topic of virtual cinematography has essentially focused on the problem of computing the best viewpoint in a virtual environment given a number of objects placed beforehand, the question of how to place the objects in the environment with relation to the camera (referred to as staging in the film industry) has received little attention. This work first proposes a staging language for both characters and cameras that extends existing cinematography languages with multiple cameras and character staging. Second, we propose techniques to operationalize and solve staging specifications given a 3D virtual environment. The novelty holds in the idea of exploring how to position the characters and the cameras simultaneously while maintaining a number of spatial relationships specific to cinematography. We demonstrate the relevance of our approach through a number of simple and complex examples [45].

7.5.4. VR Staging and Cinematography

Participants: Marc Christie [contact], Quentin Galvane.

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. In this work we propose a VR authoring system which provides intuitive ways of crafting visual sequences, both for expert animators and expert creatives in the animation and film industry. 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 [37].

7.5.5. Improving Camera tracking technologies

Participants: Marc Christie [contact], Xi Wang.

Robustness of indirect SLAM techniques to light changing conditions remains a central issue in the robotics community. With the change in the illumination of a scene, feature points are either not extracted properly due to low contrasts, or not matched due to large differences in descriptors. In this work, we propose a multi-layered image representation (MLI) in which each layer holds a contrast enhanced version of the current image in the tracking process in order to improve detection and matching. We show how Mutual Information can be used to compute dynamic contrast enhancements on each layer. We demonstrate how this approach dramatically improves the robustness in dynamic light changing conditions on both synthetic and real environments compared to default ORB-SLAM. This work focalises on the specific case of SLAM relocalisation in which a first pass on a reference video constructs a map, and a second pass with a light changed condition relocalizes the camera in the map [41], [40].

POTIOC Project-Team

7. New Results

7.1. Transition between AR and VR spaces

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

Mixed Reality systems combine physical and digital worlds, with great potential for the future of HCI. It is possible to design systems that support flexible degrees of virtuality by combining complementary technologies. In order for such systems to succeed, users must be able to create unified mental models out of heterogeneous representations. We conducted two studies focusing on the users' accuracy on heterogeneous systems using Spatial Augmented Reality (SAR) and immersive Virtual Reality (VR) displays (see Figure 4), and combining viewpoints (egocentric and exocentric). The results show robust estimation capabilities across conditions and viewpoints [31].

Figure 4. A user experiencing transitions between spatial augmented reality and virtual reality spaces.

7.2. Tangible and augmented interfaces for Schoolchildren

Participants: Philippe Giraudeau, Théo Segonds, Martin Hachet

In 2018, we have continued working on the exploration of tangible and augmented interfaces for Schoolchildren. We have notably evaluated the pedagogical potential of Teegi in a user study conducted at school [24].

We have also pursued our work on collaborative learning at school, part of the e-Tac project. In particular, based on focus group with children and practitioners, we have refined our interactive pedagogical environment, and we have implemented a new version (see Figure 5) [53]

7.3. Ambient interfaces dedicated to the awareness of energy consumption

Participants: Pierre-Antoine Cinquin, Philippe Giraudeau

Figure 5. Tangible and augmented objects to foster collaborative learning at school.

Inspired by studies in data physicalization, we explore the use of tangible and ambient interfaces to raise people's awareness of energy consumption. As a first approach, we are developing an interactive and collaborative environment named Erlen. This year, we have designed a first prototype taking the form of an Erlenmeyer flask with fluid simulation. Through manipulation, users can visualize information about their electricity consumption. This prototype was demonstrated at IHM 2018 [21]. Based on the feedback we obtained, we are actually developing a new set of individuals and shared interfaces along with new interactions.

7.4. Drones for Human interaction

Participants: Rajkumar Darbar, Anke Brock, Martin Hachet.

We have also continued working with drones. In particular, we have proposed FlyMap as a novel user experience for interactive maps projected from a drone. We iteratively designed three interaction techniques for FlyMap's usage scenarios. In a comprehensive indoor study (N = 16), we show the strengths and weaknesses of the techniques on users' cognition, task load and satisfaction. We then pilot tested FlyMap outdoors in real world conditions with four groups of participants. We show that its interactivity is exciting to users, opening the space for more direct interactions with drones [20].
We are currently exploring the use of drones to bring passive haptic feedback in immersive VR scenario. Concretely, we are building a system where drones, equipped with flat panels, co-locate themselves with virtual objects to provide physical feedbacks to VR users.

7.5. Mixed reality based interfaces for visual impaired persons

Participant: Lauren Thévin, Anke Brock

Current low-tech Orientation & Mobility (O&M) tools for visually impaired people, e.g. tactile maps, possess limitations. Interactive accessible maps have been developed to overcome these. However, most of them are limited to exploration of existing maps, and have remained in laboratories. Using a participatory design approach, we have worked closely with 15 visually impaired students and 3 O&M instructors over 6 months. We iteratively designed and developed an augmented reality map destined at use in O&M classes in special education centers. This prototype combines projection, audio output and use of tactile tokens, and thus allows both map exploration and construction by low vision and blind people. Our user study demonstrated that all students were able to successfully use the prototype, and showed a high user satisfaction. A second phase with 22 international special education teachers allowed us to gain more qualitative insights. This work shows that augmented reality has potential for improving the access to education for visually impaired people [18].

We have pursued this work to make the visual and audio augmentation of real objects easy and convenient. In a user study, six teachers created their own audio-augmentation of objects, such as a botanical atlas (Figure 6 , within 30 minutes or less. Teachers found the tool easy to use and were confident about re-using it. Participants found the resulting interactive graphics exciting to use independently of their mental imagery skills [32].

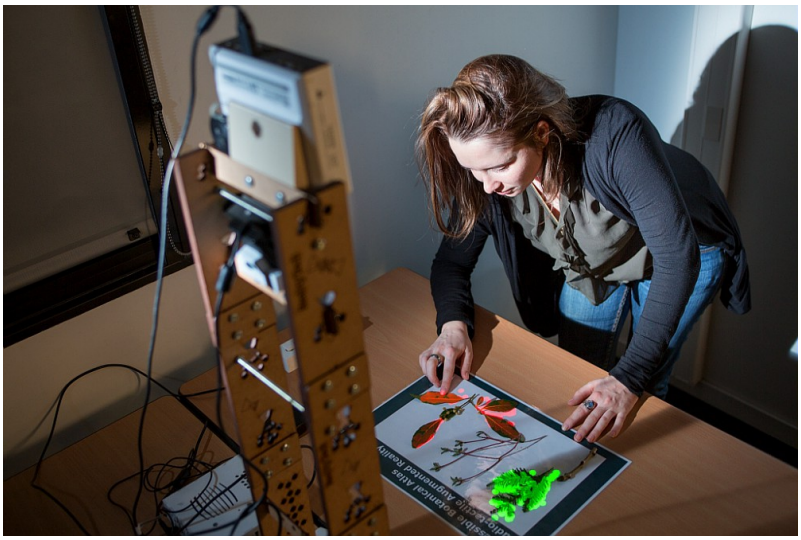


Figure 6. Multimodal Augmented Reality for visual impaired students.

7.6. Accessibility of e-learning systems

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

In 2018, 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 [11] 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. 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 9000 registered participants). https://mooc-francophone.com/cours/mooc-accessibilite-numerique/. Our first field studies demonstrate the benefits of using Aïana for disabled participants [22].

7.7. Improving EEG Signal Processing for Brain-Computer Interfaces

Participants: Aurélien Appriou, Satyam Kumar, Fabien Lotte

Figure 7. The Aïana MOOC player.

A review of classification algorithms for BCI: Most current Electroencephalography (EEG)-based Brain-Computer Interfaces (BCIs) are based on machine learning algorithms. We surveyed the BCI and machine learning literature to identify the classification approaches that have been investigated to design BCIs. We found that the recently designed classification algorithms for EEG-based BCIs can be divided into four main categories: adaptive classifiers, matrix and tensor classifiers, transfer learning and deep learning, plus a few other miscellaneous classifiers. Among these, adaptive classifiers were demonstrated to be generally superior to static ones, even with unsupervised adaptation. Transfer learning can also prove useful although the benefits of transfer learning remain unpredictable. Riemannian geometry-based methods have reached state-of-the-art performances on multiple BCI problems and deserve to be explored more thoroughly, along with tensor-based methods. Shrinkage linear discriminant analysis and random forests also appear particularly useful for small training samples settings. On the other hand, deep learning methods have not yet shown convincing and consistent improvement over state-of-the-art BCI methods. This survey was published in Journal of Neural Engineering in [14].

Exploring Modern Machine Learning Methods to Estimate Mental Workload From EEG Signals: Estimating mental workload from brain signals such as EEG has proven very promising in multiple HCI applications, e.g., to design games or educational applications with adaptive difficulty. However, currently obtained workload classification accuracies are relatively low, making the resulting estimations not fully trustable. We thus studied promising modern machine learning algorithms, including Riemannian geometry-based methods and Convolutional Neural Networks, to estimate workload from EEG signals. We studied them with both user-specific and user-independent calibration, to go towards calibration-free systems. Our results suggested that a shallow Convolutional Neural Network obtained the best performance in both conditions, outperforming state-of-the-art methods on the used data sets. This work was published as a work-in-progress in the CHI conference [19].

BCPy, an open-source python platform for offline EEG signals decoding and analysis: Although promising, BCIs are still barely used outside laboratories due to their poor robustness. Moreover, they are sensitive to noise, outliers and the non-stationarity of EEG signals. Many algorithms have been developed for EEG signals processing and classification, in order to improve BCIs robustness. We proposed BCPy, an open-source, easy-to-use python BCI platform for offline EEG signal analysis. Python is free and contains good scalable libraries for scientific computing. Moreover, Python is the major language used to implement recent advances

in ML and Deep Learning, thus making them easily available for BCI research. This work was published in the International BCI meeting [48].

Adaptive Riemannian classification methods: The omnipresence of non-stationarity and noise in EEG signals restricts the ubiquitous use of BCIs. One of the possible ways to tackle this problem is to adapt the computational model used to detect and classify different mental states. Adapting the model will possibly help us to track the changes and thus reducing the effect of non-stationarities. In this paper, we present different adaptation strategies for state of the art Riemannian geometry based classifiers. The offline evaluation of our proposed methods on two different datasets showed a statistically significant improvement over baseline non-adaptive classifiers. Moreover, we also demonstrate that combining different (hybrid) adaptation strategies generally increased the performance over individual adaptation schemes. Also, the improvement in average classification accuracy for a 3-class mental imagery BCI with hybrid adaption is as high as around 17% above the baseline non-adaptive classifier. This was published in [26].

Regularized spatial filters for EEG regression problems: In collaboration with University Freiburg, we reported on novel supervised algorithms for single-trial brain state decoding. When brain activity is assessed by multichannel recordings, spatial filters computed by the source power comodulation (SPoC) algorithm allow identifying oscillatory subspaces. In small dataset scenarios, this supervised method tends to overfit to its training data. To improve upon this, we proposed and characterize three types of regularization techniques for SPoC. Evaluating all methods on real-world data, we observed an improved regression performance mainly for datasets from subjects with initially poor performance. This was published in the Neuroinformatics journal [16].

SEREEGA: a toolbox to Simulate EEG activity: EEG is a popular method to monitor brain activity, but it is difficult to evaluate EEG-based analysis methods because no ground-truth brain activity is available for comparison. Therefore, to test and evaluate such methods, in collaboration with TU Berlin, we proposed SEREEGA, a free and open-source matlab toolbox for Simulating Event-Related EEG Activity. The toolbox is available at https://github.com/lrkrol/SEREEGA. SEREEGA unifies the majority of past simulation methods reported in the literature into one toolbox. This toolbox and its use were published in journal of neuroscience methods [13].

7.8. Understanding Brain-Computer Interfaces user Training

Participants: Léa Pillette, Camille Benaroch, Fabien Lotte

Computational models of BCI performance: Mental-Imagery based BCIs (MI-BCIs) use signals produced during mental imagery tasks to control the system. Current MI-BCIs are rather unreliable, which is due at least in part to the use of inappropriate user-training procedures. Understanding the processes underlying user-training by modelling it computationally could enable us to improve MI-BCI training protocols and adapt the latter to the profile of each user. Indeed, we developed theoretical and conceptual models of BCI performances suggesting that the users' profiles does impact their performances [12]. Our objective is to create a statistical/probabilistic model of training that could explain, if not predict, the learning rate and the performances of a BCI user over training time using user's personality, skills, state and timing of the experiment. Preliminary analyses on previous data revealed positive correlations between MI-BCI performances and mental rotation scores among two of three different studies based on the same protocol [49]. This suggests that spatial abilities play a major role in MI-BCI users' abilities to learn to perform MI tasks, which is consistent with the literature.

Modeling and measuring users' skills at MI-BCI control: Studying and improving the reliability issue of BCI requires the use of appropriate reliability metrics to quantify both the classification algorithm and the BCI user's performances. So far, Classification Accuracy (CA) is the typical metric used for both aspects. However, we argued that CA is a poor metric to study BCI users' skills. Thus, we proposed a definition and new metrics to quantify such BCI skills for MI-BCIs, independently of any classification algorithm. By re-analyzing EEG data sets with such new metrics, we indeed confirmed that CA may hide some increase in MI-BCI skills or hide the user inability to self-modulate a given EEG pattern. On the other hand, our new metrics could reveal such skill improvements as well as identify when a mental task performed by a user was no different than rest EEG. This work was published in Journal of Neural Engineering [15].

Towards measuring the impact of attention: "Attention" is a generic word encompasses alertness and sustained attentions, referring to the intensity of attention (i.e., strength), as well as selective and divided attentions, referring to its selectivity (i.e., amount of monitored information). BCI literature indicates an influence of both users' attention traits and states (i.e., respectively stable and unstable attentional characteristics) on the ability to control a BCI. Though the types of attention involved remain unclear. Therefore, assessing which types of attention are involved during BCI use might provide information to improve BCI usability. Before testing this hypothesis, we first needed to assess if the different types of attention are recognizable using EEG. Our first results suggested that indeed, using machine learning, we can discriminate attention types for each other in EEG, at least when comparing them two by two [59].

The Influence of the experimenter: Through out the research and development process of MI-BCI, human supervision (e.g., experimenters or caregivers) plays a central role. People need to present the technology to users and ensure the smooth progress of the BCI learning and use. Though, very little is known about the influence they might have on their results. Such influence is to be expected as social and emotional feedback were shown to influence MI-BCI performances and user experience. Furthermore, literature from different fields indicate an effect of experimenters, and specifically their gender, on experiment outcome. Therefore, we assessed the impact of gender on MI-BCI performances, progress and user experience. An interaction of the runs, subjects gender and experimenters gender was found to have an impact on the performances of the subjects, suggesting users learn better with female experiments [30] (see Fig. 8).

7.9. Improving BCI user performance and training

Participants: Jelena Mladenovic, Léa Pillette, Thibaut Monseigne, Fabien Lotte

The potential of learning companions: As mentioned before, current BCI training protocols do not enable every user to acquire the skills required to use BCIs. We showed that learning companions were promising tools to increase BCI user experience during training, as well as to increase the performances of users who are more inclined to work in a group. Encouraged by these first results we investigated all the other potential benefits learning companions could bring to BCI training by improving the feedback, i.e., the information provided to the user, which is primordial to the learning process and yet have proven both theoretically and practically inadequate in BCI. From these considerations, some guidelines were drawn, open challenges identified and potential solutions were suggested to design and use learning companions for BCIs [29].

Figure 8. An EEG cap is being placed on the head of a subject by an experimenter on the right while another experimenter on the left is setting up the necessary software on the computer.

Active Inference for P300 speller: Brain Computer Interface (BCI) mostly relies, on one hand, on the stability of a person's mental commands, and on the other, on the machine's capacity to interpret those commands. As a person is naturally changing and adapting all the time, the machine becomes less successful in interpreting user's commands. In turn, the machine should be able to predict and minimize undesired user fluctuations. Moreover, it should build bottom-up information about the user through physiological input (EEG observations), and influence the user by providing optimal task (action) to minimize prediction error. A novel neuroscience approach, Active (Bayesian) Inference, is a very generic and flexible computational framework that can predict user intentions through a series of optimal actions and observations. On simulated data, we have shown that Active Inference has great potential to enable the machine to co-adapt with the user, and increase performance levels in a P300 speller BCI. We further tested Active Inference on real data, and show that active inference surpasses the standard algorithms while permitting the implementation of various cases of p300 speller BCI within one single framework [57].

Towards Congruent Feedback for BCI: Congruent visual environment in MI BCI has been researched in virtual reality, giving a sense of body ownership illusion, and showed to be more robust and improve performance. On the other hand, the effects of a congruent, purely audio environment, have not yet been explicitly explored in BCI. This inspired us to explore the benefits of a task-related (congruent) and synchronised audio feedback which would comply with the user's imagined movements. We investigate the potential of such an audio feedback congruent to the task, tackling the sensory illusion of presence by providing realistic audio feedback using natural sounds. Our preliminary results show the benefits of a congruent, audio MI feedback of feet (sound of footsteps in gravel) as opposed to no congruent feedback using abstract sound [50].

Neurofeedback of daytime alertness: Neurofeedback consists in providing a subject with information about his own EEG by means of a sensory feedback (visual, auditory ...) in real-time, in order to enable cognitive learning. In collaboration with SANPSY (Pellegrin Hospital/Univ. Bordeaux), we implemented a complete Neurofeedback solution as a proof of concept that aims to determine the level of effectiveness of Neurofeedback on daytime alertness ability. Indeed, excessive daytime sleepiness (EDS) is a common complaint associated with increased accidental risk. The usual countermeasures such as blue light, caffeine or nap have been shown to be effective but have limitations. With a test on five subjects, preliminary data showed that it was possible to learn how to regulate our own EEG activity with a short number of sessions (8 sessions of 40 min). Clinical trials to confirm these results should be initiated in the course of 2019.

7.10. Physiological computing

Participants: Jelena Mladenovic, Fabien Lotte

ElectroGastoGraphy: Recent research in the enteric nervous system, sometimes called the second brain, has revealed potential of the digestive system in predicting emotions. Even though people regularly experience changes in their gastrointestinal (GI) tract which influence their mood and behavior multiple times per day, robust measurements and wearable devices are not quite developed for such phenomena. However, other manifestations of the autonomic nervous system such as electrodermal activity, heart rate, and facial muscle movement have been extensively used as measures of emotions or in biofeedback applications, while neglecting the gut. In [28], we exposed electrogastrography (EGG), i.e., recordings of the myoelectric activity of the GI tract, as a possible measure for inferring human emotions.

EEG-based neuroergonomics: In collaboration with ISAE Toulouse, we explored the use of EEG to monitor cognitive processes in real flight situation, using dry EEG sensors. We showed that doing so is possible, however with low performances, given the strong noise in signals occurring in this challenging context [23]. In general, we presented in [17] and [40] how BCIs could be useful for neuroergonomics, i.e., to estimate user interfaces ergonomics quality from neurophysiological measures. Finally, in collaboration with RIKEN BSI, Japan, we showed that emotions could be monitored to some extent in EEG signals from multiple users watching the same emotional video clips at the same time. Interestingly, emotions decoding performances were increased by using the EEG data from several users compared to using EEG from each individual user [33].

TITANE Project-Team

7. New Results

7.1. Analysis

7.1.1. Planar Shape Detection at Structural Scales

Participants: Hao Fang, Mathieu Desbrun, Florent Lafarge [contact].

Shape detection, abstraction, man-made objects, point clouds, surface reconstruction.

Interpreting 3D data such as point clouds or surface meshes depends heavily on the scale of observation. Yet, existing algorithms for shape detection rely on trial-and-error parameter tunings to output configurations representative of a structural scale. We present a framework to automatically extract a set of representations that capture the shape and structure of man-made objects at different key abstraction levels. A shape-collapsing process first generates a fine-to-coarse sequence of shape representations by exploiting local planarity. This sequence is then analyzed to identify significant geometric variations between successive representations through a supervised energy minimization. Our framework is flexible enough to learn how to detect both existing structural formalisms such as the CityGML Levels Of Details, and expert-specified levels of abstraction. Experiments on different input data and classes of man-made objects, as well as comparisons with existing shape detection methods, illustrate the strengths of our approach in terms of efficiency and flexibility. Figure 1 illustrates the goal of our method. This work has been published in the proceedings of CVPR [16].

7.1.2. Multi-task Deep Learning for Satellite Image Pansharpening and Segmentation

Participants: Andrew Khalel, Onur Tasar, Yuliya Tarabalka [contact].

This work has been done in collaboration with Dr. Guillaume Charpiat (TAU team, Inria Saclay).

Segmentation, pansharpening, multi-task, joint learning

We proposed a novel multi-task framework to learn satellite image pansharpening and segmentation jointly. Our framework is based on encoder-decoder architecture, where both tasks share the same encoder but each one has its own decoder (see Fig. 2). We compare our framework against single-task models with different architectures. Results show that our framework outperforms all other approaches in both tasks.

7.1.3. Incremental Learning for Semantic Segmentation of Large-Scale Remote Sensing Data

Participants: Onur Tasar, Pierre Alliez, Yuliya Tarabalka [contact].

This work has been done in collaboration with CNES and ACRI-ST.

Incremental learning, catastrophic forgetting, semantic segmentation, convolutional neural networks

In spite of remarkable success of the convolutional neural networks on semantic segmentation, they suffer from catastrophic forgetting: 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. For remembering, we either regularly feed samples from the stored, little fraction of the previous data or use the memory network, depending on

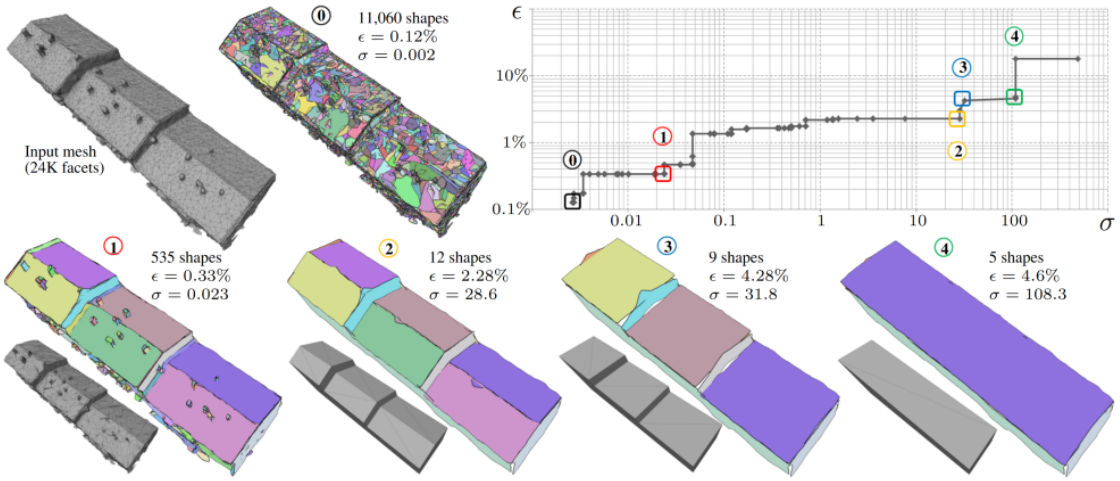


Figure 1. Planar shape detection at structural scales. Starting from 3D data (here a dense mesh generated by MultiView Stereo, top left), our algorithm produces a set of high-level representations with planar primitives (representations 1–4) describing the object at different representative structural scales (bottom). By progressively merging planar regions of an initial state (representation 0), one creates a sequence of representations whose further analysis allows for the extraction of a few structurally relevant representations (top right). Such shape representations can be used, for instance, as input for piecewise-planar reconstruction (see grey compact meshes). Note that each shape is displayed as a colored polygon computed as the α-shape of its inliers projected onto the shape.

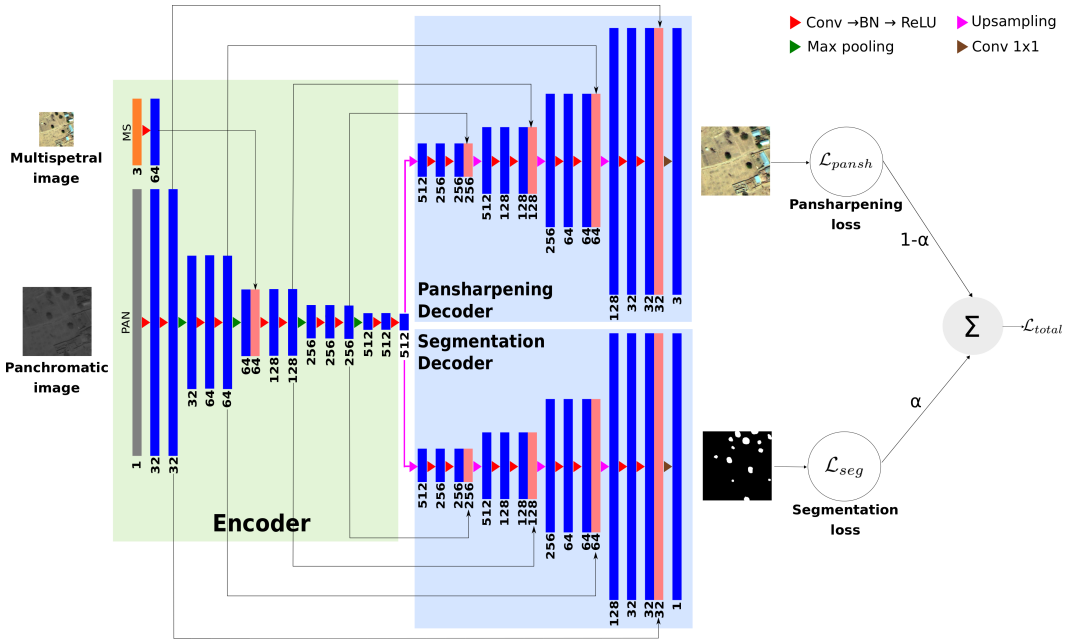


Figure 2. Overal framework for joint segmentation and pansharpening.

whether the new dat (see Fig. 3)a are collected from completely different geographic areas or from the same city. 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 submitted to IEEE Transactions on Geoscience and Remote Sensing (TGRS) and is currently on arXiV [25].

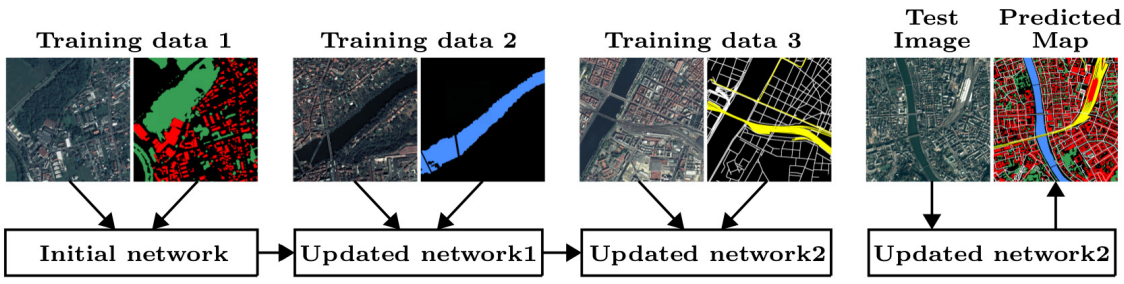


Figure 3. An example incremental learning scenario. Firstly, satellite images as well as their label maps for building and high vegetation classes are fed to the network. Then, from the second training data, the network learns water class without forgetting building and high vegetation classes. Finally, 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 comes, the network is able to detect all the classes.

7.1.4. Multimodal Image Alignment through a Multiscale Chain of Neural Networks with Application to Remote Sensing

Participants: Nicolas Girard, Yuliya Tarabalka [contact].

This work has been done in collaboration with Armand Zampieri and Dr. Guillaume Charpiat (TAO team, Inria Saclay).

Multimodal, Alignment, Registration, Remote sensing

We tackle here the problem of multimodal image non-rigid registration, which is of prime importance in remote sensing and medical imaging. The difficulties encountered by classical registration approaches include feature design and slow optimization by gradient descent. By analyzing these methods, we note the significance of the notion of scale. We design easy-to-train, fully-convolutional neural networks able to learn scale-specific features. Once chained appropriately, they perform global registration in linear time, getting rid of gradient descent schemes by predicting directly the deformation.

We show their performance in terms of quality and speed through various tasks of remote sensing multimodal image alignment. In particular, we are able to register correctly cadastral maps of buildings as well as road polylines onto RGB images, and outperform current keypoint matching methods (see Fig. 4). This work has been published in the proceedings of ECCV [20].

7.1.5. Aligning and Updating Cadaster Maps with Aerial Images by Multi-Task, Multi-Resolution Deep Learning

Participants: Nicolas Girard, Yuliya Tarabalka [contact].

This work has been done in collaboration with Dr. Guillaume Charpiat (TAO team, Inria Saclay).

Alignment, Registration, Multi-task, Multi-resolution

A large part of the world is already covered by maps of buildings, through projects such as OpenStreetMap. However when a new image of an already covered area is captured, it does not align perfectly with the buildings of the already existing map, due to a change of capture angle, atmospheric perturbations, human error when annotating buildings or lack of precision of the map data. Some of those deformations can be partially corrected, but not perfectly, which leads to misalignments. Additionally, new buildings can appear in the image. Leveraging multi-task learning, our deep learning model aligns the existing building polygons to the new image through a displacement output, and also detects new buildings that do not appear in the cadaster through a segmentation output (see Fig. 5). It uses multiple neural networks at successive resolutions to output a displacement field and a pixel-wise segmentation of the new buildings from coarser to finer scales. We also apply our method to buildings height estimation, by aligning cadaster data to the rooftops of stereo images.

7.2. Reconstruction

7.2.1. Kinetic Polygonal Partitioning of Images

Participants: Jean-Philippe Bauchet, Florent Lafarge [contact].

Polygons, image segmentation, object contouring, kinetic framework

Recent works showed that floating polygons can be an interesting alternative to traditional superpixels, especially for analyzing scenes with strong geometric signatures, as man-made environments. Existing algorithms produce homogeneously-sized polygons that fail to capture thin geometric structures and over-partition large uniform areas. We propose a kinetic approach that brings more flexibility on polygon shape and size. The key idea consists in progressively extending pre-detected line-segments until they meet each other. Our experiments demonstrate that output partitions both contain less polygons and better capture geometric structures than those delivered by existing methods. We also show the applicative potential of the method when used as preprocessing in object contouring. Figure 6 illustrates the goal of our method. This work has been published in the proceedings of CVPR [15].

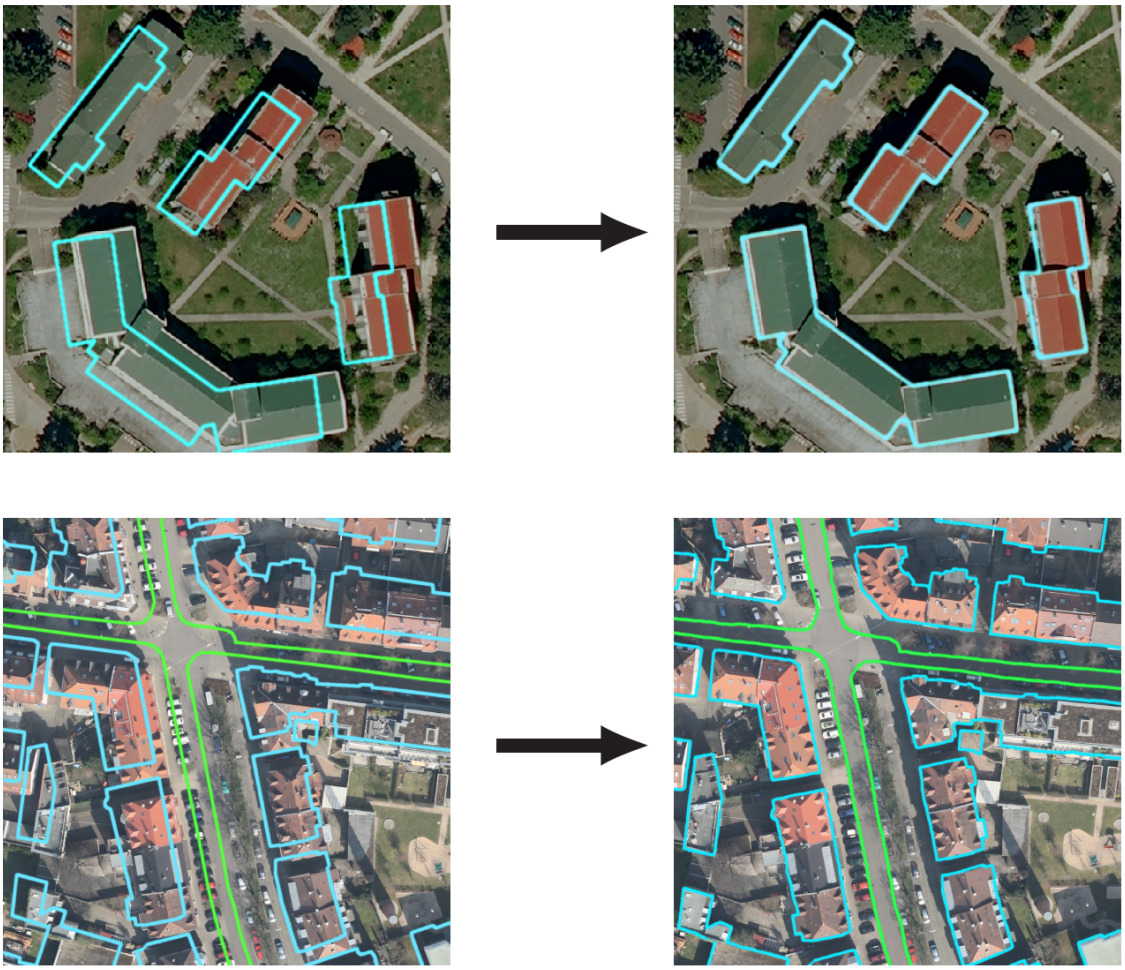


Figure 4. Multi-modal alignment.

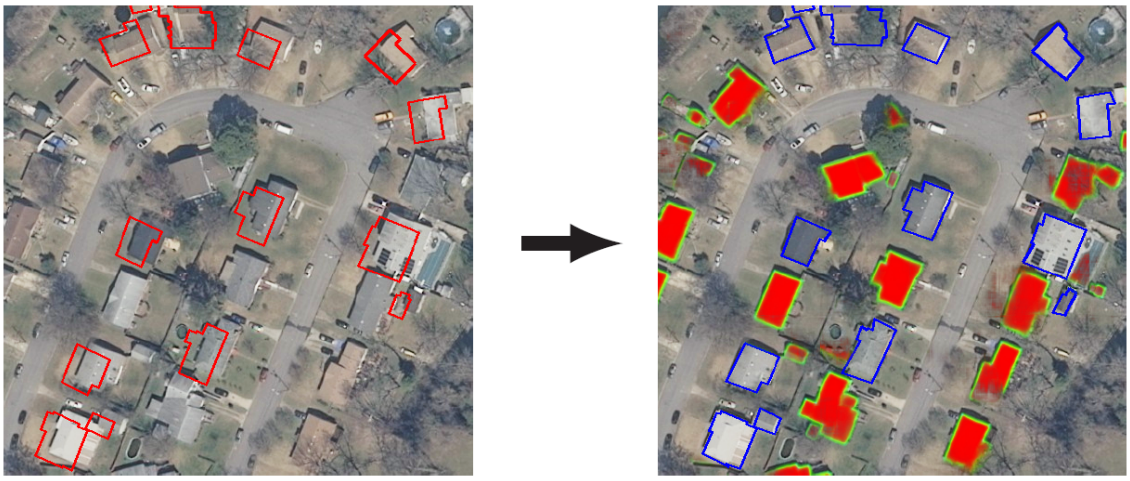


Figure 5. Multi-task learning with 2 tasks: multi-modal alignment and semantic segmentation.

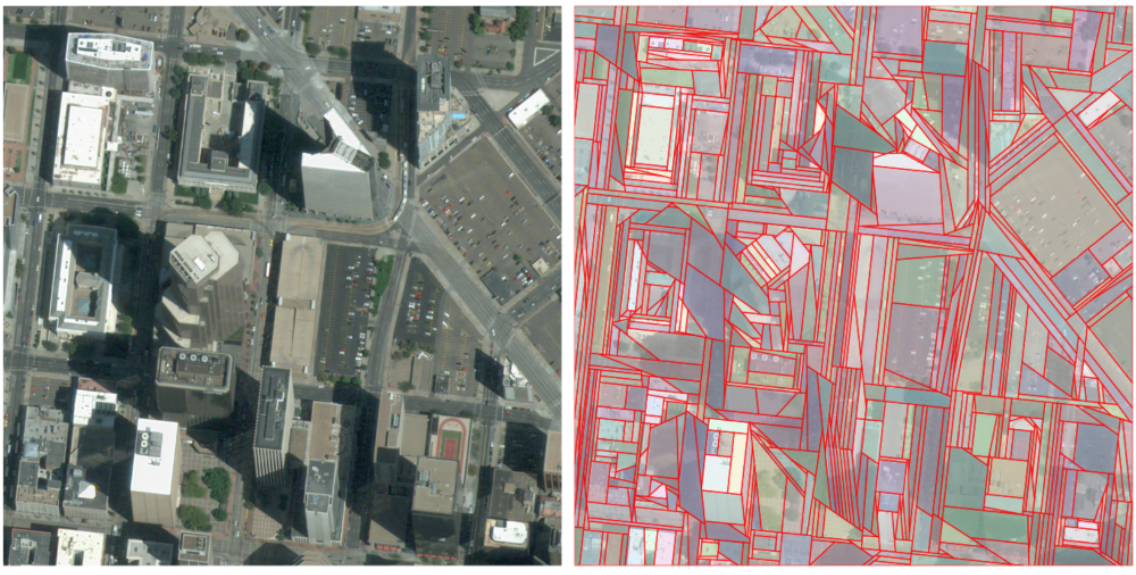


Figure 6. Kinetic partitioning into polygons. Our algorithm decomposes an image (left) into a partition of convex polygons (right). While superpixel-based methods impose homogeneously-sized regions, our polygons are more meaningful, capturing both large components and thin lineic structures that compose, for instance, urban scenes.

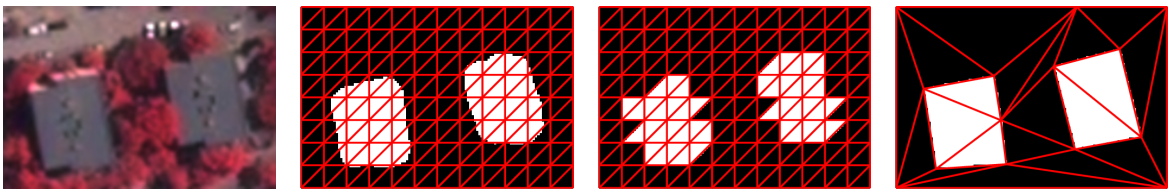
7.2.2. Polygonization of Binary Classification Maps Using Mesh Approximation with Right Angle Regularity

Participants: Onur Tasar, Pierre Alliez, Yuliya Tarabalka [contact].

Work in collaboration with Emmanuel Maggiori.

Polygonization, vectorization, remote sensing, classification maps, mesh approximation, right angles

One of the most popular and challenging tasks in remote sensing applications is the generation of digitized representations of Earth's objects from satellite raster image data. A common approach to tackle this challenge is a two-step method that first involves performing a pixel-wise classification of the raster data, then vectorizing the obtained classification map. We propose a novel approach, which recasts the polygonization problem as a mesh-based approximation of the input classification map, where binary labels are assigned to the mesh triangles to represent the building class. A dense initial mesh is decimated and optimized using local edge and vertex-based operators in order to minimize an objective function that models a balance between fidelity to the classification map in ℓ_1 norm sense, right angle regularity for polygonized buildings, and final mesh complexity (see Fig. 7). Experiments show that adding the right angle objective yields better representations quantitatively and qualitatively than previous work and commonly used polygon generalization methods in remote sensing literature for similar number of vertices. This work was published at IGARSS [19].



(a) (b) (c) (d)

Figure 7. Input image and example labeled meshes. (a) Input image, (b) Initial fine lattice, (c) Initial and (d) Optimized labeled triangle meshes. The triangles labeled as building are indicated by white.

7.2.3. End-to-End Learning of Polygons for Remote Sensing Image Classification

Participants: Nicolas Girard, Yuliya Tarabalka [contact].

High-resolution aerial images, polygon, vectorial, regression, deep learning, convolutional neural networks

While geographic information systems typically use polygonal representations to map Earth's objects, most state-of-the-art methods produce maps by performing pixelwise classification of remote sensing images, then vectorizing the outputs. This work studies if one can learn to directly output a vectorial semantic labeling of the image. We here cast a mapping problem as a polygon prediction task, and propose a deep learning approach which predicts vertices of the polygons outlining objects of interest. Experimental results on the Solar photovoltaic array location dataset show that the proposed network succeeds in learning to regress polygon coordinates, yielding directly vectorial map outputs (see Fig. 8). This work has been published in the proceedings of IGARSS [14].

7.3. Approximation

7.3.1. Curved Optimal Delaunay Triangulation

Participants: Mathieu Desbrun, Pierre Alliez [contact].

Work in collaboration with Leman Feng (Ecole des Ponts ParisTech), Hervé Delingette (EPIONE) and Laurent Busé (AROMATH).

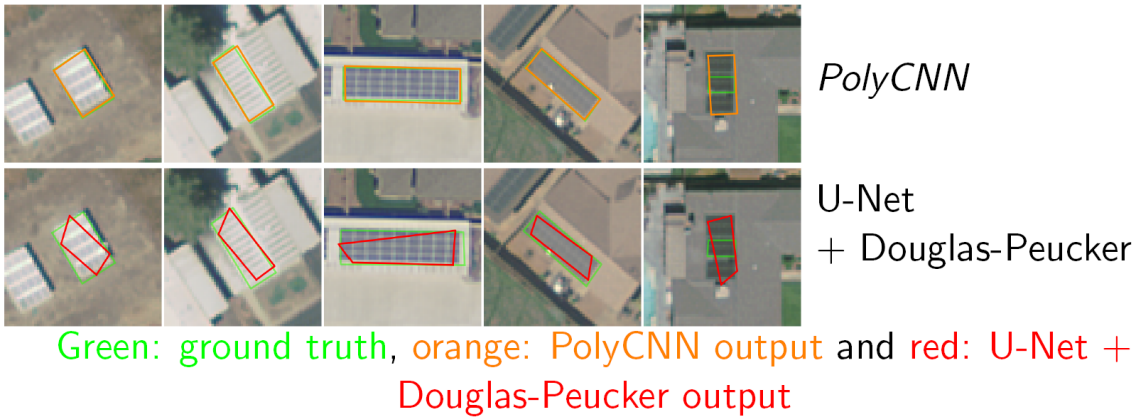


Figure 8. Results of polygon reconstruction with a deep neural network.

Higher-order meshing, Optimal Delaunay Triangulations, higher order finite elements, Bézier elements.

Meshes with curvilinear elements hold the appealing promise of enhanced geometric flexibility and higher-order numerical accuracy compared to their commonly-used straight-edge counterparts. However, the generation of curved meshes remains a computationally expensive endeavor with current meshing approaches: high-order parametric elements are notoriously difficult to conform to a given boundary geometry, and enforcing a smooth and non-degenerate Jacobian everywhere brings additional numerical difficulties to the meshing of complex domains. In this paper, we propose an extension of Optimal Delaunay Triangulations (ODT) to curved and graded isotropic meshes. By exploiting a continuum mechanics interpretation of ODT instead of the usual approximation theoretical foundations, we formulate a very robust geometry and topology optimization of Bézier meshes based on a new simple functional promoting isotropic and uniform Jacobians throughout the domain. We demonstrate that our resulting curved meshes can adapt to complex domains with high precision even for a small count of elements thanks to the added flexibility afforded by more control points and higher order basis functions (see Figure 9). This work has been published in the proceedings of ACM SIGGRAPH conference [12].

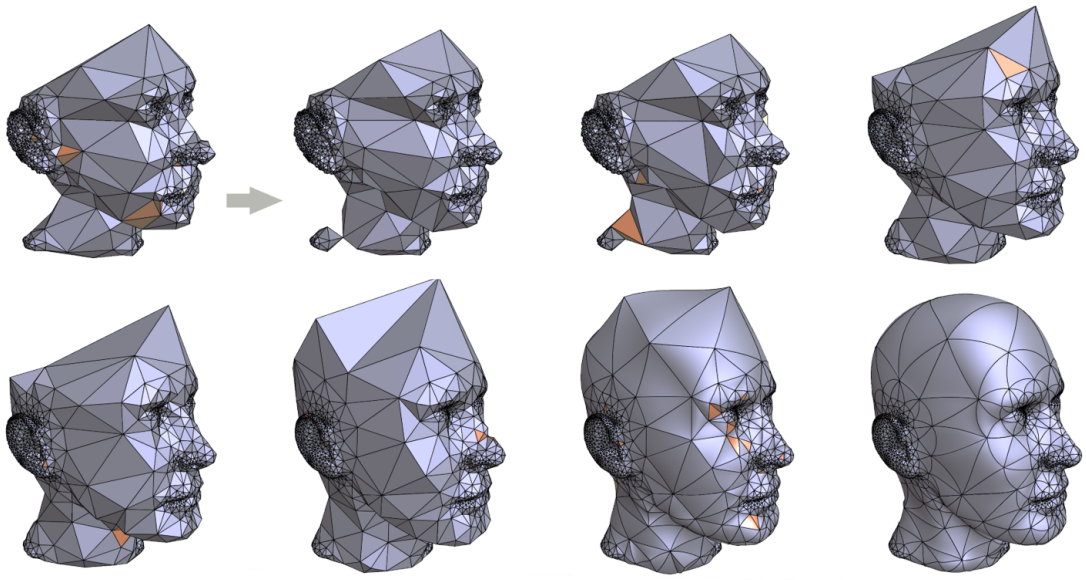


Figure 9. Generation of a curved optimal Delaunay triangulation.

ALMAnaCH Team

6. New Results

6.1. Syntax modelling and treebank development

Participants: Djamé Seddah, Benoît Sagot, Éric Villemonte de La Clergerie, Emilia Verzeni, Wigdan Abbas Mekki Medeni, Elias Benaïssa, Farah Essaidi, Amal Fethi.

- In 2018, members of ALMAnaCH have finalised a conversion of the biggest annotated data set for French, the French Treebank, to Universal Dependencies 2.3, the now *de facto* standard for syntactic annotations [27]. The same group was also deeply involved in a proposal co-written with others leaders of the field [25], aiming at representing morpho-syntactic ambiguities from user-generated content and morphologically-rich languages. This proposal was implemented via the development of language specific analysers and data-driven normalised lexica [26].
- As part of the ANR Parsiti project, the development of gold standards for North-African dialectal Arabic has seen great progresses and is coming to a pre-release date in the first semester of 2019. This work involved more than 24 man.months over the last 12 months and will culminate with a multi-layered corpus of about 2000 sentences that is made of user-generated content with a highly variable dialect that contains up to 36% of French words and mixed syntax with Arabic. In order to assess the quality of the translation produced by the Parsiti project, we also included a translation layer (North-African Arabic-French) as well as all expected morpho-syntactic and syntactic annotations, following the state-of-the-art in terms of annotations. Papers are currently being written and will target the main NLP conferences of early 2019.
- In parallel to the last item, we also translated to English half of the French Social Media Bank which was developed in our previous project [92]. A morpho-syntactic annotation layer was added. The crucial difficulty was to maintain a symmetry in term of style and level of languages between French user-generated content and its English counterpart. This data set is currently being used in the Parsiti project in order to evaluate the MT models currently being developed by the LIMSI partner.

6.2. Modeling of language variability via diachronic embeddings and extra-linguistic contextual features

Participants: Djamé Seddah, Benjamin Muller, Ganesh Jawahar, Benoît Sagot, Éric Villemonte de La Clergerie.

Following ALMAnaCH's participation in the 2017 CoNLL shared task on heavily multilingual dependency parsing in the *Universal Dependency* (hereafter UD) framework (we ranked 3rd/33 on part-of-speech tagging and 6th/33 on parsing), the team has taken part in the 2018 edition of the shared task. This year, most of the work was carried out by junior members of the team, for whom it was an interesting opportunity to gain experience on the development of NLP architectures and their deployment in the context of a shared task. It was also the opportunities to test new ideas.

We developed a neural dependency parser and a neural part-of-speech tagger, which we called 'ELMoLex' [21]. We augmented the deep Biaffine (BiAF) parser [64] with novel features to perform competitively: we utilize an in-domain version of ELMo features [77], which provide context-dependent word representations; we utilised disambiguated, embedded, morphosyntactic features extracted from our UD-compatible lexicons [26], which complements the existing feature set. In addition to incorporating character embeddings, ELMoLex leverages pre-trained word vectors, ELMo and morphosyntactic features (whenever available) to correctly handle rare or unknown words which are prevalent in languages with complex morphology. ELMoLex ranked 11th in terms of the Labeled Attachment Score metrics (70.64%) and the Morphology-aware LAS metrics (55.74%), and ranked 9th in terms of Bilexical dependency metric (60.70%). In an extrinsic evaluation setup, ELMoLex ranked 7th for Event Extraction, Negation Resolution tasks and 11th for Opinion Analysis task in terms of F1 score.

6.3. Modelling of language variability via diachronic embeddings and extra-linguistic contextual features

Participants: Djamé Seddah, Ganesh Jawahar, Éric Villemonte de La Clergerie, Benoît Sagot.

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 in 2018 into studying language variability (i.e. how the language evolve over time and how this evolution is tied to socio-demographic and dynamic network variables). Taking advantages of the SoSweet corpus (220 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. We devised a novel attentional model, based on Bernouilli word embeddings, that are conditioned on contextual extra-linguistic (social) features such as network, spatial and socio-economic variables, which are associated with Twitter users, 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. A paper on this work is currently under review.

6.4. Standardisation of Natural Language data

Participants: Laurent Romary, Jack Bowers, Charles Riondet, Mohamed Khemakhem, Benoît Sagot, Loïc Grobol.

One essential aspect of working with human traces as they occur in digital humanities at large and in natural language processing in particular, is to be able to re-use any kind of primary content and further enrichments thereof. The central aspect of re-using such content is the development and applications of reference standards that reflect the best state of the art in the corresponding domains. In this respect, our team is particularly attentive to the existing standardisation background when both producing language resources or devloping NLP components. Furthermore, our specific leading roles in the domain of standardisation in both the Parthenos and EHRI EU projects as well as in related initiatives (TEI consortium, ISO committee TC 37, DARIAH lexical working group) has allowed to make progress along the following lines:

- Contributing to the revision of the ISO 24613 standard (Lexical Markup Framework) in the form of a multipart standard covering, for the time being, the core model (ISO 24613-1), machine readable dictionaries (ISO 24613-2), etymology (ISO 24613-3) and a TEI based serialisation (ISO 24613-4). Several members of the team have been particularly active as experts in the definition of the first two parts, which are now at publication and DIS stage respectively [0] and are co-editors of parts 3 and 4;
- Proposal for a reference TEI subset for integrating dictionary content: in the context of the DARIAH working group on lexical resources, a first release of the *TEI Lex 0*[0] was issued in September 2018 integrating the continuous work of the group over the the 2016-2018 period and already taken up by the infrastructure project ELEXIS [0] as its reference back-office format. This work is also the basis for the output format of Grobid-Dictionaries [71];
- Finalisation of the ISO proposal on reference annotation (ISO 24617-9): the team has been leading the work on the definition of the Reference Annotation Framework (RAF) [0] which is now at DIS ballot stage and already implemented in several concrete annotation projects[19], [43]. The standard is feature complete from a linguistic point of view (from simple co-reference to complex bridging anaphora phenomena) and compliant with the TEI stand-off annotation module [59] from the point of view of its implementation [66];

[0] See the ISO/TC 37/SC 4 work current work program under https://www.iso.org/committee/297592/x/catalogue/p/0/u/1/w/0/d/0
[0] https://github.com/DARIAH-ERIC/lexicalresources
[0] https://elex.is
[0] https://www.iso.org/standard/69658.html

- Large-scale implementation of international standard for the documentation of the Mixtepec-Mixtec language (see section 6.11);
- Proposing a customisation architecture for the EAD international standard: EAD (Encoding Archival Description [0]) is used worldwide in cultural heritage institution to describe and exchange collection level information. In the context of the EHRI project, where we had to design a mechanism for integrating heterogeneous implementations of EAD-based data, we used the TEI ODD specification language to re-design and subset the international EAD specification to precisely provide interoperability conditions within the project[14];
- Release of the SSK (Standardisation Survival Kit), a generic environment for describing standards-based digital humanities research scenarios: the SSK is an online platform for describing research scenarios developed within the Parthenos project[40] and now deployed as a service hosted by the French national Huma-Num infrastructure [0]. The SSK has been developed as a completely open project [0], where the scenarios are themselves described as TEI-based representations[51], [35], [50].

6.5. Entity-fishing: a generic named entity recognition and disambiguation for digital humanities projects

Participants: Marie Puren, Charles Riondet, Laurent Romary, Luca Foppiano, Tanti Kristanti.

Since several years (starting at the beginning of the EU Cendari project in 2012 [75]) we have been working on the provision of a generic named-entity recognition and disambiguation module (NERD) called *entity-fishing*[18] as a stable on-line service. The work we have achieved demonstrates the possible delivery of sustainable technical services as part of the development of research infrastructures for the humanities in Europe. In particular, our results contribute not only to DARIAH, the European digital research infrastructure for the arts and humanities, but also to OPERAS, the European research infrastructure for the development of open scholarly communication in the social sciences and humanities. Deployed as part of the French national infrastructure Huma-Num, the service provides an efficient state-of-the-art implementation coupled with standardised interfaces allowing easy deployment in a variety of potential digital humanities contexts. In 2018, we have specifically integrated *entity-fishing* within the H2020 HIRMEOS project where several open access publishers have used the service in their collections of published monographs as a means to enhance retrieval and access.

To this end, we have set up a common layer of services on top of several existing e-publishing platforms for Open Access monographs. The *entity extraction* task was deployed over a corpus of monographs provided by the HIRMEOS partners, with the following coverage:

- 4000 books in English and French from Open Edition Books
- 2000 titles in English and German from OAPEN
- 162 books in English from Ubiquity Press
- 765 books (606 in German, 159 in English) from the University of Göttingen

The introduction of *entity-fishing* has undergone different levels of integration. The majority of the participating publishers provided additional features in their user interface, using the data generated by *entity-fishing*, for example, as search facets for persons and locations to help users narrow down their searches and obtain more precise results.

entity-fishing has been developed in Java and it has been designed for fast processing on text and PDF, with relatively limited memory and to offer relatively close to state-of-the-art accuracy (as compared with other NERD systems). The accuracy f-score for disambiguation is currently between 76.5 and 89.1 on standard datasets (ACE2004, AIDA-CONLL-testb, AQUAINT, MSNBC) (Table 1) [74].

[0] https://en.wikipedia.org/wiki/Encoded_Archival_Description
[0] http://ssk.huma-num.fr
[0] https://github.com/ParthenosWP4/SSK

Table 1. Accuracy measures

	ACE 2004	AIDA CONLL-testb	AQUAINT	MSNBC
Priors	83.1	66.1	80.3	71.1
entity-fishing	83.5	76.5	**89.1**	86.7
Wikifier	83.4	77.7	86.2	85.1
DoSeR	**90.7**	78.4	84.2	91.1
AIDA	81.5	77.4	53.2	78.2
Spotlight	71.3	59.3	71.3	51.1
Babelfy	56.1	59.2	65.2	60.7
WAT	80.0	84.3	76.8	77.7
(Ganea & Hofmann, 2017)	88.5	**92.2**	88.5	**93.7**

The objective, however, is to provide a generic service that has a steady throughput of 500-1000 words per second or one PDF page of a scientific article in 1-2 seconds on a medium range (4CPU, 3Gb Ram) Linux server.

From the point of view of the technical deployment itself, we have provided all the necessary components of a sustainable service:

- release and publish *entity-fishing* as open source software [0];
- deploy the service in the DARIAH infrastructure through HUMA-NUM [0];
- produce evaluation data and metrics for content validation.

6.6. From GROBID to GROBID-Dictionaries

Participants: Luca Foppiano, Mohamed Khemakhem, Laurent Romary, Pedro Ortiz Suárez, Alba Marina Malaga Sabogal.

GROBID is an open source software suite initiated in 2007 by Patrice Lopez with the purpose of extracting metadata automatically from scholarly papers available in PDF. Over the years, it has developed into a rich information extraction environment, and deployed in many Inria projects, but also national and international services, such as HAL (front-end meta-data extraction from uploaded scholarly publications). It is a central piece for our information extraction activities and we have been particularly active in 2018 in the following domains:

- General contributions to GROBID [0]:
  - Major refactoring and design improvements
  - fixes, tests, documentation and update of the pdf2xml fork for Windows
  - added and improved several models in collaboration with CERN (e.g. for the recognition of arXiv identifier)
  - Further tests on the specific case of bibliographic documents[32]
- Contribution to GROBID-Dictionaries [0]: the lexical GROBID extension has been implemented and tested on modern and multilingual dictionaries[23]. In the context of several collaborative activities, GROBID-Dictionaries has been applied on several documentary sources:
  - Early editions of the The Petit Larrousse Illustré in the context of the Nénufar project[45], [29]

[0] http://github.com/kermitt2/nerd
[0] http://nerd.huma-num.fr/nerd/
[0] https://github.com/kermitt2/grobid
[0] https://github.com/MedKhem/grobid-dictionaries

- Further experiments on etymological dictionaries from the Berlin Brandenburg Academy of Sciences
- Experiments on entry-based documents such as manuscript catalogues (with University of Neuchâtel)[16] and the French address Directory Bottin from the end of the XIXth Century[22]

These various experiments have been accompanied by an intense training and hand-on activity in the context in particular of the French research network CAHIERS (Huma-Num consortium), the Lexical Data Master Class and a series of workshop organised in South Africa under the auspices of a national linguistic documentation program. Finally, 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

The experience gained in the develoment and application of GROBID-Dictionaries has been the basis for the recently accepted ANR BASNUM project which aims at automatically structuring and enriching of the Dictionnaire universel (DU) by Antoine Furetière, in its 1701 edition rewritten by Basnage de Beauval and the doctoral work of Pedro Ortiz.

6.7. Resources, models and tools for coreference resolution

Participants: Loïc Grobol, Éric Villemonte de La Clergerie.

This year we performed many experiments, some of them detailed in [28], targeting end-to-end coreference systems for spontaneous oral French. More precisely, for several mention-pair coreference detection models, we tried to assess their sensibility to various features of coreference chains and their viability for end-to-end systems, compared to the more recent antecedent scoring models.

Also, one of our objective being to assess the usefulness of syntactic features for coreference detection, we enriched the coreference annotations of the ANCOR corpus with both automatically produced dependency syntax annotations and improved speech transcription. All these annotation were wrapped in a TEI-compliant XML format as described in [20] (see also 6.4).

Finally, we have been working on neural architectures for coreference detection, building upon some recent state of the art techniques. They are based on embeddings for general text span and we try to make them more scalable through efficient uses of the local context but also more tunable to different document types and language variation. The base idea is to complete pre-training by training on related lower-level tasks such as entity-mention detection.

6.8. Computational history through information extraction from archive texts

Participants: Éric Villemonte de La Clergerie, Marie Puren, Charles Riondet, Alix Chagué, Marie-Laurence Bonhomme.

From two different DH projects emerged some interesting research questions related to the extraction of information from archival documents, in particular the management of the diversity of document types and structures and on the contrary the acquisition of detailed information from a regular visual structure.

In the context of the ANR TIME-US, whose goal is to reconstruct the "time-budgets" of textile workers in France (18th - early 20th centuries), we worked on the creation of a digitization workflow to acquire structured textual data from a wide range of printed and handwritten materials: professional court records (like *Prud'Hommes*), Police reports on strikes or early sociological studies such as the *Monographies de Le Play* This workflow has been presented at the ADHO DH conference in Mexico (see the presentation here: [34].The set up of this workflow is a prerequisite for further experiments and processing to extract information that can be exploited by historians, such as the relation between working tasks, the time spent by workers to perform them and the price they are paid for this time.

Another project was intiated in collaboration with the EPHE and the French National Archives, in the framework of the convention signed between Inria and the Ministry of Culture. This project is called LECTAUREP (for *LECTure AUtomatique de REPertoires*, and is aimed at extracting the information recorded in the registries of Parisian notaries, held by the National Archives. This project it at the intersection of NLP and Computer Vision because one of the main objectives is to extract information from the physical layout of the documents, presented as tables. Another issue is to be able to recognize with accuracy an important diversity of handwritten scripts. The final goal of LECTAUREP is to give access to researchers the information contained in these records, in particular the name of the persons involved in cases recorded by notaries, their addresses and the nature of the case (wills, powers of attorney, wedding contracts, etc.). An initial report has been produced (see [39]), and the project will continue in 2019 with the release of the extracted information (named entities, geolocation, typology, etc) into a structured database.

6.9. Discovering correlations between parser features and neurological observations

Participants: Éric Villemonte de La Clergerie, Murielle Fabre, Pauline Brunet.

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 fMRI data resulting from listening of the novel Le Petit Prince.

In 2018, Pauline Brunet, during her Master thesis, has worked on developing the infrastructure (scripts and formats) for the integration of the features, and the use of these features for computing correlations with fMRI data. A first set of features has been identified and collected from the novel and from its processing by ALMAnaCH tools (namely FRMG as an instance of a symbolic TAG-based parser and Dyalog-SR, as an instance of an hybrid feature-based neural-based dependency parser). A first dataset of fMRI scan was received to assess the infrastructure and get some preliminary results.

The work is now being continued with the arrival as a post-doc of Murielle Fabre (November 2018). With the expected arrival of the second half of the scans, she will explore more features, use her expertise to interpret the correlations, and guide the choice of new features to be tested. Since her arrival, she has in particular focused on Multi-Word Expressions (MWEs), in particular to be comparable with results published on the English side of the project. We have also identified several kinds of parsing architectures to test, in relation with various complexity parameters: (1) LSTM (two layers), (2) RNNG (with a partile filter), (3) Dyalog-SR et (4) FRMG (TAG).

In order to be in phase (and comparable) with our US partners, we have started to assemble 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 an 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 represents a balanced counterpart to the American COCA corpus. [0]

Both Éric de La Clergerie and Murielle Fabre attended the annual meeting of the CRCNS network (Berkeley, June 2018).

6.10. Evaluating the quality of text simplification

Participants: Louis Martin, Benoît Sagot, Éric Villemonte de La Clergerie.

[0] https://corpus.byu.edu/coca/

In 2018, our collaboration on text simplification with the Facebook Artificial Intelligence Research lab in Paris (in particular with Antoine Bordes) has started in practice. It has taken the form of a CIFRE PhD. In this context, in 2018, we dedicated important efforts to the problem of the evaluation of text simplification (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 judgment of grammaticality and meaning preservation, whereas simplicity is best evaluated by basic length-based metrics [24].

6.11. Advances in descriptive, computational and historical linguistics

Participants: Benoît Sagot, Laurent Romary, Jack Bowers, Rebecca Blevins.

ALMAnaCH members have resumed their work in descriptive, computational and historical linguistics, an important way to ensure that NLP models and tools are robust to the diversity of world languages, as well as a way to apply NLP models and tools for contributing to research in linguistics. Three of 2018 advances in this regard are the following:

- In the context of the doctoral work of Jack Bowers, a first release of a global documentation of the Mixtepec-Mixtec language has been released which covers, multilayered annotated spoken and written resources as well as a reference lexical resource covering both basic word descriptions and elaborate semantic and etymological (word formation) content [13];
- Work on language description and computational morphology for Romansh Tuatschin in collaboration with Géraldine Walther (Universität Zürich) was pursued, following the work published in 2017 [99]. A new interest in the quantitative, corpus-based study of code switching in this language has emerged in collaboration with Claudia Cathomas (Universität Zürich), leading to preliminary results to be published in 2019;
- We resumed our work in (classical) etymology in collaboration with Romain Garnier (Université de Limoges, Institut Universitaire de France), with a focus not only on (Ancient) Greek and its substrates, but also, more specifically, on Anatolian languages that could be amongst said substrates. In particular, we proposed that Lydian could be the source language for a number of Greek words lacking a good etymology in the literature [31], which motivated Rebecca Blevins's internship on the development of a lexicon of the Lydian language. We also published new etymological results at the (Proto-)Indo-European level [37].

6.12. Language resources and NLP tools for Medieval French

Participants: Éric Villemonte de La Clergerie, Mathilde Regnault, Benoît Sagot.

The main objectives of the ANR project "Profiterole" are 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 like such a corpus (because of diachronic, dialectal, geographic, stylistic and genre-based variation, among other types of linguistic variation). To this end, we have continued previous experiments in morpho-syntactic 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 syntactic annotation (i.e. 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. The development of the OFrLex lexicon has started within the Alexina framework, following the Lefff lexicon for contemporary French [5]. It already allowed for preliminary experiments. On the other hand, we conducted parsing experiments with neural models (DyALog's SRNN models). Note that members of the ALMAnaCH team participated in the CoNLL dependency parsing Shared Task 2018, which included an Old French dataset (see section 6.2).

COML Team

6. New Results

6.1. Speech and Audio Processing from the Raw Waveform

State-of-the-art speech technology systems (e.g., ASR and TTS) rely on fixed, hand-crafted features such as mel-filterbanks to preprocess the waveform before the training pipeline. This is at odds with recent work in machine vision where hand-crafter features (SIFT, etc) have been succesfully replaced by features derived from raw pixels trained jointly with a downstream task. In this line of work, we explored how a similar approach could be undertaken for audio and speech processing.

- In [24], we train a bank of complex filters that operates at the level of the raw speech signal and feeds into a convolutional neural network for phone recognition. These time-domain filterbanks (TD-filterbanks) are initialized as an approximation of MFSC, and then fine-tuned jointly with the remaining convolutional network. We perform phone recognition experiments on TIMIT and show that for several architectures, models trained on TD-filterbanks consistently out-perform their counterparts trained on comparable MFSC. We get our best performance by learning all front-end steps, from pre-emphasis up to averaging. Finally, we observe that the filters at convergence have an asymmetric impulse response while preserving some analyticity.
- In [25], we study end-to-end systems trained directly from the raw waveform, building on two alternatives for trainable replacements of mel-filterbanks that use a convolutional architecture. The first one is inspired by gammatone filterbanks [4], [9], and the second one by the scattering transform [24]. We propose two modifications to these architectures and systematically compare them to mel-filterbanks, on the Wall Street Journal dataset. The first modification is the addition of an instance normalization layer, which greatly improves on the gammatone-based trainable filterbanks and speeds up the training of the scattering-based filterbanks. The second one relates to the low-pass filter used in these approaches. These modifications consistently improve performances for both approaches, and remove the need for a careful initialization in scattering-based trainable filterbanks. In particular, we show a consistent improvement in word error rate of the trainable filterbanks relatively to comparable mel-filterbanks. It is the first time end-to-end models trained from the raw signal significantly outperform mel-filterbanks on a large vocabulary task under clean recording conditions.
- Recent progress in deep learning for audio synthesis opens the way to models that directly produce the waveform, shifting away from the traditional paradigm of relying on vocoders or MIDI synthesizers. Despite their successes, current state-of-the-art neural audio synthesizers such as WaveNet and SampleRNN [12], [8] suffer from prohibitive training and inference times because they are based on autoregressive models that generate audio samples one at a time at a rate of 16kHz. In this work [26], we study the more computationally efficient alternative of generating the waveform frame-by-frame with large strides. We present SING, a lightweight neural audio synthesizer for the original task of generating musical notes given desired instrument, pitch and velocity. Our model is trained end-to-end to generate notes from nearly 1000 instruments with a single decoder, thanks to a new loss function that minimizes the distances between the log spectrograms of the generated and target waveforms. On the generalization task of synthesizing notes for pairs of pitch and instrument not seen during training, SING produces audio with significantly improved perceptual quality compared to a state-of-the-art autoencoder based on WaveNet [4] as measured by a Mean Opinion Score (MOS), and is about 32 times faster for training and 2,500 times faster for inference.

6.2. Development of cognitively inspired algorithms

Speech and language processing in humans infants and adults is particularly efficient. We use these as sources of inspiration for developing novel machine learning and speech technology algorithms. In this area, our results are as follows:

- In [22], we summarize the accomplishments of a multi-disciplinary 6-weeks workshop organized by E. Dupoux (PI) at Carnegy Mellon Univerrsity (Pittsburgh), funded through the Jelinek Memorial Summer Workshop Program of Johns Hopkins University. The workshop explored the computational and scientific issues surrounding the discovery of linguistic units (subwords and words) in a language without orthography. We studied the replacement of orthographic transcriptions by images and/or translated text in a well-resourced language to help unsupervised discovery from raw speech.
- Developing speech technologies for low-resource languages has become a very active research field over the last decade. Among others, Bayesian models have shown some promising results on artificial examples but still lack of in situ experiments. In [20] , we apply state-of-the-art Bayesian models to unsupervised Acoustic Unit Discovery (AUD) in a real low-resource language scenario. We also show that Bayesian models can naturally integrate information from other resourceful languages by means of informative prior leading to more consistent discovered units. Finally, discovered acoustic units are used, either as the 1-best sequence or as a lattice, to perform word segmentation. Word segmentation results show that this Bayesian approach clearly outperforms a Segmental-DTW baseline on the same corpus.
- Fixed-length embeddings of words are very useful for a variety of tasks in speech and language processing. In [19], we systematically explore two methods of computing fixed-length embeddings for variable-length sequences. We evaluate their susceptibility to phonetic and speaker-specific variability on English, a high resource language, and Xitsonga, a low resource language, using two evaluation metrics: ABX word discrimination and ROC-AUC on same-different phoneme n-grams. We show that a simple downsampling method supplemented with length information can be competitive with the variable-length input feature representation on both evaluations. Recurrent autoencoders trained without supervision can yield even better results at the expense of increased computational complexity.
- Recent studies have investigated siamese network architectures for learning invariant speech representations using same-different side information at the word level. In [21], we investigate systematically an often ignored component of siamese networks: the sampling procedure (how pairs of same vs. different tokens are selected). We show that sampling strategies taking into account Zipf's Law, the distribution of speakers and the proportions of same and different pairs of words significantly impact the performance of the network. In particular, we show that word frequency compression improves learning across a large range of variations in number of training pairs. This effect does not apply to the same extent to the fully unsupervised setting, where the pairs of same-different words are obtained by spoken term discovery. We apply these results to pairs of words discovered using an unsupervised algorithm and show an improvement on state-of-the-art in unsupervised representation learning using siamese networks.
- Unsupervised spoken term discovery is the task of finding recurrent acoustic patterns in speech without any annotations. Current approaches consists of two steps: (1) discovering similar patterns in speech, and (2) partitioning those pairs of acoustic tokens using graph clustering methods. In, [23] we propose a new approach for the first step. Previous systems used various approximation algorithms to make the search tractable on large amounts of data. Our approach is based on an optimized k-nearest neighbours (KNN) search coupled with a fixed word embedding algorithm. The results show that the KNN algorithm is robust across languages, consistently outperforms the DTW-based baseline, and is competitive with current state-of-the-art spoken term discovery systems.

6.3. 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).

- Two PhDs were defended this year. In [14], Adriana Guavara Rukoz presented a computational model of the perception of non-native speech contrasts based on standard ASR pipelines is presented. An adaptation of the model is proposed to account for forced-choice classification psycholinguistic experiments and directly reproduced classical results. The general finding is that, suprisingly, the acoustic model part of a phone recognizer is sufficient to account for experimental data, even those apparently related to phonotactic properties of the native language. The 'language model' part does not improve the correlation with adult data (if anything, it degrades it). Yet the match between model and human is not perfect, and is was hypothetized that improvement in the acoustic model could help. In [13], Julia Maria Carbajal presented a study of the effect of multilingual exposure on language acquisition. She used a computational model of language separation based on i-vectors to reproduce some of the known effects of phonological distance on language discrimination in infants.
- In [16], we investigate whether infant-directed speech (IDS) facilitates lexical learning when compared to adult-directed speech (ADS). To study this, we compare the distinctiveness of the lexicon at two levels, acoustic and phonological, using a large database of spontaneous speech in Japanese. At the acoustic level we show that, as has been documented before for phonemes, the realizations of words are more variable and less discriminable in IDS. At the phonological level, we find that despite a slight increase in the number of phonological neighbors, the IDS lexicon contains more distinctive words (such as onomatopeias). Combining the acoustic and phonological metrics together in a global discrimination score, the two effects cancel each other out and the IDS lexicon winds up being as discriminable as its ADS counterpart. We discuss the implication of these findings for the view of IDS as hyperspeech, i.e., a register whose purpose is to facilitate language acquisition.
- Existing theories of cross-linguistic phonetic category perception agree that listeners perceive foreign sounds by mapping them onto their native phonetic categories. Yet, none of the available theories specify a way to compute this mapping. As a result, they cannot provide systematic quantitative predictions and remain mainly descriptive. Here [17], Automatic Speech Recognition (ASR) systems are used to provide a fully specified mapping between foreign and native sounds. This is shown to provide a quantitative model that can account for several empirically attested effects in human cross-linguistic phonetic category perception.
- Spectacular progress in the information processing sciences (machine learning, wearable sensors) promises to revolutionize the study of cognitive development. In [15], we analyse the conditions under which 'reverse engineering' language development, i.e., building an effective system that mimics infant's achievements, can contribute to our scientific understanding of early language development. We argue that, on the computational side, it is important to move from toy problems to the full complexity of the learning situation, and take as input as faithful reconstructions of the sensory signals available to infants as possible. On the data side, accessible but privacy-preserving repositories of home data have to be setup. On the psycholinguistic side, specific tests have to be constructed to benchmark humans and machines at different linguistic levels. We discuss the feasibility of this approach and present an overview of current results.

6.4. 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.

- In [18], we present BabyCloud, a platform for capturing, storing and analyzing daylong audio recordings and photographs of children's linguistic environments, for the purpose of studying infant's cognitive and linguistic development and interactions with the environment. The proposed platform connects two communities of users: families and academics, with strong innovation potential for each type of users. For families, the platform offers a novel functionality: the ability for parents to follow the development of their child on a daily basis through language and cognitive

metrics (growth curves in number of words, verbal complexity, social skills, etc). For academic research, the platform provides a novel means for studying language and cognitive development at an unprecedented scale and level of detail. They will submit algorithms to the secure server which will only output anonymized aggregate statistics. Ultimately, BabyCloud aims at creating an ecosystem of third parties (public and private research labs...) gravitating around developmental data, entirely controlled by the party whose data originate from, i.e. families.

MULTISPEECH Project-Team

7. New Results

7.1. Explicit Modeling of Speech Production and Perception

Participants: Anne Bonneau, Vincent Colotte, Denis Jouvet, Yves Laprie, Slim Ouni, Agnès Piquard-Kipffer, Théo Biasutto-Lervat, Sara Dahmani, Ioannis Douros, Valérian Girard, Thomas Girod, Anastasiia Tsukanova.

7.1.1. Articulatory modeling

7.1.1.1. Articulatory models and synthesis

Since articulatory modeling, i.e. representing the geometry of the vocal tract with a small number of parameters, is a key issue in articulatory synthesis the improvement of the articulatory models remains an important objective. This year we put emphasis on thin articulators as the epiglottis and velum. Indeed, the delineation of those contours often leads to erroneous transverse dimensions (too thin or too thick contours) which generates some artificial swelling deformations. Before the determination of the deformation modes, the central lines of the velum and epiglottis are extracted in the images use to build the model. The deformation modes thus only concern the central line, which prevents artificial swelling factors to emerge from the factor analysis. A reconstruction algorithm has been developed to obtain the contour from the central line.

7.1.1.2. Acoustic simulations

One of the issues in articulatory synthesis is to assess the impact of the geometric simplifications that are made on the vocal tract so as to enable faster acoustic simulation and to decrease the number of parameters required to approximate the vocal tract shape. The other issue concerns the impact of the plane wave assumption. The idea consists of comparing the signal or spectrum synthesized via numerical acoustic simulation against the one measured on a real human subject. However, this requires that both geometric and corresponding acoustic data are available at the same time. This can be achieved with MRI data when the acquisition duration is sufficiently short to allow the speaker to phonate the sound during the whole acquisition. The MRI acquisition protocol has thus been optimized on the new Siemens Prisma MRI machine of Nancy hospital so as to reduce the acquisition time to 7 seconds, which makes it possible for the subject to produce a sound throughout the acquisition. The acoustic simulation was achieved by using the Matlab K-wave package, either from the entire 3D volume extracted from the MRI data, or from the 2D shape extracted from the mid-sagittal plane. Several simplifications have been carried out (with or without the epiglottis, with or without the velum...) so as to assess their acoustic impacts. These simulations only concern vowels because these sounds can be sustained by subjects and the MRI machine noise does not change the position of formant frequencies dramatically. This work has been carried out in cooperation with IADI laboratory.

7.1.1.3. Exploitation of dynamic articulatory data

The size of the dynamic database (recorded last year in the Max Planck Institute for Biophysical Chemistry in Göttingen), in the form of MRI films of the mid-sagittal plane acquired at 55 Hz, is about 200.000 images. Even if the long term objective is to exploit the whole database, efforts were dedicated to manual delineation of contours in some films with the idea of using those data to train a machine learning technique. Several students were trained, and in total more than 1000 images have been delineated. The corresponding films have been exploited to achieve articulatory copy synthesis by improved acoustic simulations developed last year.

7.1.1.4. Acoustic-to-articulatory inversion

Deriving articulatory dynamics from the acoustic speech signal is a recurrent topic in our team. This year, we have investigated whether it is possible to predict articulatory dynamics from phonetic information without having the acoustic speech signal. The input data may be considered as not sufficiently rich acoustically, as there is probably no explicit coarticulation information, but we expect that the phonetic sequence provides compact yet rich knowledge. We have experimented a recurrent neural network architecture, where we have trained the model with an electromagnetic articulography (EMA) corpus, and have obtained good performances similar to the state-of-the-art articulatory inversion from line spectral frequencies (LSF) features [21].

7.1.2. Expressive acoustic and visual synthesis

7.1.2.1. Expressive speech

A comparison between emotional speech and neutral speech has been carried on using a small corpus of acted speech. The analysis was focused on the way pronunciations and prosodic parameters are modified in emotional speech, compared to neutral style [20].

Experiments with deep learning-based approaches for expressive speech synthesis are described in 7.2.4.2 .

7.1.2.2. Expressive audiovisual synthesis and lipsync

This year, we have acquired audiovisual 3D corpus (using the optitrack system, using 8 cameras) for a set of emotions acted by a professional actress. We recorded 6 basic emotions: joy, fear, disgust, sadness, anger, surprise; in addition to neutral speech. The corpus contains 5000 utterances (2000 utterances for the neutral speech and 500 utterances per emotion). The visual and acoustic data have been processed, segmented and labeled spatially and temporally. An important aspect of the work was to study the evaluation of the quality of the animation of a 3D talking head where the animation is generated from the acquired 3D data. For this purpose, we studied the relevance of root mean square error (RMSE) measure which is classically used to evaluate the error of the prediction. Our preliminary results confirmed that RMSE can be irrelevant in our field, as we may not reach critical articulatory target, and we still obtain very low RMSE. Thus the audiovisual intelligibility of the system would be low. To improve the results, we have worked on improving the 3D model controls using better key-shapes and reduced redundant and confusing blendshapes.

The processed neutral-speech data have been used to train a deep neural network to predict from speech and linguistic information the trajectories of the animation controls of the talking head, which is the core of the lipsync system. We have also used this expressive-speech data to train a DNN-based TTS to synthesize expressive audiovisual speech from text. Currently, we are performing extensive testing and validation of the results.

7.1.3. Categorization of sounds and prosody for native and non-native speech

7.1.3.1. Visual clues in speech perception and production

We continue our research focused on the importance of multimodal speech combining oral and visual clues. We investigated identification and production of morpho-syntactic skills in ten deaf children (severe with cochlear implant using French cued-speech LPC - *Langue française Parlée Complétée*) and ten age-matched children with typical development. Our goal was to examine the production of morpho-syntactic structures in auditory channel versus audiovisual speech. Five conditions were observed: audiovisual conditions with a 3D avatar speaking or coding oral language with LPC versus a human speaker with or without LPC and auditory channel. We used the 3D avatar coding set up in the ADT Handicom project. Statistical analysis and interpretation of results is ongoing.

7.1.3.2. Reading and related skills norms

We set-up standardized norms on the development of reading and related skills in French: EVALEC Primaire software (in collaboration with the LPC - *Laboratoire de Psychologie Cognitive*, UMR 7290, *Aix-Marseille Université*). This year, LPC collected new data at the end of grade 5 (about 100 children) and added them to those previously collected at the end of grades 1–4, about 100 children for each level [69]. EVALEC primaire software includes five tests focused on written word processing, recording both accuracy scores and processing time (time latency and vocal response duration for the reading aloud tests). EVALEC primaire software also includes tests of phonemic and syllabic awareness, phonological short-term memory, and rapid naming. These data would allow researchers and speech therapists to assess the reading and reading-related skills of dyslexic children as compared to average readers.

7.1.3.3. Analysis of non-native pronunciations

We have examined the effects of L1/L2 interferences at the segmental level, and of the lack of fluency at the sentence level, on the realizations of French final fricatives by German learners. Due to L1/L2 interference, German speakers tend to devoice French final fricatives. A well-known effect of the lack of L2 mastering is the decrease of the speech articulation rate, which lengthens the average duration of segments. In order to better apprehend the impact of categorization and fluency, we selected four series of consonants from the IFCASL corpus, i.e. voiced and unvoiced fricatives uttered by French native and German non-native speakers. The realizations of French unvoiced consonants uttered by German speakers are essentially dependent on fluency, whereas the realizations of voiced consonants by the same speakers are dependent on both fluency and categorization. We evaluated a set of acoustic cues related to the voicing distinction -including consonant duration and periodicity-, and submitted the data to a hierarchical clustering analysis. Results, discussed as a function of speaker's level and prosodic boundaries, confirmed the mutual importance of fluency and segmental categorization on non-native realizations [22].

Within the METAL project, work is on-going for integrating speech processing technology in an application to help learning foreign language and for experimenting it with middle and high school students learning German. This includes tutoring aspects using 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. Preliminary experiments have shown the poor quality of speech signals recorded from groups of students in classrooms.

7.2. Statistical Modeling of Speech

Participants: Vincent Colotte, Antoine Deleforge, Dominique Fohr, Irène Illina, Denis Jouvet, Odile Mella, Romain Serizel, Emmanuel Vincent, Md Sahidullah, Guillaume Carbajal, Ken Déguernel, Diego Di Carlo, Adrien Dufraux, Raphaël Duroselle, Mathieu Fontaine, Nicolas Furnon, Amal Houidhek, Ajinkya Kulkarni, Nathan Libermann, Aditya Nugraha, Manuel Pariente, Laureline Perotin, Sunit Sivasankaran, Nicolas Turpault, Imene Zangar.

7.2.1. Source localization and separation

Emmanuel Vincent has co-edited a 500-page book on audio source separation and speech enhancement, which provides a unifying view of array processing, matrix factorization, deep learning and other methods, with application to speech and music [64]. We also contributed to five chapters in that book [60], [62], [59], [54], [61] and three chapters in another book [53], [56], [55].

7.2.1.1. Source localization

In multichannel scenarios, source localization and source separation are tightly related tasks. We introduced the real and imaginary parts of the acoustic intensity vector in each time-frequency bin as suitable input features for deep learning based speaker localization [37]. We analyzed the inner working of the neural network using a methodology called layerwise relevance propagation, which points the time-frequency bins on which the network relies to output a given location [68]. We defined a new task called text-informed speaker localization, which consists of localizing the speaker uttering a known word or sentence such as the wake-up word of a hands-free voice command system in a situation when other speakers are overlapping. We proposed a method to address this task, where a phonetic alignment is obtained, converted into an estimated time-frequency mask, and fed to a convolutional neural network together with interchannel phase difference features in order to localize the desired speaker [43]. We published a new dataset using a microphone array embedded in an unmanned aerial vehicle in [45], organized an international sound source localization challenge associated to this dataset and participated to the 2018 LOCATA sound source localization challenge. We published a book chapter on audio-motor integration, showing an application to sound source localization with robots [52].

7.2.1.2. Room acoustics modeling

In a given room, each possible position of the microphones and the sources corresponds to different room transfer functions. The goal of room acoustic modeling is to model the manifold formed by these transfer functions. Past studies have focused on learning a supervised mapping between the relative transfer function and the source location for localization purposes. We introduced the reverse task consisting of learning a mapping between the source location and the corresponding relative transfer function, which may be used as a prior on the relative transfer function for source separation purposes. We proposed a semi-supervised algorithm to learn this mapping in a situation when the location of each relative transfer function measurement is not precisely known [48]. We also started investigating the estimation and modeling of early acoustic echoes. In [39] we showed how their knowledge could improve performance of sound source separation algorithms. In [36] we proposed a new method to estimate them blindly from multichannel recordings with much higher precision than conventional blind channel identification methods.

7.2.1.3. Deep neural models for source separation and echo suppression

We pursued our research on the use of deep learning for multichannel source separation [5]. We introduced a method that exploits knowledge of the source locations in order to estimate multichannel Wiener filters for two or more sources [38]. We explored several variants of the multichannel Wiener filter, which turned out to result in better speech recognition performance on the CHiME-3 dataset [17]. We also used deep neural networks for reducing the residual nonlinear echo after linear acoustic echo cancellation [23] and started extending this approach to joint reverberation, echo, and noise reduction. Finally, we recently started exploring the case where the microphones composing a multichannel array are not distributed according to a predefined geometry and do not have a common sampling clock.

7.2.1.4. Alpha-stable modeling of audio signals

This year, our work on heavy tails distribution has witnessed a significant advance with the development of a multichannel model that is able to account for the inter-channel delays and time difference of arrivals in an alpha-stable framework, hence benefiting from the inherent robustness of such distributions. This work has been submitted to the IEEE transactions on Signal Processing by Mathieu Fontaine and is still under review. Its main applications are: i/ the separation of multichannel sources, for which we have demonstrated a superiority with respect to the multichannel Wiener filter in the oracle setting, and ii/ localizations of heavy tailed sources, where we worked on the theoretical foundations

7.2.1.5. Beyond Gaussian modeling of audio signals

The team has investigated a number of alternative probabilistic models to the symmetric local complex Gaussian (LCG) model for audio source separation. An important limit of LCG is that most signals of interest such as speech or music do not exhibit Gaussian distributions but heavier-tailed ones due to their important dynamic. In [31] we proposed a new sound source separation algorithm using heavy-tailed alpha stable priors for source signals. Experiments showed that it outperformed baseline Gaussian-based methods on under-determined speech or music mixtures. Another limitation of LCG is that it implies a zero-mean complex prior on source signals. This induces a bias towards low signal energies, in particular in under-determined settings. With the development of accurate magnitude spectrogram models for audio signals using deep neural networks, it becomes desirable to use probabilistic models enforcing stronger magnitude priors and better accounting for phases. In [35], we presented the BEADS (Bayesian Expansion Approximating the Donut Shape) model. The prior considered is a mixture of isotropic Gaussians regularly placed on a zero-centered complex circle. We showed it outperformed LCG on an informed source separation task.

7.2.1.6. Interference reduction

Our work on interference reduction focused this year in scaling our previous work to full-length recording. This has been achieved thanks to a new method we proposed, which estimates the interference reduction parameters based on random projections of the full length recordings [25]. This technique scales linearly with the duration of the recording, making it usable in real-world use-cases.

The book chapter we published on audio-motor integration, shows an application to ego-noise reduction for robots [52]. In the context of robotics, ego-noise refers to the acoustic noise produced in a robot's microphones by its own movement.

7.2.2. Acoustic modeling

7.2.2.1. Robust acoustic modeling

Achieving robust speech recognition in reverberant, noisy, multi-source conditions requires both speech enhancement and separation and 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 [1]. We oversaw the collection of a new dataset sponsored by Google, which considers a 'dinner party' scenario. Twenty parties of four people, who know each other well, were recorded in their own homes using 2 binaural in-ear microphones per participant and 6 distant Kinects, for a total duration of about 50 h. We organized the CHiME-5 Challenge based on these data [19]. We also participated in the collection of two French datasets for ambient assisted living applications as part of the voiceHome [11] and VOCADOM [51] projects.

7.2.2.2. Ambient sounds

We are constantly surrounded by sounds and we rely heavily on these sounds to obtain important information about what is happening around us. Our team has been involved in the community on ambient sound recognition for the past few years. In collaboration with Johannes Kepler University (Austria) and Carnegie Mellon University (USA), we co-organized a task on large-scale sound event detection as part of the Detection and Classification of Acoustic Scenes and Events (DCASE) 2018 Challenge [40]. It focused on the problem of learning from audio segments that are either weakly labeled or not labeled, targeting domestic applications. In this context, we work on semi-supervised sampling strategies to create triplets (a triplet is composed of the current sample, a so-called positive sample from the same class as the current sample and a negative sample from a different class) and studied their application to train triplet networks for audio tagging.

7.2.2.3. Speech/Non-speech detection

Automatic Speech Recognition (ASR) of multimedia content such as videos or multi-genre broadcasting requires a correct extraction of speech segments. We explored the efficiency of deep neural models for speech/non-speech segmentation. We used a bidirectional LSTM model to obtain speech/non-speech probabilities and a decision module (4-state automaton with safety margins). Compared to a Gaussian Mixture Model (GMM) based speech/non-speech segmenter, the results achieved on the MGB British Challenge data, show a reduction of the ASR word error rate (23.7% versus 29.4%). We have also trained models for the Arabic and French languages.

7.2.2.4. Transcription systems

Within the AMIS project, speech recognition systems have been developed for the transcription of videos in French, English and Arabic. They have been integrated with other components (such as translation and summarization) to allow for the summarization of videos in a target language [44], [29], [28].

7.2.2.5. Speaker recognition

Speaker recognition is the task of recognizing a person from its voice. The performances of speaker recognition systems severely degrade due to several practical challenges such as the limited amount of speech data, real-world noises and spoofing. We explored the efficiency of DNN-based distance metric learning methods for speaker recognition in short duration conditions. Currently, we are developing a neural network architecture that gives phone-invariant speaker embeddings for robust speaker recognition. We also participated in the NIST speaker recognition evaluation 2018 as a part of the I4U consortium. The speaker recognition technology is vulnerable to spoofing attacks where mimicked voice, synthetic speech, or playback voice is used to get illegitimate access. We are investigating whether technology-assisted speaker selection can help in improving mimicry attack [67]. In [24], we proposed an enhanced baseline system for replay spoofing detection with ASVspoof 2017 dataset. In [26], we demonstrated that playback speech enhanced with DNN-based speech enhancement method can severely degrade the speaker recognition and countermeasure performance as compared to the conventional replay attacks with voice samples from covert recording. We also proposed

a common feature and back-end fusion scheme for the integration of spoofing countermeasures and speaker recognition [47]. Currently, we are co-organizing the third edition of automatic speaker verification spoofing challenge (ASVspoof 2019) where our newly developed cost function [32] will be adopted for the performance assessment of integrated systems. In the context of multimodal authentication with the voice as a modality, we investigated the optimization of speech features for audio-visual synchrony detection [41].

7.2.2.6. Language identification

With respect to language identification, the current research activity focuses on lightly supervised or unsupervised domain adaptation. The goal is to adapt a language identification system optimized for a given transmission channel to a new transmission channel.

7.2.3. Language modeling

7.2.3.1. Out-of-vocabulary proper name retrieval

Despite recent progress in developing Large Vocabulary Continuous Speech Recognition Systems (LVCSR), these systems suffer from Out-Of-Vocabulary words (OOV). In many cases, the OOV words are Proper Nouns (PNs). The correct recognition of PNs is essential for broadcast news, audio indexing, etc. We addressed the problem of OOV PN retrieval in the context of broadcast news LVCSR. We focused on dynamic (document dependent) extension of LVCSR lexicon. To retrieve relevant OOV PNs, we proposed to use a very large multipurpose text corpus: Wikipedia. This corpus contains a huge number of PNs. These PNs are grouped in semantically similar classes using word embedding. We used a two-step approach: first, we selected OOV PN pertinent classes with a multi-class Deep Neural Network (DNN). Secondly, we ranked the OOVs of the selected classes. The experiments on French broadcast news show that a bi-directional Gated Recurrent Unit model outperforms other studied models. Speech recognition experiments demonstrate the effectiveness of the proposed methodology [18].

7.2.3.2. Updating speech recognition vocabularies

Within the AMIS project, the update of speech recognition vocabularies has been investigated using web data collected over a time period similar to that of the collected videos, for three languages: French, English and Arabic. Results have been analyzed globally, and also with respect to names only. This analysis has shown the poor coverage of the names by the baseline lexicons, and has also demonstrated the benefits of the updated lexicons, both in term of WER reduction and OOV rate reduction [14].

7.2.3.3. Music language modeling

Similarly to speech, music involves several levels of information, from the acoustic signal up to cognitive quantities such as composer style or key, through mid-level quantities such as a musical score or a sequence of chords. The dependencies between mid-level and lower- or higher-level information can be represented through acoustic models and language models, respectively. Ken Déguernel defended his PhD on automatic music improvisation [10] and he proposed a polyphonic music improvisation approach that takes the structure of the musical piece at multiple time scales into account [12]. We also explored the ability of a conventional recurrent neural network with moving history to account for long-term dependencies in music melodies, and compared it with two new architectures with growing or parallel history [50].

7.2.3.4. Automatic detection of hate speech

Nowadays, Twitter, LinkedIn, Facebook and YouTube are very popular for communicating ideas, beliefs, feelings or any other form of information. At the same time, the dark side of these new technologies has led to an increase in hate speech or racism. Our work seeks to study hate speech in user-generated contents in France, which thus requires French resources. We plan to design a hate speech corpus and a lexicon in French; whereas such hate speech lexicons exist for other languages, no such tool can be found in French. We began, on English data, to develop a new methodology to automatically detect hate speech, based on machine learning and Neural Networks. Human detection of this material is unfeasible since the contents to be analyzed are huge. Current machine learning methods use only certain task specific features to model hate speech. We propose to develop an innovative approach to combine these pieces of information into a multi-feature approach so that the weaknesses of the individual features are compensated by the strengths of other features. We began a collaboration with the CREM laboratory in Metz and Saarland University.

7.2.4. Speech generation

7.2.4.1. Arabic speech synthesis

Work on Arabic speech synthesis was carried out within a CMCU PHC project with ENIT (École Nationale d'Ingénieurs de Tunis, Tunisia), using HMM and NN based approaches applied to Modern Standard Arabic language. Speech synthesis systems rely on a description of speech segments corresponding to phonemes, with a large set of features that represent phonetic, phonologic, linguistic and contextual aspects. When applied to Modern Standard Arabic, two specific phenomena have to be taken in account: vowel quantity and gemination. This year, we studied thoroughly the modeling of these phenomena. Results of objective and subjective evaluations showed that the use of a deep neural architecture in speech synthesis (more specifically in predicting the speech parameters) enhanced the accuracy of acoustic modelling so that the quality of generated speech is better than that of HMM-based speech synthesis [30], [13].

Deep neural network (DNN) approaches have been further investigated for the modeling of phoneme duration. According to the specific phenomena of the Arabic language, we proposed a class-specific modeling of the phoneme durations. An objective evaluation showed that the proposed approach leads to a more accurate modeling of the phoneme duration (compared to HMM-based or MERLIN DNN-based approaches) [49].

7.2.4.2. Expressive acoustic synthesis

Expressive speech synthesis using parametric approaches is constrained by the style of the speech corpus used. We carried out a preliminary study on developing expressive speech synthesis for a new speaker voice without requiring a specific recording of expressive speech by this new speaker. For that, we focused on deep neural network based layer adaptation for investigating the transfer the expressive characteristics to a new speaker for which only neutral speech data is available. Such transfer learning mechanism should accelerate the efforts towards exploiting existing expressive speech corpora. However, there is a trade-off between the knowledge transfer of expressivity characteristics and the retaining of the speaker's identity in the synthesized speech.

7.3. Uncertainty Estimation and Exploitation in Speech Processing

Participants: Irène Illina, Denis Jouvet, Emmanuel Vincent, Yassine Boudi, Baldwin Dumortier, Elodie Gauthier, Mathieu Hu, Lou Lee, Anne-Laure Piat-Marchand.

7.3.1. Uncertainty and acoustic modeling

7.3.1.1. Uncertainty in noise-robust speech and speaker recognition

In many real-world conditions, the target speech signal overlaps with noise and some distortion remains after speech enhancement. The framework of uncertainty decoding assumes that this distortion has a Gaussian distribution and seeks to estimate its covariance matrix and propagate it through the acoustic model for robust ASR [4]. We introduced new Gaussian mixture model-derived (GMMD) uncertainty features for robust DNN-based acoustic model training and decoding, which are computed as the difference between the closed-form GMM log-likelihoods obtained with vs. without uncertainty. We concatenated the GMMD features with conventional acoustic features and showed that they improve ASR performance on both the CHiME-2 and CHiME-3 datasets [15].

7.3.1.2. Uncertainty in other applications

Besides the above application, we finalized our exploration of uncertainty modeling for wind turbine control. Baldwin Dumortier defended his PhD thesis on this topic [9].

7.3.2. Uncertainty and phonetic segmentation

In the METAL project, experiments are planned to investigate further the use of speech technologies for foreign language learning in middle and high schools. Besides adapting acoustic models to teenager voices, current work investigates the reliability of speech technologies for analyzing student pronunciations, and for detecting miss-pronunciations. Also, besides making the pronunciation diagnostics more reliable, the aim is to elaborate robust strategies that will make it possible to handle sets of unreliable individual results, and still be able to provide a relevant feedback on recurrent miss-pronunciations.

7.3.3. Uncertainty and prosody

The analysis of prosodic correlates of discourse particles has continued. Some additional data has been annotated. The automatic word and phonetic segmentation of the discourse particles has been manually checked and corrected when necessary. Once more, this has shown that automatic segmentation is not perfect, especially on spontaneous speech recording in real conditions. For each discourse particle, prosodic characteristics of occurrences of each pragmatic function (conclusive, introductive, etc.) were automatically extracted. For each discourse particle and each pragmatic function, the most frequent F0 patterns were retained as the representative forms. Results show that a pragmatic function, common to several discourse particles, gives rise to a uniform prosodic marking [34].

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.

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. Besides the principle and its application to image denoising [105], we demonstrated the potential of the approach to speedup linear inverse problems [104], and a comprehensive journal paper was published in 2016 [107]. Pioneering identifiability results have been obtained in the Ph.D. thesis of Luc Le Magoarou [108].

We further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms [106], and studied their approximation error [109]. In a journal paper published this year [16] we empirically show that $\mathcal{O}(n \log n)$ approximate implementations of Graph Fourier Transforms are possible for certain families of graphs. This opens the way to substantial accelerations for Fourier Transforms on large graphs.

The FAμST software library (see Section 6) was first released as Matlab code primarily for reproducibility of the experiments of [107]. A C++ version is being developed to provide 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 have 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. A special case of the proposed structure is the widespread separable dictionary, named SuKro, which was evaluated experimentally last year on an image denoising application [81]. The general approach, coined HOSUKRO for High Order Sum of Kronecker products, has been shown this year to reduce empirically the sample complexity of dictionary learning, as well as theoretical complexity of both the learning and the sparse coding operations [27].

Combining faster matrix-vector products with screening techniques. We combined accelerated matrix-vector multiplications offered by FAμST / HOSUKRO matrix approximations with dynamic screening [57], 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 [80]. We then adapted an existing screening rule to this new framework and define a general procedure to leverage the advantages of both strategies. This year we completed a comprehensive preprint submitted for publication in a journal [49] 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 [28].

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 [76], 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 [77], 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. This year, a condensed version of these results has been prepared which is now accepted for publication [14].

7.1.3. Algorithmic exploration of large-scale Compressive Learning via Sketching

Participants: Rémi Gribonval, Antoine Chatalic, Antoine Deleforge.

Main collaborations: Patrick Perez (Technicolor R&I France, Rennes), Anthony Bourrier (formerly Technicolor R&I France, Rennes; then GIPSA-Lab, Grenoble), Antoine Liutkus (ZENITH Inria project-team, Montpellier), Nicolas Keriven (ENS Paris), Nicolas Tremblay (GIPSA-Lab, Grenoble), 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 Mixture Estimation. When fitting a probability model to voluminous data, memory and computational time can become prohibitive. We proposed during the Ph.D. thesis of Anthony Bourrier [58], [61], [59], [60] to fit a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical generalized moments of the underlying probability distribution. Deriving a reconstruction algorithm by analogy with compressive sensing, we experimentally showed that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. The Ph.D. thesis of Nicolas Keriven [97] consolidated extensions to non-isotropic Gaussians, with a new algorithm called CL-OMP [96] and large-scale experiments demonstrating its potential for speaker verification

[95]. A journal paper was published this year [15], with an associated toolbox for reproducible research (see SketchMLBox, Section 6).

Sketching for Compressive Clustering and beyond. In 2016 we started a new endeavor to extend the sketched learning approach beyond Gaussian Mixture Estimation.

First, we showed empirically that sketching can be adapted to compress a training collection while allowing large-scale *clustering*. The approach, called "Compressive K-means", uses CL-OMP at the learning stage [98]. This year, we showed that in the high-dimensional setting one can substantially speedup both the sketching stage and the learning stage by replacing Gaussian random matrices with fast random linear transforms in the sketching procedure [23].

An alternative to CL-OMP for cluster recovery from a sketch 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 [62]. During his first year of Ph.D., Antoine Chatalic visited the group of Phil Schiter to further investigate this topic, and a journal paper is in preparation.

We also demonstrated that sketching can be used in blind source localization and separation, by learning mixtures of alpha-stable distributions [32], see details in Section 7.5.3 .

Finally, sketching provides a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum. A conference paper establishing theoretical privacy guarantees (with the *differential privacy* framework) and exploring the utility / privacy tradeoffs of Compressive K-means has been submitted for publication.

7.1.4. Theoretical results on Low-dimensional Representations, Inverse problems, and Dimension Reduction

Participants: Rémi Gribonval, Clément Elvira.

Main collaboration: Mike Davies (University of Edinburgh, UK), Gilles Puy (Technicolor R&I France, Rennes), Yann Traonmilin (Institut Mathématique de Bordeaux), 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)

Inverse problems and compressive sensing in Hilbert spaces.

Many inverse problems in signal processing deal with the robust estimation of unknown data from underdetermined linear observations. Low dimensional models, when combined with appropriate regularizers, have been shown to be efficient at performing this task. Sparse models with the ℓ^1-norm or low-rank models with the nuclear norm are examples of such successful combinations. Stable recovery guarantees in these settings have been established using a common tool adapted to each case: the notion of restricted isometry property (RIP). We published a comprehensive paper [20] establishing generic RIP-based guarantees for the stable recovery of cones (positively homogeneous model sets) with arbitrary regularizers. We also described a generic technique to construct linear maps from a Hilbert space to $\mathbb{R}^m$ that satisfy the RIP [121]. These results have been surveyed in a book chapter published this year [46]. In the context of nonlinear inverse problems, we showed that the notion of RIP is still relevant with proper adaptation [42].

Optimal convex regularizers for linear inverse problems. The ℓ^1-norm is a good convex regularization for the recovery of sparse vectors from under-determined linear measurements. No other convex regularization seems to surpass its sparse recovery performance. We explored possible explanations for this phenomenon by defining several notions of " best " (convex) regularization in the context of general low-dimensional recovery and showed that indeed the ℓ^1-norm is an optimal convex sparse regularization within this framework [43]. A journal paper is in preparation with extensions concerning nuclear norm regularization for low-rank matrix recovery and further structured low-dimensional models.

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 has been completed is under revision for a journal [88].

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 [40], 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 [40]. A journal paper is in preparation describing a full class of kernels for which such an analysis holds, in particular for higher dimensional parameters.

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 [89]. 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. [51] Somewhere in the proofs appears an apparently new characterization of proximity operators of (nonconvex) penalties as subdifferentials of convex potentials [50].

7.1.5. Algorithmic Exploration of Sparse Representations for Neurofeedback

Participant: Rémi Gribonval.

Claire Cury, Pierre Maurel & Christian Barillot (VISAGES 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 9.1.1.1), in collaboration with the VISAGES team, we validated a technique to estimate brain neuronal activity by combining EEG and fMRI modalities in a joint framework exploiting sparsity [118]. This year we 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. A conference paper has been submitted.

7.1.6. Sparse Representations as Features for Heart Sounds Classification

Participant: Nancy Bertin.

Main collaborations: Roilhi Frajo Ibarra Hernandez, Miguel Alonso Arevalo (CICESE, Ensenada, Mexico)

A heart sound signal or phonocardiogram (PCG) is the most simple, economical and non-invasive tool to detect cardiovascular diseases (CVD), the main cause of death worldwide. During the visit of Roilhi Ibarra, we proposed a pipeline and benchmark for binary heart sounds classification, based on his previous work on a sparse decomposition of the PCG [91]. We improved the feature extraction architecture, by combining features derived from the Gabor atoms selected at the sparse representation stage, with Linear Predictive Coding coefficients of the residual. We compared seven classifiers with two different approaches in presence of multiple hearts beats in the recordings: feature averaging (proposed by us) and cycle averaging (state-of-the-art). The feature sets were also tested when using an oversampling method for balancing. The benchmark identified systems showing a satisfying performance in terms of accuracy, sensitivity, and Matthews correlation coefficient, with best results achieved when using the new feature averaging strategy together with oversampling. This work was accepted for publication in an international conference [30].

7.1.7. An Alternative Framework for Sparse Representations: Sparse "Analysis" Models

Participants: Rémi Gribonval, Nancy Bertin, Clément Gaultier.

Main collaborations: Srdan Kitic (Orange, Rennes), Laurent Albera and Siouar Bensaid (LTSI, Univ. Rennes)

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of *few* columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome. Building on our pioneering work on the cosparse model [87], [117] [8], successful applications of this approach to sound source localization, brain imaging and audio restoration have been developed in the team during the last years [99], [101], [100], [55]. Along this line, two main achievements were obtained this year. First, and following the publication in 2016 of a journal paper embedding in a unified fashion our results in source localization [5], a book chapter gathering our contributions in physics-driven cosparse regularization, including new results and algorithms demonstrating the versatility, robustness and computational efficiency of our methods in realistic, large scale scenarios in acoustics and EEG signal processing, was published this year [45]. Second, we continued extending the cosparse framework on audio restoration problems [85], [84], [82], especially improvements on our released real-time declipping algorithm (A-SPADE - see Section 6.2) and extension to multichannel data [29].

7.2. Activities on Waveform Design for Telecommunications

Peak to Average Power Ratio (PAPR), Orthogonal Frequency Division Multiplexing (OFDM), Generalized Waveforms for Multi Carrier (GWMC), Adaptive Wavelet Packet Modulation (AWPM)

7.2.1. Multi-carrier waveform systems with optimum PAPR

Participant: Rémi Gribonval.

Main collaboration: Marwa Chafii, Jacques Palicot, Carlos Bader (SCEE team, CentraleSupelec, Rennes)

In the context of the TEPN (Towards Energy Proportional Networks) Comin Labs project (see Section 9.1.1.2), in collaboration with the SCEE team at Supelec (thesis of Marwa Chafii [63], defended in October 2016 and co-supervised by R. Gribonval, and awarded with the GDR ISIS/GRETSI/EEA thesis prize, see Section 5.1.1), we investigated a problem related to dictionary design: the characterization of waveforms with low Peak to Average Power Ratio (PAPR) for wireless communications. This is motivated by the importance of a low PAPR for energy-efficient transmission systems.

A first stage of the work consisted in characterizing the statistical distribution of the PAPR for a general family of multi-carrier systems, [67], [65], [66]. We characterized waveforms with optimum PAPR [68], [64] as well as the tradeoffs between PAPR and Power Spectral Density properties of a wavelet modulation scheme [70]. Our design of new adaptive multi-carrier waveform systems able to cope with frequency-selective channels while minimizing PAPR gave rise to a patent [69] and was published this year [22], [13].

7.3. 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 [11] was dedicated to learning techniques for the design of new efficient methods for large-scale image search and indexing. A first step was to propose techniques for approximate nearest neighbor search exploiting quantized sparse representations in learned dictionaries [92]. The thesis then explored structured binary codes, computed through supervised learning with convolutional neural networks [93]. This year, we integrated these two components in a unified end-to-end learning framework where both the representation and the index are learnt [31]. These results have led to a patent application.

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 out- put 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 has been submitted for publication.

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 journal paper is in preparation.

7.4. Emerging activities on Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

7.4.1. Locally-Linear Inverse Regression

Participant: Antoine Deleforge.

Main collaborations: Florence Forbes (MISTIS Inria project-team, Grenoble), Emeline Perthame (HUB team, Institut Pasteur, Paris), Vincent Drouard, Radu Horaud, Sileye Ba and Georgios Evangelidis (PERCEPTION Inria project-team, Grenoble)

A general problem in machine learning and statistics is that of *high- to low-dimensional mapping*. In other words, given two spaces $\mathbb{R}^D$ and $\mathbb{R}^L$ with $D \gg L$, how to find a relation between these two spaces such that given a new observation vector $y \in \mathbb{R}^D$ its associated vector $x \in \mathbb{R}^L$ can be estimated? In *regression*, a set of training pairs $\{(y_n, x_n)\}_{n=1}^N$ is used to learn the relation. In *dimensionality reduction*, only vectors $\{y_n\}_{n=1}^N$ are observed, and an intrinsic low-dimensional representation $\{x_n\}_{n=1}^N$ is sought. In [73], we introduced a probabilistic framework unifying both tasks referred to as *Gaussian Locally Linear Mapping* (GLLiM). The key idea is to learn an easier other-way-around locally-linear relationship from x to y using a joint Gaussian Mixture model on x and y. This mapping is then easily reversed via Bayes' inversion. This framework was notably applied to hyperspectral imaging of Mars [71], head pose estimation in images [79], sound source separation and localization [72], and virtually-supervised acoustic space learning (see Section 7.6.1). This year, in [19], we introduced the *Student Locally Linear Mapping* (SLLiM) framework. The use of heavy-tailed Student's t-distributions instead of Gaussian ones leads to more robustness and better regression performance on several datasets.

7.4.2. 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 [54]), we proposed over the last three years a series of algorithms leveraging the competitive cosparse approach, which offers a very appealing trade-off between reconstruction performance and computational time [100], [102] [6]. The work on cosparse audio declipping which was awarded the Conexant best paper award at the LVA/ICA 2015 conference [102] resulted in a software release in 2016. In 2017, this work was extended towards advanced (co)sparse decompositions, including several forms of structured sparsityand towards their application to the denoising task.In particular, we investigated the incorporation of the so-called "social" structure constraint [103] into problems regularized by a cosparse prior [84], [85], and exhibited a common framework allowing to tackle both denoising and declipping in a unified fashion [82].

In 2018, a new algorithm for joint declipping of multichannel audio was derived and published [29]. Extensive experimental benchmarks were conducted, questioning the previous state-of-the-art habits in degradation levels (usually moderate to inaudible) and evaluation (small datasets, SNR-based performance criteria) and setting up new standards for the task (large and diverse datasets, severe saturation, perceptual quality evaluation) as well as guidelines for the choice of the best variant (sparse or cosparse, with or without structural time-frequency constraints...) depending on the data and operational conditions. These new results will be included in an ongoing journal paper, to be submitted in 2019.

7.5. 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 [122] 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 [10].

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 voiceHome and INVATE projects (see Sections 7.5.1 , 7.5.2); more exploratory work towards new approaches diverging away from local Gaussian modeling (Section 7.5.3) ; formulating and addressing a larger class of problems related to localization and separation, in the contexts of robotics (Section 7.5.4) and virtual reality (Section 7.5.2). Eventually, one of these new tracks, audio scene analysis with machine learning, evolved beyond the "localization and separation" paradigm, and is the subject of a new axis of research presented in Section 7.6 .

7.5.1. 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, we were part of the voiceHome project (2015-2017, see Section 9.1.4), an industrial collaboration 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). We benchmarked, improved and optimized existing localization and separation tools to the particular context of this application, worked on a better interface between source localization and source separations steps and on optimal initialization scenarios, and reduced the latency and computational burden of the previously available tools, highlighting operating conditions were real-time processing is achievable. Automatic selection of the best microphones subset in an array was investigated. A journal publication including new data (extending the voiceHome Corpus, see Section 6.1), baseline tools and results, submitted to a special issue of Speech Communication, was published this year [12].

Accomplished progress and levers of improvements identified thanks to this project resulted in the granting of an Inria ADT (Action de Développement Technologique), which started in September 2017, for a new development phase of the FASST software (see Section 6.5). 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), and to industrial transfer (see Section 8.1.1).

7.5.2. 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 9.1.2) started in November 2016, we investigated a new application of source separation to sound re-spatialization from Higher Order Ambisonics (HOA) signals [86], 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 led to a conference paper submission in 2018. We continued extending our methods to hybrid acquisition scenarios, where the separation of HOA signals can be informed by complementary close-up microphonic signals. Future work will include subjective evaluation of the developed workflows.

7.5.3. Beyond the Local Complex Gaussian Model

Participant: Antoine Deleforge.

Main collaboration: Nicolas Keriven (ENS Paris), Antoine Liutkus (ZENITH Inria project-team, Montpellier)

The team has also recently investigated a number of alternative probabilistic models to the local complex Gaussian (LCG) model for audio source separation. An important limit of LCG is that most signals of interest such as speech or music do not exhibit Gaussian distributions but heavier-tailed ones due to their important dynamic [110]. We provided a theoretical analysis of some limitations of the classical LCG-based multichannel Wiener filter in [21]. In [32] we proposed a new sound source separation algorithm using heavy-tailed alpha stable priors for source signals. Experiments showed that it outperformed baseline Gaussian-based methods on under-determined speech or music mixtures. Another limitation of LCG is that it implies a zero-mean complex prior on source signals. This induces a bias towards low signal energies, in particular in under-determined settings. With the development of accurate magnitude spectrogram models for audio signals such as nonnegative matrix factorization [120][9] or more recently deep neural networks [119], it becomes desirable to use probabilistic models enforcing strong magnitude priors. In [75], we explored deterministic magnitude models. An approximate and tractable probabilistic version of this referred to as BEADS (Bayesian Expansion Approximating the Donut Shape) was presented this year [33]. The source prior considered is a mixture of isotropic Gaussians regularly placed on a zero-centered complex circle.

7.5.4. Applications to Robot Audition

Participants: Nancy Bertin, Antoine Deleforge.

Main collaborations: Aly Magassouba, Pol Mordel and François Chaumette (LAGADIC Inria project-team, Rennes), Alexander Schmidt and Walter Kellermann (University of Erlangen-Nuremberg, Germany)

Implicit Localization through Audio-based Control. In robotics, the use of aural perception has received recently a growing interest but still remains marginal in comparison to vision. Yet audio sensing is a valid alternative or complement to vision in robotics, for instance in homing tasks. Most existing works are based on the relative localization of a defined system with respect to a sound source, and the control scheme is generally designed separately from the localization system. In contrast, the approach that we investigated in the context of Aly Magassouba's Ph.D. (defended in December 2016) focused on a sensor-based control approach. Results obtained in the previous years [116], [114], [115] were encompassed and extended in two journal papers published this year [17], [18]. In particular, we obtained new results on the use of interaural level difference as the only input feature of the servo, a counter-intuitive result outside the robotic context. We also showed the robustness, low-complexity and independence to Head Related Transfer Function (HRTF) of the approach on humanoid robots.

Sound Source Localization with a Drone. Flying robots or drones have undergone a massive development in recent years. Already broadly commercialized for entertainment purpose, they also underpin a number of exciting future applications such as mail delivery, smart agriculture, archaeology or search and rescue. An important technological challenge for these platforms is that of localizing sound sources in order to better analyse and understand their environment. For instance, how to localize a person crying for help in the context of a natural disaster? This challenge raises a number of difficult scientific questions. How to efficiently embed

a microphone array on a drone? How to deal with the heavy ego-noise produced by the drone's motors? How to deal with moving microphones and distant sources? Victor Miguet and Martin Strauss tackled part of these challenges during their masters' internships. A light 3D-printed structure was designed to embed a USB sound card and a cubic 8-microphone array under a Mikrokopter drone that can carry up to 800 g of payload in flights. Noiseless speech and on-flights ego-noise datasets were recorded. The data were precisely annotated with the target source's position, the state of each drone's propellers and the drone's position and velocity. Baseline methods including multichannel Wiener filtering, GCC-PHAT and MUSIC were implemented in both C++ and Matlab and were tested on the dataset. Up to 5° speech localization accuracy in both azimuth and elevation was achieved under heavy-noise conditions (-5 dB signal-to-noise-ratio). The dataset was made publicly available at dregon.inria.fr and was presented together with the results in [37].

7.6. Towards comprehensive audio scene analysis

Source localization and separation, machine learning, room geometry, room properties, multichannel audio classification

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.

7.6.1. Virtually-Supervised Auditory Scene Analysis

Participants: Antoine Deleforge, Nancy Bertin, Diego Di Carlo, Clément Gaultier, Rémi Gribonval.

Main collaborations: Ivan Dokmanic (University of Illinois at Urbana-Champaign, Coordinated Science Lab, USA), Saurabh Kataria (IIT Kanpur, India).

Classical audio signal processing methods strongly rely on a good knowledge of the *geometry* of the audio scene, *i.e.*, what are the positions of the sources, the sensors, and how does the sound propagate between them. The most commonly used *free field* geometrical model assumes that the microphone configuration is perfectly known and that the sound propagates as a single plane wave from each source to each sensor (no reflection or interference). This model is not valid in realistic scenarios where the environment may be unknown, cluttered, dynamic, and include multiple sources, diffuse sounds, noise and/or reverberations. Such difficulties critical hinders sound source separation and localization tasks.

Recently, two directions for advanced audio geometry estimation have emerged and were investigated in our team. The first one is physics-driven [45]. This approach implicitly solves the wave propagation equation in a given simplified yet realistic environment assuming that only few sound sources are present, in order to recover the positions of sources, sensors, or even some of the wall absorption properties. However, it relies on partial knowledge of the system (e.g. room dimensions), limiting their real-world applicability so far. The second direction is data-driven. It uses machine learning to bypass the use of a physical model by directly estimating a mapping from acoustic features to source positions, using training data obtained in a real room [72], [74]. These methods can in principle work in arbitrarily complex environments, but they require carefully annotated training datasets. Since obtaining such data is time consuming, the methods are usually working well for one specific room and setup, and are hard to generalize in practice.

We proposed a new paradigm that aims at making the best of physics-driven and data-driven approaches, referred to as *virtually acoustic space travelling* (VAST) [83], [94]. The idea is to use a physics-based room-acoustic simulator to generate arbitrary large datasets of room-impulse responses corresponding to various acoustic environments, adapted to the physical audio system at hand. We demonstrated that mappings learned from these data could potentially be used to not only estimate the 3D position of a source but also some acoustical properties of the room [94]. We also showed that a virtually-learned mapping could robustly localize sound sources from real-world binaural input, which is the first result of this kind in audio source localization [83]. The VAST datasets and approaches made the bed of several new works in 2018, including real-world

source localization on a wider range of settings (LOCATA test data on various microphone arrays) and echo estimation (see below).

7.6.2. Room Properties: Estimating or Learning Early Echoes

Participants: Antoine Deleforge, Nancy Bertin, Diego Di Carlo.

Main collaborations: 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 [35] 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 [78]. In 2018, two different approaches to this problem were explored.

In [34] we proposed an analytical method for early echoes estimation. This method builds on the framework of finite-rate-of-innovation sampling. The approach operates directly in the parameter-space of echo locations and weights, and enables near-exact blind and off-grid echo retrieval from discrete-time measurements. It is shown to 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 and convex relaxations, extensions to sparse-spectrum signals and noisy measurements, and applications to dereverberation and audio-based room shape reconstruction.

As a concurrent approach exploration, the PhD thesis of Diego Di Carlo aims at applying the VAST framework to the blind estimation of acoustic echoes, or other room properties (such as reverberation time, acoustic properties at the boundaries, etc.) This 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 submitted to an international conference. Future work will consider 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.

7.6.3. 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 months 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 a baseline algorithm for multi-channel joint localization and classification of audio events, currently under development.

7.7. 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 [56], in particular the System & Contrast model [1].

7.7.1. Tensor-based Representation of Sectional Units in Music

Participant: Frédéric Bimbot.

This work was primarily carried out by Corentin Guichaoua, former PhD student with Panama, now with IRMA (CNRS UMR 7501, Strasbourg).

Following Kolmogorov's complexity paradigm, modeling the structure of a musical segment can be addressed by searching for the compression program that describes as economically as possible the musical content of that segment, within a given family of compression schemes.

In this general framework, packing the musical data in a tensor-derived representation enables to decompose the structure into two components : (i) the shape of the tensor which characterizes the way in which the musical elements are arranged in an n-dimensional space and (ii) the values within the tensor which reflect the content of the musical segment and minimize the complexity of the relations between its elements.

This approach has been studied in the context of Corentin Guichaoua's PhD [90] where a novel method for the inference of musical structure based on the optimisation of a tensorial compression criterion has been designed and experimented.

This tensorial compression criterion exploits the redundancy resulting from repetitions, similarities, progressions and analogies within musical segments in order to pack musical information observed at different timescales in a single n-dimensional object.

The proposed method has been introduced from a formal point of view and has been related to the System & Constrast Model [1] as a extension of that model to hypercubic tensorial patterns and their deformations.

From the experimental point of view, the method has been tested on 100 pop music pieces (RWC Pop database) represented as chord sequences, with the goal to locate the boundaries of structural segments on the basis of chord grouping by minimizing the complexity criterion. The results have clearly established the relevance of the tensorial compression approach, with F-measure scores reaching 70 % on that task [41]

7.7.2. 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 [111] defined as the shortest displacement of notes, in semitones, between a pair of chords. Other chord representations and relations are possible, as studied in [113] 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 [112] 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 will be included in Corentin Louboutin's PhD, which is planned to be defended early 2019.

7.7.3. 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.

These experiments have been conducted on the Lakh MIDI dataset and more particularly on a subset of 8308 monophonic 16-bar melodic segments. The obtained results indicate a non-uniform distribution of modeling capabilities prediction of recurrent networks, suggesting the utility of non-ergodic models for the generation of melodic segments [38].

Ongoing work is extending these findings to the design of specific NN architectures, sto account for this non-invariance of information across musical segments.

7.7.4. 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.

SEMAGRAMME Project-Team

6. New Results

6.1. Syntax-Semantics Interface

Participants: Maxime Amblard, William Babonnaud, Philippe de Groote, Bruno Guillaume, Guy Perrier, Sylvain Pogodalla, Valentin Richard.

6.1.1. Abstract Categorial Grammars

Although Abstract Categorial Grammars have well established formal properties that make them suitable for language modeling, some missing features hinder their practical use. For instance, in order to have a compact description of grammatical properties such as number agreement between the subject and the verb of a sentence, a very common approach is to have syntactic descriptions augmented with feature value matrices. Having such a mechanism in Abstract Categorial Grammars requires a lot of attention in order to avoid impacting their computational properties (a previous approach using dependent types showed that, if too general, the problem may become intractable [64]). We have been working on theoretical approaches to this problem from different perspectives: looking for a computationally adequate type extension of the formalisms, and using the composition capabilities of the framework.

We also have been working on a unifying and general framework, provided by a categorical generalization of Abstract Categorial Grammars [50]. The goal is to get a unified approach to several semantic modeling, and to add numerical methods to the formalism.

6.1.2. Syntax-Semantics Interface as Graph Rewriting

In their book (English version: [22] and French version: [21]), Guillaume Bonfante (LORIA, Université de Lorraine), Bruno Guillaume and Guy Perrier devote two chapters to the usage of the Graph Rewriting formalism in the modeling of Syntax-Semantics Interface. Chapter 4 presents two existing semantics formalisms and shows how they can be encoded as graphs: Abstract Meaning Representation (AMR) [33] and Dependency Minimal Recursion Semantics (DMRS) [43], [42]. Chapter 5 described two Graph Rewriting Systems proposed by the authors to build semantics graphs in these two formalisms from syntactic dependencies.

6.1.3. 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. William Babonnaud is currently developing such a calculus.

6.2. Discourse Dynamics

Participants: Maxime Amblard, Timothée Bernard, Clément Beysson, Maria Boritchev, Philippe de Groote, Bruno Guillaume, Pierre Ludmann, Michel Musiol.

6.2.1. Dynamic Logic

We have revisited the type-theoretic dynamic logic introduced in [3]. We have shown how a slightly richer notion of continuation together with an appropriate notion of polarity results in a richer and more powerful framework. In particular, it allows new dynamic connectives and quantifiers to be defined in a systematic way. This work has been presented as an invited talk at the *LACompLing 2018* symposium [11].

6.2.2. Discourse Relations

A text as a whole must exhibit some coherence that makes it more than just a bag of sentences. This coherence hinges on discourse relations (DRs), that express the articulations between the different segments of the text. Typical DRs include relations of *Contrast*, *Consequence* or *Explanation*. The most direct and reliable way to express a DR is to use a discourse connective (e.g., *because*, *instead*, *for example*). These lexical items have specific syntactic, semantic and pragmatic properties, the study of which is the subject of Timothée Bernard's PhD thesis.

Some discourse connectives (typically, adverbial connectives such as *so* or *otherwise*) have only one syntactic argument. It then seems natural to use an anaphora mechanism to retrieve the other argument from the context. This proposal has been formalized in [12] by means of continuation-based type theoretic dynamic logic. In this model, the semantic arguments of a DR are considered to be abstract entities akin to Davidsonian events. This approach raises difficulties when the argument of DR is a negative sentence. Indeed, according to the standard analysis of negation in event semantics, a negative sentence does not introduce any specific event. In order to circumvent this problem, we have developped a logical theory of *negative events* [13], [17], [29].

6.2.3. Dynamic Generalized Quantifiers

Clement Beysson has continued his work on dynamic generalized quantifiers as denotations of the (French) determiners. In this context, he has studied several issues raised by the modeling of plural determiners. In particular, the opposition between distributive and collective interpretations suggests that intrinsically dynamic plural determiners should introduce plural discourse referents that stand for collection of entities. In order to formalize this notion, he has studied several theories of plurality: mereology, plural logic, and second-order logic.

6.2.4. Dialogue Modeling

Maxime Amblard and Maria Boritchev develop a dynamic approach of dialogue modelling. One of the main difference between discourse and dialogue is the interactions between the speakers. To do so, they introduce a formal approach to compositional processing of questions and answers. They address dialogue lexicality issues starting from the formal definitions of so-called Düsseldorf Frame Semantics given in [51]. They introduce a view of dialogues as compositions of negotiation phases that can be studied separately one from another while linked by a common dialogue context (accessible to all participants of a dialogue). They apply Inquisitive Semantics [39] in that context.

Maxime Amblard and Maria Boritchev works on the categorisation of questions and answers and apply some machine learning approches for automatic classification. They present the architecture of the model, especially how to handle these phenomena with logical representations in [14]. Their view is to narrow the problem of identifying incomprehension in dialogue to the one of finding logical incoherences in speech act combinations as the one we found in the SLAM project (ongoing project of the Sémagramme team on interviews with schizophrenics). They also start to build a new corpus - DinG (Discourse in Dialogue) - based on record and transcript plays to the settlers of Catan board game.

Maxime Amblard also started a cooperation with CLASP, especially with Robin Cooper, Ellen Breitholtz and Chris Howes. They work on the synchronisation of the representation of dialogue modelling with the previous proposals and Type-Theoretic-Records (TTR) [41]. They apply the solution on extracts from two corpora where patients with schizophrenia are involved.

6.2.5. Pathological Discourse Modelling

Michel Musiol obtained a part-time delegation in the Semagramme team. This proximity makes possible to set up a more active dialogue 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.

Moreover, the diagnostic tools used today by the professional community (clinical and psychiatric) are of limited expertise for the effective identification of the signs of the pathology for at least two reasons: on the one hand, they are much too imprecise on the side of the recognition of Language Impairment and Thought Disorder (no underlying linguistic and psycholinguistic theories); on the other hand, they do not take into account (either theoretically or technically) the discursive structure within which these disorders are expressed. The objective of this research program is therefore also to anticipate the development of diagnostic tools for the psychiatric and psychological community.

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. The project has progressed on three different operational levels: building new resources, editing a volume (Springer) on the SLAM project in 2019 and improving the representation model.

An agreement is being deployed with the psychiatric hospital of Aix-en-Provence. The on-site staff administered a test protocol to the entire test group of 60 people. Transcripts are in progress, which will provide a significant amount of data to work on for the project. Thanks to the involvement of a medical staff, the recovery of new data appears well advanced. In the same perspective, contacts are being made with the Psychotherapeutic Centre in Nancy.

In addition, Maxime Amblard carried out a one-week international mobility at CLASP thanks to a mobility grant from the French Embassy in Sweden. Discussions were initiated with these colleagues for the development of projects using formal semantic models for the analysis of interaction with schizophrenic patients.

6.3. Common Basic Resources

Participants: Maxime Amblard, Clément Beysson, Philippe de Groote, Bruno Guillaume, Maxime Guillaume, Guy Perrier, Sylvain Pogodalla, Nicolas Lefebvre.

6.3.1. Application of Graph Rewriting to Natural Language Processing

Guillaume Bonfante, Bruno Guillaume and Guy Perrier collected their work on the application of graph rewriting to Natural Language Processing (NLP) in a book written in French [21] and translated to English [22] by the editor. This book shows how graph rewriting can be used as a computational model adapted to NLP. Currently, there is no standard model for graph rewriting and, as such, the authors have conceived one that is specifically adapted to NLP, proposing their own implementation: the GREW system. In addition to the application to Syntax-Semantic Interface mentioned above, the book presents applications in syntactic parsing and in syntactic corpus conversion.

In [5], Guillaume Bonfante and Bruno Guillaume describe some mathematical properties of the Graph Rewriting framework used in GREW. The previous experiments on NLP tasks have shown that Graph Rewriting applications to Natural Language Processing do not require the full computational power of the general Graph Rewriting setting. The most important observation is that all graph vertices in the final structures are in some sense "predictable" from the input data and so, it is possible to consider the framework of Non-size increasing Graph Rewriting. The paper concerns the theoretical aspect of termination with respect to this calculus. It is shown that uniform termination is undecidable and that non-uniform termination is decidable. We define termination techniques based on weight, we prove the termination of weighted rewriting systems and we give complexity bounds on derivation lengths for these rewriting systems.

6.3.2. Building Linguistics Resources with Crowdsourcing

In the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions, Karën Fort (Sorbonne Université), Bruno Guillaume, Matthieu Constant (ATILF, Nancy), Nicolas Lefebvre and Yann-Alan Pilatte (Sorbonne Université) presented the results obtained in crowdsourcing French speakers' intuition concerning multi-word expressions (MWEs) [15]. They developed a slightly gamified crowdsourcing platform, part of which is designed to test users' ability to identify MWEs with no prior training. The participants perform relatively well at the task, with a recall reaching 65% for MWEs that do not behave as function words.

6.3.3. Corpus Annotation

Kim Gerdes (Sorbonne nouvelle, Paris 3), Bruno Guillaume, Sylvain Kahane (Université Paris Nanterre) and Guy Perrier proposed a surface-syntactic annotation scheme called Surface Universal Dependencies (SUD) that is near-isomorphic to the Universal Dependencies (UD) annotation scheme. The SUD scheme follows distributional criteria for defining the dependency tree structure and the naming of the syntactic functions [16]. Rule-based graph transformation grammars allow for a bi-directional transformation of UD into SUD. The back-and-forth transformation can serve as an error-mining tool to assure the intra-language and inter-language coherence of the UD treebanks. The UD corpora are available on gitlab.inria.fr.

Bruno Guillaume and Guy Perrier used the GREW system for the development of the French part of the Universal Dependencies project (UD) [32]. They focused in particular on correcting the annotation of two French corpora, *UD_French-GSD* and *UD_French-Sequoia*. For the correction, they first used the tool Grew-match (based on the pattern matching part of GREW) to detect error patterns, but also the GREW rewriting rule system to transform the annotation from one format to another one [19]. Version 2.3 of the UD corpora was released on 15 November 2018.

6.3.4. FR-Fracas

Maxime Amblard, Clement Beysson, Philippe de Groote, Bruno Guillaume and Sylvain Pogodalla continue their work on the FR-Fracas project. There are two major levels of processing that are significant in the use of a computational semantics framework: semantic composition, for the construction of meanings, and inference, either to exploit those meanings, or to assist the determination of contextually sensitive aspects of meanings. FraCas is an inference test suite for evaluating the inferential competence of different NLP systems and semantic theories. Providing an implementation of the inference level was beyond the scope of FraCaS, but the test suite nevertheless provides an overview of a useful and theory- and system-independent semantic tool [40].

There currently exists a multilingual version of the resource for Farsi, German, Greek, and Mandarin. Sémagramme completed the translation into French of the test suite. All translations were subject to a bidding phase by two project members. Then the cases that were identified as difficult were discussed by all project members. An adjudication step finally ensured the quality of the translation. In order to evaluate the inference mechanism triggered by the translated sentences, a web interface is being developed.

6.3.5. Large Coverage Abstract Categorial Grammars

Maxime Amblard, Maxime Guillaume, and Sylvain Pogodalla have worked on the automatic translation of large coverage Tree-Adjoining grammars into Abstract Categorial Grammars. On the theoretical side, this work hinges on the encoding proposed by Philippe de Groote and Sylvain Pogodalla [69], [63]. On the implementation side, the starting point are TAG grammars generated from meta-grammars by XMG [44], [61]. This generates Abstract Categorial grammars containing about 23 000 entries, and was used as a test bed for the ACGtk toolkit, some parts of which have been rewritten to scale up.

AUCTUS Team

7. New Results

7.1. Posture and motion capture by smart textile

The objective of the work is to design a jacket made of smart textile, without the use of built-in sensors, to determine the posture of the operator.

We propose an innovative solution based on the electrical properties of a stretchable conductive tissue which is used in the manufacture of a smart garment. We use the Electrical Impedance Tomography (EIT) to reconstruct the resistance change of the conductive tissue during tissue extension/deformation caused by human movement. The conductive tissue is placed at strategic points of the jacket (e.g., elbow, shoulder). The model that describes the correlation between the operator's posture/motion and tissue deformation is difficult to obtain analytically. Neural networks are being used to associate the different postures and movements measured by the reference device with the electric field measured in the smart textile. After the learning phase, the neural network is able to predict articular angle with an accuracy of + -5 degrees from tissue extension/deformation only.

Following the successful validation on the first prototype, a request of the patent was drafted and submitted on November 6, 2018 under the number FR1860192 (Smart textile adapted for motion and/ or deformation detection). At the same time, we submitted an experiment project to COERLE. The experiments are planned for next year. This study will allow us to acquire a big database for the learning of artificial neural networks in order to try to propose a unique and stable solution of human posture capture by the smart textile, whatever the anthropometric parameters.

7.2. Appropriate design of kinematic chains

The goal of this research is to develop efficient and reliable tools based on the appropriate design framework using interval analysis that are capable of handling variations and uncertainties for the analysis and synthesis of serial kinematics chains. A primary application for this tool is to accurately model the true workspaces of the redundant human arm by imposing realistic joint constraints that may be obtained experimentally. The appropriate design framework makes it possible to model variations and uncertainties in the kinematics chains to describe families of mechanisms (e.g., sets of arms) and to understand the performance of the family. Through studying a person's usage of their available workspace on a given task, it is theorized that a task expert will make greater use of their available workspace to minimize the risk of fatigue, while a task amateur will confine themselves to a smaller region of their available workspace which will result in expedited fatigue. By understanding the range of motions of a family of task experts, collaborative robotics can be effectively incorporated to assist with the task. A C++ software library, titled the Kinematic Chain Appropriate Design Library, is being developed to efficiently model serial kinematics chains, where the main difficulty is to properly formulate the kinematic equations and incorporate additional constraints so that the problem can be quickly solved using interval analysis methods. The library will be capable of completely solving the forward and inverse kinematics problems, generating certified descriptions of various workspaces, and synthesizing appropriate design solutions.

7.3. Filtering method for human motion analysis

We have developed a series of filters to estimate the states of a dynamic system from a series of incomplete or noisy measurements for the analysis of human motion. They are also used for data fusion or for filtering noisy data from a model, especially for a Kinect and Orbbec sensor. In our case, we first developed an extended Kalman filter [13] that we improved to take into account the singularities of representations of the human kinematic module, the estimation of users' physiological parameters as well as the calibration of measurement systems. In addition, different strategies have been implemented to ensure the real-time operation of the filter, and the addition of joint constraints to improve the accuracy of the results.

In a second step, we implemented an interesting alternative technique for filtering time series. It consists of performing singular spectrum analysis. Due to the multidimensional nature of the type of data we use a specific version of this technique called Multivariate or Multidimensional Singular Spectrum Analysis (MSSA) [19].

This technique is based on a method called *decomposition into main components* which aims to compress the data both on their temporal and physical dimensions. Excellent results have been obtained.

7.4. A software architecture for the analysis of human movement and the prevention of musculoskeletal risk

Robot Operating System (ROS) is used to build the architecture of an in situ system for analyzing the movement of industrial operators. The system, presented in [5], allows us to manage data processing and modules for evaluating and recognizing a human's actions.

The ROS architecture has been chosen to guarantee a certain modularity in our system. More specifically, our objectives are to receive and merge any type of data. We want to set up an agile system that can be used in real time or in remote calculation. We also plan to use our architecture for human-robot interaction

7.5. Hamiltonian Monte Carlo with boundary reflections, and application to polytope volume calculations

In this work [7], we studied HMC with reflections on the boundary of a domain, providing an enhanced alternative to Hit-and-run (HAR) to sample a target distribution in a bounded domain. We make three contributions. First, we provide a convergence bound, paving the way to more precise mixing time analysis. Second, we present a robust implementation based on multi-precision arithmetic – a mandatory ingredient to guarantee exact predicates and robust constructions. Third, we use our HMC random walk to perform polytope volume calculations, using it as an alternative to HAR within the volume algorithm by Cousins and Vempala. The tests, conducted up to dimension 50, show that the HMC RW outperforms HAR.

This work is a collaboration with Frédéric Cazals and Augustin Chevallier from the ABS team at Inria Sophia-Antipolis. Augustin Chevallier visited our team on May 17-18, 2018. Volume calculation is a topic of interest for AUCTUS in light of the volume of configuration spaces.

7.6. Classification of cobotic systems

A new classification of cobotic systems has been proposed [1]. As there are many different ways to classify robots (robotic architecture, size, autonomy, moving ability, adaptability, etc.) and to classify human work or human roles, classifying cobotic systems (the teams formed by a robot and a human operator) is a complex problem. We proposed to focus on information exchanges and interactions among the robot, the human operator and objects of the environment. The graph describing these interactions provides interesting clues to classify cobotic systems. For example, in the surgical robotics and drone domains, the human operator is typically teleoperating (no direct contact with the environment) with constant information exchanges between him and the robot. For that reason, the graph describing these interactions called "scheme of interactions" is very specific. Further on, the description with a scheme of interactions seems particularly appropriate for cobotic systems classification. Several schemes present discriminant features that allow the qualification and naming of the cobotic systems. It is thus possible to identify the symbiotic system, with a constant information exchange and an efficient work sharing (drone), the augmented human case (work with exoskeleton), the subcontracting case, the assistance to effort case and the intelligent assistance case.

7.7. Use of Bayesian networks for situation awareness risks prediction

In all domains involving complex human systems interactions, such as the robotic domain, human errors may have dramatic impacts. These errors are often linked to situation awareness issues. We recently proposed a new method to predict situation awareness errors in training simulations [2]. It is based on Endsley's model and the 8 "situation awareness demons" that she described. The predictions are determined thanks to a Bayesian network and Noisy-Or nodes. A maturity model is introduced to come up with the initialization problem. The NASA behavioral competency model is also used to take individual differences into account.

7.8. Classification of human actions

It is important for the decomposition of human industrial activities to recognize and classify elementary gestures (a possible decomposition for measuring difficulty is described in section 8.3 or classical methods in industry such as MTM Methods Time Measurement). Due to the temporal nature of the signals, it is necessary to use a type of deep networks that manage this type of data. Recursive networks are therefore used where past observations influence the current prediction. Among recent deep network research, the so-called *long-short term memory* (LSTM) cells, represented here, seem well adapted. Unlike a simple recursive network where only data from the previous time is used for a new prediction, an LSTM cell can store data over a much longer period of time. With each prediction, the *forget gate* can decide to authorize the use or forget a previously observed data. We tested our algorithms on a classic benchmark (NTU RGB+D). In order to obtain interesting recognition rates, we showed that it was necessary to use the filters explained in section 7.3 to determinate the number of learning movements. Other less data-intensive methods are to be tested.

Chroma Project-Team

7. New Results

7.1. Bayesian Perception

Participants: Christian Laugier, Lukas Rummelhard, Jean-Alix David, Thomas Genevois, Jerome Lussereau, Nicolas Turro [SED], Jean-François Cuniberto [SED].

7.1.1. Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT) Framework

Participants: Lukas Rummelhard, Jerome Lussereau, Jean-Alix David, Thomas Genevois, Christian Laugier, Nicolas Turro [SED].

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 [83] and the associated risks of collision [85], 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 [84]. In 2018, this framework, and the corresponding software, has continued to be the core of many important industrial partnerships and academic contributions [17] [18] [16] [15] [45] [47], and to be the subject of important developments, both in terms of research and engineering. Some of those recent evolutions are detailed below.

- CMCDOT evolutions : important developments in the CMCDOT, in terms of calculation methods and fundamental equations, were introduced and tested this year. These developments could lead, in the coming months, to the proposal of a new patent, then to academic publications. These changes introduced, among other evolutions, a much higher update frequency, greater flexibility in the management of transitions between states (and therefore a better system reactivity), as well as the management of a high variability in sensor frequencies (for each sensor over time, and in the set of sensors). The technical documents describing those developments are currently being redacted, and will be described in the next annual report.
- Multi-sensor integration in the Ground Estimator : the module of dynamic estimation of the shape of the ground and data segmentation, based solely on the sensor point clouds (no prior map information), the first step of data interpretation in CMCDOT framework, has been developed since 2016, patented and published in 2017. The corresponding software, until this year, could not take into account more than one sensor. In case of multiple sensors, several different modules were to be launched, their respective occupancy grids then fused, not only increasing the global computation use, but also preventing each sensor from benefiting from the ground models generated by the others. This point was corrected this year, by introducing the management of multiple input sensors, unifying the ground estimation in a single model, thus leading to improved performance, both in terms of calculation and results.
- Velocity display : in the CMCDOT framework, velocity of every element of the scene is inferred at a cell level, without object segmentation. This low-level velocity estimation is one of the most original and important aspects of the method, and should be displayed accordingly. A velocity display module, displaying for each occupied cell of the grid the average of the estimated velocity, generating colors depending on the intensity and the orientation, has been developed, see Fig. 5 .
- 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 2018, on the optimization of the software and methods to be embedded on low energy consumption embedded boards (now Nvidia Jetson TX2).

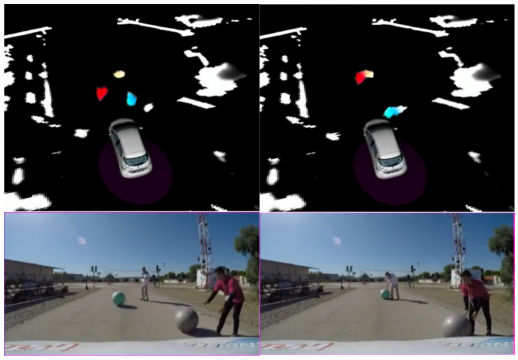


Figure 5. Image from the Velocity Display module : in every occupied cell of the grid, the average velocity is represented by a color code, the hue being based on the orientation, the saturation on its norm. A static cell is white, a cell moving in the same direction as the vehicle is red, in the opposite direction in blue. In the grid can be seen the moving balloons, the pedestrians being static.

- IROS 2018 Autonomous Driving event : https://hal.inria.fr/medihal-01963296v1 As already mentioned in the highlights of the year, the experimental Zoe platform, funded by IRT Nanoelec, has participated at IROS2018 in the Autonomous Vehicle Demonstrations, a full day of demonstration of autonomous vehicle capacities from various research centers. During this successful event, it has been presented and demonstrated on live conditions the effectiveness of the embedded CMCDOT framework, in connection with the newly developed control and decision making systems.

7.1.2. Simulation based validation

Participants: Thomas Genevois, Lukas Rummelhard, Nicolas Turro [SED], Christian Laugier, Anshul Paigwar, Alessandro Renzaglia.

Since 2017, we are working to address the concept of *simulation based validation* in the scope of the EU Enable-S3 project, with the objective of searching for novel approaches, methods, tools and experimental methodology for validating BOF-based algorithms. For that purpose, we have collaborated with the Inria Tamis team (Rennes) and with Renault for developing the simulation platform that is used in the test platform. The simulation of both the sensors and the driving environment are based on the Gazebo simulator. A simulation of the prototype car and its sensors has also been realized, meaning that the same implementation of *CMCDOT* can handle both real data and simulated data. The test management component that generates random simulated scenarios has also been developed. Output of *CMCDOT* computed from the simulated scenarios are recorded by *ROS* and analyzed through the Statistical Model Checker *(SMC)* developed by the Inria Tamis team. In [41], we presented the first results of this work, where a decision-making approach for intersection crossing (see Section 7.2.3) has been analyzed. In particular new KPIs expressed as Bounded Linear Temporal Logic (BLTL) formula have been defined. Temporal formulas allow a finer formulation of KPIs by taking into account the evolution of the metrics during time. A further work in this direction will be done in the next months to provide new results on the validation of the perception algorithm, namely for the velocity estimation and collision risk assessment. For this part, we are also exploring the advantages and potentiality of a new open-source vehicle simulator (Carla), which would allow considering more realistic scenarios with respect to Gazebo. This work on simulation-based validation will be continued in 2019.

Previously, in 2017, CHROMA has developed a model of the Renault Zoe demonstrator within the simulation framework Gazebo. In 2018, we have improved it to keep it up-to-date after several evolutions of the actual demonstrator. Namely, the drivers of the simulated lidars and the control law have been updated. Thus the model now provides the outputs corresponding to a simulated Inertial Measurement Unit.

7.1.3. Control and navigation

Participants: Thomas Genevois, Lukas Rummelhard, Jerome Lussereau, Jean-Alix David, Christian Laugier, Nicolas Turro [SED], Rabbia Asghar.

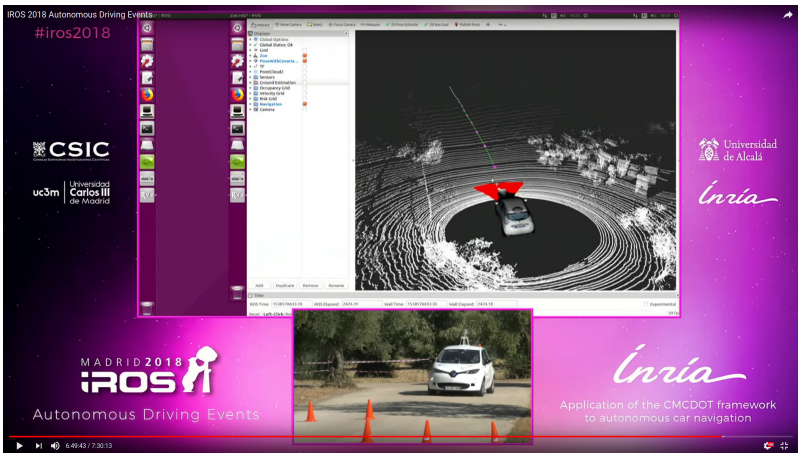


Figure 6. Image taken from the live diffusion of the Autonomous Vehicles event at IROS2018. The demonstrator Renault Zoe is about to go through an obstacle course

In 2018, we have updated the Renault Zoe demonstrator in collaboration with the LS2N (Laboratoire des Sciences Numérique de Nantes). The control codes have been transferred to the micro-controllers of the car for a faster and more precise control. An electric signal has been added to identify when the driver acts on the manual controls of the car. Finally the control law of the vehicle has been modified in order to consider a command in acceleration. These modifications allowed us to improve the software we use to control the vehicle. We have improved our implementation of DWA (Dynamic Window Approach) local planner in order to handle acceleration commands. This local planner has also been modified to take in account maxima of lateral acceleration and to integrate a path following module in its cost function. Thanks to this, the new version of this program provides a smooth command for a combination of path following and obstacle avoidance with the demonstrator Renault Zoe. This has been showed at the Autonomous Vehicle Demonstration event at IROS2018, Madrid, Figure 6 [46].

We have also experimented a driving assistant for autonomous obstacle avoidance. We showed that it is possible on the Renault Zoe demonstrator to let a driver drive manually the car and then, when a collision risk is identified, to take over the control with the autonomous drive and perform an avoidance maneuver. A simple ADAS [0] system has been developed for this purpose. In addition, we have developed on the Renault Zoe demonstrator, a localization system which merges the data of wheel speed, accelerometer, gyrometer, magnetometer and GPS into a position estimation. This relies on an Extended Kalman Filter. This will probably be extended later to consider the localization with respect to roads identified on a map.

Finally a Dijkstra Algorithm have been tested in simulation to define a global navigation path allowing management of waypoints to give to the DWA planner for local navigation.

7.2. Situation Awareness & Decision-making

Participants: Christian Laugier, Olivier Simonin, Jilles Dibangoye, David Sierra-Gonzalez, Mathieu Barbier, Victor Romero-Cano [Universidad Autónoma de Occidente, Cali, Colombia], Ozgur Erkent, Christian Wolf.

[0] Advanced Driving Assistance System

7.2.1. Dense & Robust outdoor perception for autonomous vehicles

Participants: Christian Laugier, Victor Romero-Cano, Özgür Erkent, Christian Wolf.

Robust perception plays a crucial role in the development of autonomous vehicles. While perception in normal and constant environmental conditions has reached a plateau, robustly perceiving changing and challenging environments has become an active research topic, particularly due to the safety concerns raised by the introduction of autonomous vehicles to public streets. Solving the robustness issue in road and urban perception applications is the first challenge. Then, it is also mandatory to develop an appropriate framework for extracting relevant semantic information. Our approach is to reason about vision-based data and the output of our grid-based multi-sensors perception approach (see previous section).

The work presented in this section has partly been done in 2017 and completed in 2018, in the scope of our collaboration with Toyota Motor Europe (TME). The main objective was to develop a framework for integrate the outcomes of the deep learning methods with a well-established area, occupancy grids obtained with a Bayesian filtering method in the grid space.

In this work, we are interested in 2D egocentric representations. We propose a method, which estimates an occupancy grid containing detailed semantic information. The semantic characteristics include classes like *road, car, pedestrian, sidewalk, building, vegetation, etc.*. To this end, we leverage and fuse information from multiple sensors including Lidar, odometry and monocular RGB video. To benefit from the respective advantages of the two different methodologies, we propose a hybrid approach leveraging i) the high-capacity of deep neural networks as well as ii) Bayesian filtering, which is able to model uncertainty in a unique way.

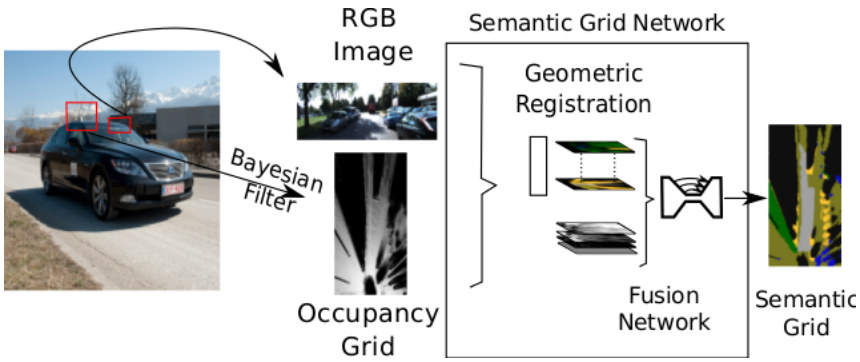


Figure 7. The Semantic Grid framework.

In the system depicted by Figure 7 , Bayesian particle filtering processes the Lidar data as well as odometry information from the vehicle's motion in order to robustly estimate an egocentric bird's eye view in the form of an occupancy grid. This grid contains a 360° view of the environment around the car and integrates information from the observation history through temporal filtering; however, it does not include fine-grained semantic classes.

Deep Learning is used for two different tasks in our work. Firstly, a deep network performs semantic segmentation of monocular RGB images. This network has been pre-trained on large scale datasets for image classification and fine-tuned on the vehicle datasets. Secondly, a deep network fuses the occupancy grid with the segmented image of the projective view in order to estimate the semantic grid. Since the occupancy grid is dense, the semantic grid is also expected to be dense. We pay particular attention to correctly model the transformation from the egocentric projective view of the RGB image to the bird's eye view of the occupancy grid as input to the neural network. This work was filed for a patent [98] and published in [28], [14].

Novel approach: Semantic Grid Estimation with a Hybrid Bayesian and Deep Neural Network Approach.

Current and future work in the scope of our collaboration with TME, aims at constructing *Semantic Occupancy Grids*. We propose a hybrid approach, which combines the advantages of Bayesian filtering and deep neural networks. Bayesian filtering provides robust temporal/geometrical filtering and integration and allows for modelling of uncertainty. RGB information and deep neural networks provide knowledge about the semantic class labels like *sideway* vs *road*. The fusion process is fully learned and due to dense structure of occupancy grid, we can construct a dense semantic grid even if we have a sparse point cloud.

7.2.2. Towards Human-Like Motion Prediction and Decision-Making in Highway Scenarios

Participants: David Sierra González, Victor Romero-Cano, Özgür Erkent, Jilles Dibangoye, Christian Laugier.

The objective is to develop human-like motion prediction and decision-making algorithms to enable automated driving in highways. This research work is done in the scope of the Inria-Toyota long-term cooperation on Autonomous Driving and of the PhD thesis work of David Sierra González.

Previous work from our team has shown the predictive potential of driver behavioral models learned from demonstrations using Inverse Reinforcement Learning (IRL) [87] [88]. Unfortunately, these models are hard to learn from real-world driving data due to the inability of traditional IRL algorithms to handle continuous state spaces and dynamic environments. To facilitate this task, we have proposed in 2018 an approximated IRL algorithm for driver behavior modeling that successfully scales to continuous spaces with moving obstacles, by leveraging a spatio-temporal trajectory planner [35]. The proposed algorithm was validated using real-world data gathered with an instrumented vehicle. As an example, Figure 8 shows the similarity between the trajectory obtained using a driver model learned with the proposed method and that of a real human driver in a highway overtake scenario. Current efforts are directed towards integrating the learned behavioral models and the predictive models developed in the scope of this project into a decision-making framework for highways. David Sierra González will defend his PhD thesis in March 2019.

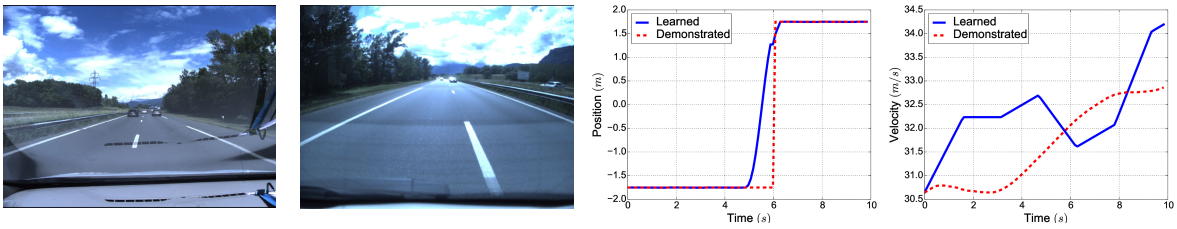


Figure 8.

Comparison of the trajectory obtained with a driver model learned from demonstrated driving data using the method proposed in [35] and that of a human driver for a typical highway overtake scenario
a. Front view at $t = 5.0$ b. Back view at $t = 5.0$ c. Position prediction d. Velocity prediction

7.2.3. Decision-making for safe road intersection crossing

Participants: Mathieu Barbier, Christian Laugier, Olivier Simonin.

Road intersections are probably the most complex segment in a road network. Most major accidents occur at intersections, mainly caused by human errors due to failures in fully understanding the encountered situations. Indeed, as drivers approach a road intersection, they must assess the situation and quickly adapt their behaviour accordingly. When this task is performed by a computer, the available information is partial and uncertain.

Any decision requires the system to use this information as well as taking into account the behaviour of other drivers to avoid collisions. However, metrics such as collision rate can remain low in an interactive environment because of other driver's actions. Consequently, evaluation metrics must depend on other driving aspects.

In this framework, we developed a decision-making mechanism and designed metrics to evaluate such a system at road intersection crossing [22]. For the former, a Partially Observable Markov Decision Process (POMDP) is used to model the system with respect to uncertainties in the behaviour of other drivers. For the latter, different key performance indicators are defined to evaluate the resulting behaviour of the system in different configurations and scenarios. The approach has been demonstrated within an automotive grade simulator.

Current work aims at increasing the complexity of the scenario, to include pedestrians and more vehicles, and improving the model used for the dynamics of the vehicle and the observation of the physical state to get closer to real world scenarios.

This work has been carried out in the framework of the PhD thesis of Mathieu Barbier, which will be defended in the first trimester of 2019.

7.3. 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.

7.3.1. Visual-inertial structure from motion

Participants: Agostino Martinelli, Alexander Oliva, Alessandro Renzaglia.

During this year, we have obtained the full analytic solution of the cooperative visual inertial sensor fusion problem in the case of two agents, starting from the closed-form solution obtained in the last years (this latter solution will be published on the journal of Autonomous Robots [76]). Additionally, we also validated this solution with real experiments and in particular we showed that the analytic solution significantly outperforms our previous closed-form solution in [76]. The new analytic solution has been accepted for publication by the IEEE Robotics and Automation Letters [13].

Specifically, we obtained the analytic solution of the problem by first proving that, this sensor fusion problem, is equivalent to a simple polynomial equations system that consists of several linear equations and three polynomial equations of second degree. The analytic solution of this polynomial equations system was easily obtained by using an algebraic method (developed by Bernard Mourrain, the leader of AROMATH at Inria Sophia Antipolis). The power of the analytic solution is twofold. From one side, it allows us to determine the relative state between the agents (i.e., relative position, speed and orientation) without the need of an initialization. From another side, it provides fundamental insights into all the theoretical aspects of the problem. During this year, we focused on the first issue. Our next objective is to exploit the analytic solution to obtain basic structural properties of the problem.

7.3.2. 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. The mathematics apparatus necessary to obtain this solution is very sophisticated and is based on Ricci calculus, borrowed from theoretical physics. On the other hand, this mathematics can be avoided in the case of driftless systems and characterized by a single unknown input.

All the results (i.e., in the general case that also accounts for a drift and more than one unknown input) are fully described in a book available on ArXiv (arXiv:1704.03252).

During this year, my effort was devoted to make the analytic derivation of the solution palatable for a large audience (in particular, without knowledge of Ricci calculus). Hence, I focused on the simple case of a single unknown input and without drift. This solution has been published on a full paper on the IEEE Transaction on Automatic Control [75].

Regarding the general case available on ArXiv (arXiv:1704.03252), I was invited by the SIAM to write a book, palatable for a large audience. The scope of writing this book, is to present to the control theory and information theory communities a very powerful mathematics framework borrowed from theoretical physics. This could provide the possibility of revisiting many aspects of the control and information theory and bring new fundamental results, open new research domains etc. In this sense the book could be the kick-off of a new season of research in control and information theory. This will be the objective of the next years.

7.4. Motion-planning in human-populated environment

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) . These works are presented here after.

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 car or platoons (PhD work of P. Vasishta). In 2017 we 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 [95], [94]. In 2018, an extension to the Growing Hidden Markov Model (GHMM) method 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. 9 .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 [36] and won the **best student paper** of the conference.

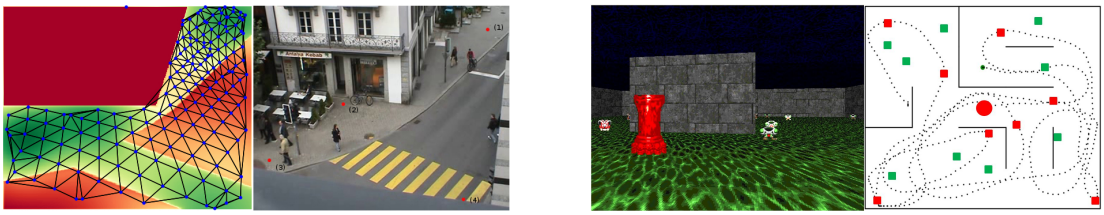


Figure 9.

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 learning task-based motion planning.

7.4.2. 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 human populated 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. recognize the current activities of all the actors in this environment) or which can be given in an encoded form as additional input. Addressing these problems requires competences in computer vision, machine learning, and robotics (navigation and paths planning).

We started this work in the end of 2017, following the arrival of C. Wolf, through combinations of reinforcement learning and deep learning. The underlying scientific challenge here is to automatic learn representations which allow the agent to solve multiple sub problems require 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, 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 who started on October 2018, see illustration Fig. 9 .b.

7.4.3. Human-flows modeling and social robots

Participants: Jacques Saraydaryan, Fabrice Jumel, Olivier Simonin, Benoit Renault, Laetitia Matignon, Christian Wolf.

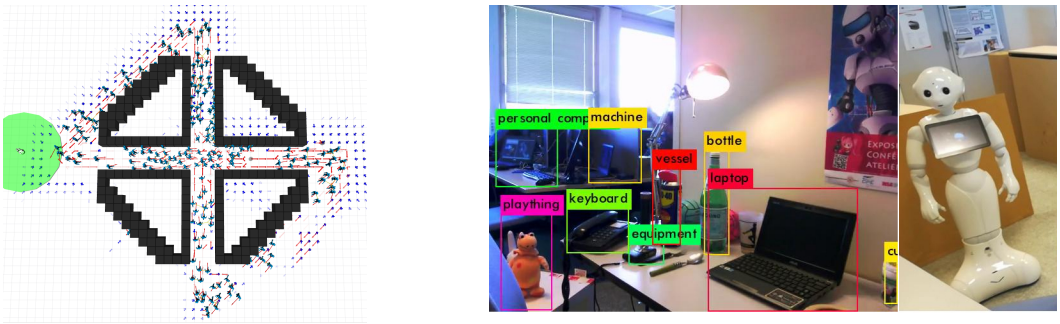


Figure 10.

(a) Flow-grid mapping in a cross-corridor where 200 moving pedestrians turns
(b) Object detection with Pepper based on vision/deep learning techniques.

In order to deal with robot/humanoid navigation in complex and populated environments such as homes, we investigate since 2 years several research avenues :

- Mapping humans flows. We defined a statistical learning approach (ie. a counting-based grid model) exploiting only data from robots embedded sensors. See illustration in Fig. 10 .a and publication [66].
- Path-planning in human flows. We revisited the A* path-planning cost function under the hypothesis of the knowledge of a flow grid. See publication [66].
- In 2018 we started to study NAMO problems (Navigation Among Movable Obstacles) by considering populated environments and muti-robot cooperation. After his Master thesis on this subject, Benoit Renault started a PhD in Chroma focusing on the extension of NAMO algorithms to such dynamic environments.

- RoboCup competition. In the context of the **RoboCup** international competition, we created the 'LyonTech' team, joining members from Chroma (INSA/CPE/UCBL). We investigated several humanoid tasks in home environments with our Pepper robot : social aware architecture, decision making and navigation, deep-learning based human and object detection (see Fig. 10 .b), human-robot interaction. In July 2018, we participated for the first time to the RoboCup and reaching the 5th rank of the SSL league (Pepper@home). We also published our social-aware architecture to the RoboCup Conference [31]. In October 2018, we qualified for the next final phase of RoboCup SSL (Pepper) to be organized on July 2019, in Sydney.

7.5. Decision Making in Multi-Robot Systems

7.5.1. Multi-robot planning in dynamic environments

7.5.1.1. Global-local optimization in autonomous multi-vehicles 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 [53]. We were surprise 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.

7.5.1.2. 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 in 2016 [81], 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 : Multi-robot routing has been the main testbed in the domain of multi-robot task allocation, where 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 submitted to RSS'2019 conference.
- 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, like standard SSI or SSI with regret clearing, especially in optimizing the target allocation w.r.t. the target coverage in time and the robot resource usage (e.g. minimizing the worst mission cost). First results have been published in [34].

- 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 (e.g. the base station). 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. We developed a simulator, allowing experiments that show the approaches have different performances and robustness. This work will be submitted to IROS'2019 conference.
- PHC DRONEM [0] : We started a collaboration in 2017 with the team of Prof. Gabriela Czibula from Babes-Bolyai University in Cluj-Napoca, Romania. The DRONEM project focuses on optimization and online adaptation of the multi-cycle patrolling with machine learning (RL) techniques in order to deal with the arrival of new targets in the environment.

7.5.1.3. Middleware for open multi-robot systems

Participants: Stefan Chitic, Julien Ponge [INSA/CITI, Dynamid], Olivier Simonin.

Multi-robots systems (MRS) require dedicated software tools and models to face the complexity of their design and deployment. In the context of the PhD work of Stefan Chitic, we addressed service self-discovery and property proofs in an ad-hoc network formed by a fleet of robots. This led us to propose a robotic middleware, SDfR, that is able to provide service discovery, see [54]. In 2017, we defined a tool-chain based on timed automata, called ROSMDB, that offers a framework to formalize and implement multi-robot behaviors and to check some (temporal) properties (both offline and online). Stefan Chtic defended his Phd thesis on March 2018 [11].

7.5.2. Multi-robot Coverage and Mapping

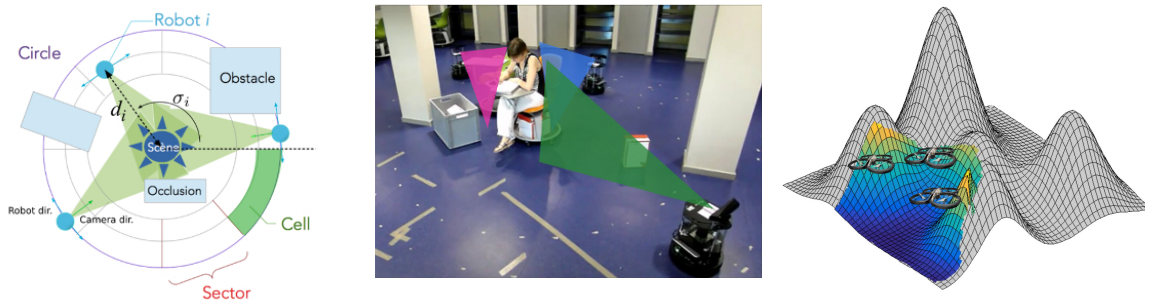


Figure 11. (a) Concentric navigation model and (b) its experimental setup. (c) Illustration of the local search method for multi-UAV coverage.

7.5.2.1. Human scenes observation

Participants: Laetitia Matignon, Olivier Simonin, Stephane d'Alu, Christian Wolf.

[0] Hubert Curien Partnership

Solving complex tasks with a fleet of robots requires to develop generic strategies that can decide in real time (or time-bounded) efficient and cooperative actions. This is particularly challenging in complex real environments. To this end, we explore anytime algorihms and adaptive/learning techniques.

The "CROME" and "COMODYS" [0] projects [0] are motivated by the exploration of the joint-observation of complex (dynamic) scenes by a fleet of mobile robots. In our current work, the considered scenes are defined as a sequence of activities, performed by a person in a same place. Then, mobile robots have to cooperate to find a spatial configuration around the scene that maximizes the joint observation of the human pose skeleton. It is assumed that the robots can communicate but have no map of the environment and no external localisation.

To attack the problem, we proposed an original concentric navigation model allowing to keep easily each robot camera towards the scene (see fig. 11 .a). This model is combined with an incremental mapping of the environment and exploration guided by meta-heuristics in order to limit the complexity of the exploration state space. Results have been published in AAMAS'2018 [32]. An extended version has been submitted to the Journal JAAMAS.

For experiment with multi-robot systems, we defined an hybrid metric-topological mapping. Robots individually build a map that is updated cooperatively by exchanging only high-level data, thereby reducing the communication payload. We combined the on-line distributed multi-robot decision with this hybrid mapping. These modules has been evaluated on our platform composed of several Turtlebots2, see fig. 11 .b. This robotic architecture has been presented in [77] (ECMR). A Demo has been done in AAMAS'2018 international conference [33].

7.5.2.2. Multi-UAV Visual Coverage of Partially Known 3D Surfaces

Participants: Alessandro Renzaglia, Olivier Simonin, Jilles Dibangoye, Vincent Le Doze.

It has been largely proved that the use of Unmanned Aerial Vehicles (UAVs) is an efficient and safe way to deploy visual sensor networks in complex environments. In this context, a widely studied problem is the cooperative coverage of a given environment. In a typical scenario, a team of UAVs is called to achieve the mission without a perfect knowledge on the environment and needs to generate the trajectories on-line, based only on the information acquired during the mission through noisy measurements. For this reason, guaranteeing a global optimal solution of the problem is usually impossible. Furthermore, the presence of several constraints on the motion (collision avoidance, dynamics, etc.) as well as from limited energy and computational capabilities, makes this problem particularly challenging.

Depending on the sensing capabilities of the team (number of UAVs, range of on-board sensor, etc.) and the dimension of the environment to cover, different formulations of this problem can be considered. We firstly approached the deployment problem, where the goal is to find the optimal static UAVs configuration from which the visibility of a given region is maximized. A suitable way to tackle this problem is to adopt derivative-free optimization methods based on numerical approximations of the objective function. In 2012, Renzaglia et al. [82] proposed an approach based on a stochastic optimization algorithm to obtain a solution for arbitrary, initially unknown 3D terrains (see fig. 11 .c). However, adopting this kind of approaches, the final configuration can be strongly dependent on the initial positions and the system can get stuck in local optima very far from the global solution. We identified that a way to overcome this problem can be found in initializing the optimization with a suitable starting configuration. An a priori partial knowledge on the environment is a fundamental source of information to exploit to this end. The main contribution of our work is thus to add another layer to the optimization scheme in order to exploit this information. This step, based on the concept of Centroidal Voronoi Tessellation, will then play the role of initialization for the on-line, measurement-based local optimizer. The resulting method, taking advantages of the complementary properties of geometric and stochastic optimization, significantly improves the result of the previous approach 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. This work led to a paper submitted to AAMAS 2019 [0], currently under review. The development of a realistic simulation environment based on

[0] COoperative Multi-robot Observation of DYnamic human poSes

[0] Funded by a LIRIS transversal project in 2016-2017 and a FIL project in 2017-2019 (led by L. Matignon)

Gazebo is an important on-going activity in Chroma and will allow us to further test the approach and to prepare the implementation of this algorithm on the real robotic platform of the team.

We are currently also investigating the dynamic version of this problem, where the information is collected along the trajectories and the environment reconstruction is obtained from the fusion of the total visual data.

7.5.3. Sequential decision-making

This research is the follow up of a group led by Jilles S. Dibangoye carried out during the last three 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.

7.5.3.1. Optimally solving cooperative and competitive games as continuous Markov decision processes

Participants: Jilles S. Dibangoye, Olivier Buffet [Inria Nancy], Vincent Thomas [Inria Nancy], Christopher Amato [Univ. New Hampshire], François Charpillet [Inria Nancy, Larsen team].

Our major findings this year include:

1. (Theoretical) – As an extension of [58] in the cooperative case [44], we characterize the optimal solution of partially observable stochastic games.
2. (Theoretical) – We further exhibit new underlying structures of the optimal solution for both cooperative and non-cooperative settings.
3. (Algorithmic) – We extend a non-trivial procedure in [27] for computing such optimal solutions when only an incomplete knowledge about the model is available.

This work proposes a novel 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 [50] for zs-POSGs, and coined them maximin (resp. minimax) equations. Even more importantly, we demonstrated Von Neumann & Morgenstern's minimax theorem [99] [100] holds in zs-POSGs. We further proved that value functions—solutions of maximin (resp. minimax) equations—yield special structures. More specifically, the maximin value functions are convex whereas the minimax value functions are concave. Even more surprisingly, we prove that for a fixed strategy, the optimal value function is linear. Together these findings allow us to extend planning and learning 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 [27].

7.5.3.2. Learning to act in (continuous) decentralized partially observable Markov decision process

Participants: Jilles S. Dibangoye, Olivier Buffet [Inria Nancy].

During the last year, we investigated deep and standard reinforcement learning for solving decentralized partially observable Markov decision processes. Our preliminary results include:

1. (Theoretical) Proofs that the optimal value function is linear in the occupancy-state space, the set of all possible distributions over hidden states and histories.
2. (Algorithmic) Value-based and policy-based (deep) reinforcement learning for common-payoff partially observable stochastic games.

[0] A. Renzaglia, J. Dibangoye, V. Le Doze and O. Simonin, "Multi-UAV Visual Coverage of Partially Known 3D Surfaces: Voronoi-based Initialization to Improve Local Optimizers", International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2019, *under review*.

This work addresses a long-standing open problem of Multi-Agent Reinforcement Learning (MARL) in decentralized stochastic control. MARL previously applied to finite decentralized decision making with a focus on team reinforcement learning methods, which at best lead to local optima. In this research, we build on our recent approach [44], which converts the original problem into a continuous-state Markov decision process, allowing knowledge transfer from one setting to the other. In particular, we introduce the first optimal reinforcement learning method for finite cooperative, decentralized stochastic control domains. We achieve significant scalability gains by allowing the latter to feed deep neural networks. Experiments show our approach can learn to act optimally in many finite decentralized stochastic control problems from the literature [43], [26].

7.5.3.3. Study of policy-gradient methods for decentralized stochastic control

Participants: Guillaume Bono, Jilles S. Dibangoye, Laëtitia Matignon, Olivier Simonin, Florian Peyreron [VOLVO Group, Lyon].

This work is part of the Ph.D. thesis in progress of Guillaume Bono, with VOLVO Group, in the context of the INSA-VOLVO Chair. The work aims at investigating an attractive family of reinforcement learning methods, namely policy-gradient and more generally actor-critic methods for solving decentralized partially observable Markov decision processes. Our preliminary results include:

1. (Theoretical) Proofs of the policy-gradient theorems for both total- and discounted-reward criteria in decentralized stochastic control.
2. (Algorithmic) (deep) actor-critic reinforcement learning methods for centralized and decentralized stochastic control.

Reinforcement Learning (RL) for decentralized partially observable Markov decision processes (Dec-POMDPs) is lagging behind the spectacular breakthroughs of single-agent RL. That is because assumptions that hold in single-agent settings are often obsolete in decentralized multi-agent systems. To tackle this issue, we investigate the foundations of policy gradient methods within the centralized training for decentralized control (CTDC) paradigm. In this paradigm, learning can be accomplished in a centralized manner while execution can still be independent. Using this insight, we establish policy gradient theorem and compatible function approximations for decentralized multi-agent systems. Resulting actor-critic methods preserve the decentralized control at the execution phase, but can also estimate the policy gradient from collective experiences guided by a centralized critic at the training phase. Experiments demonstrate our policy gradient methods compare favorably against standard RL techniques in benchmarks from the literature [42], [23]. Guillaume Bono also designed a simulator for urban logistic reinforcement learning, namely SULFR [39].

7.5.3.4. 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 [56] and iterative local search [97]. 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.

DEFROST Project-Team

7. New Results

7.1. Dynamic control of soft robots

The objective is to design a closed-loop strategy to control the dynamics of soft robots. We model the soft robot using the Finite Element Method, which leads to work with large-scale systems that are difficult to control. No unified framework exist to control these robots, especially when considering their dynamics. The main contribution of our work is a reduced order model-based control law, that consists in two main features: a reduced state feedback tunes the performance while a Lyapunov function guarantees the stability of the large-scale closed-loop systems. The method is generic and usable for any soft robot, as long as a FEM model is obtained. Simulation and real robots experiments show that we can control and reduce the settling time of the soft robot and make it converge faster without oscillations to a desired position. It can make the robot converge faster and with reduced oscillations to a desired equilibrium state in the robot's work-space. These results have been presented at the European Control Conference [24] and accepted for publication in Robotics and Automation Letters [8].

7.2. Vision-based force sensing for soft robots

This paper proposes a new framework of external force sensing for soft robots based on the fusion of vision-based measurements and Finite Element Model (FEM) techniques. A precise mechanical model of the robot is built using real-time FEM to describe the relationship between the external forces acting on the robot and the displacement of predefined feature points. The position of these feature points on the real robot is measured using a vision system and is compared with the equivalent feature points in the finite element model. Using the compared displacement, the intensities of the external forces are computed by solving an inverse problem. Based on the developed FEM equations, we show that not only the intensities but also the locations of the external forces can be estimated. A strategy is proposed to find the correct locations of external forces among several possible ones. The method is verified and validated using both simulation and experiments on a soft sheet and a parallel soft robot (both of them have non-trivial shapes). The good results obtained from the experimental study demonstrate the capability of our approach.

7.3. Fast, generic and reliable control and simulation of soft robots using model order reduction

Obtaining an accurate mechanical model of a soft deformable robot compatible with the computation time imposed by robotic applications is often considered as an unattainable goal. This paper should invert this idea. The proposed methodology offers the possibility to dramatically reduce the size and the online computation time of a Finite Element Model (FEM) of a soft robot. After a set of expensive offline simulations based on the whole model, we apply snapshot-proper orthogonal decomposition to sharply reduce the number of state variables of the soft robot model. To keep the computational efficiency, hyper-reduction is used to perform the integration on a reduced domain. The method allows to tune the error during the two main steps of complexity reduction. The method handles external loads (contact, friction, gravity...) with precision as long as they are tested during the offline simulations. The method is validated on two very different examples of FE models of soft robots and on one real soft robot. It enables acceleration factors of more than 100, while saving accuracy, in particular compared to coarsely meshed FE models and provides a generic way to control soft robots.

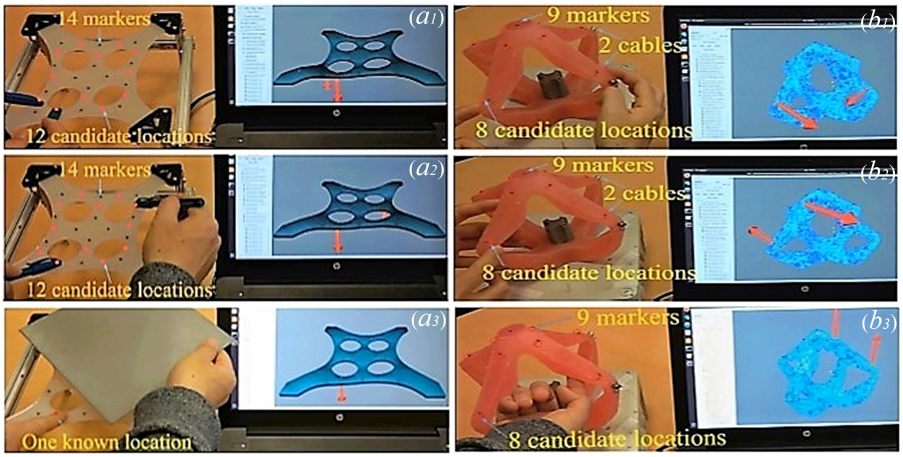


Figure 5. External force sensing for soft objects

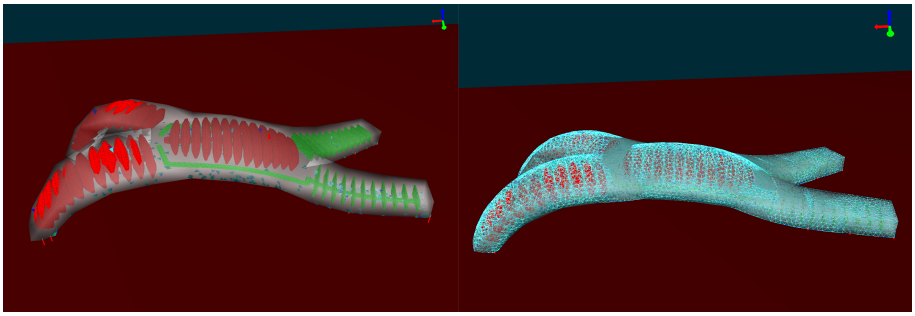


Figure 6. Pneumatic Soft Robot fine simulation versus its surrogate reduced representation manageable in real-time.

7.4. FEM-based kinematics and closed-loop control of soft, continuum manipulators

This paper presents a modeling methodology and experimental validation for soft manipulators to obtain forward and inverse kinematic models under quasistatic conditions. It offers a way to obtain the kinematic characteristics of this type of soft robots that is suitable for offline path planning and position control. The modeling methodology presented relies on continuum mechanics which does not provide analytic solutions in the general case. Our approach proposes a real-time numerical integration strategy based on Finite Element Method (FEM) with a numerical optimization based on Lagrangian Multipliers to obtain forward and inverse models. To reduce the dimension of the problem, at each step, a projection of the model to the constraint space (gathering actuators, sensors and end-effector) is performed to obtain the smallest number possible of mathematical equations to be solved. This methodology is applied to obtain the kinematics of two different manipulators with complex structural geometry. An experimental comparison is also performed in one of the robots, between two other geometric approaches and the approach that is showcased in this paper. A closed-loop controller based on a state estimator is proposed. The controller is experimentally validated and its robustness is evaluated using Lyapunov stability method.

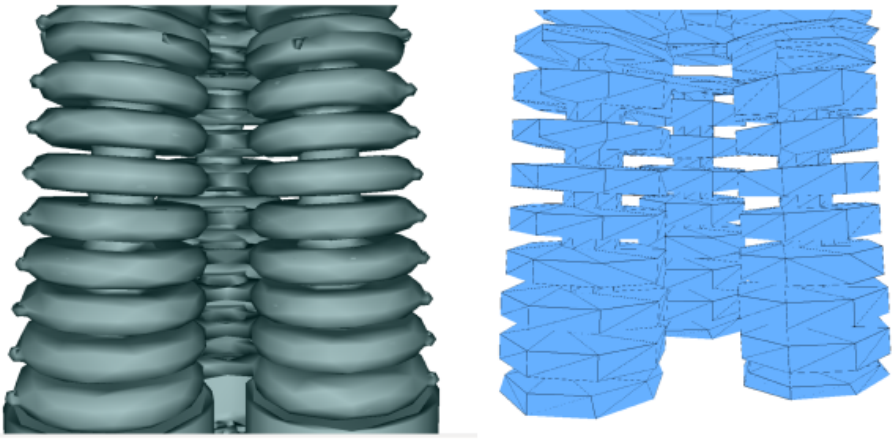


Figure 7. Visual model of the manipulator and the underlying finite element model.

7.5. FEM-based Deformation Control for Dexterous Manipulation of 3D Soft Objects

In this project, that was organize through a collaboration with Fanny Ficuciello from University of Naples and Antoine Petit from Mimesis team in Strasbourg we developed a method for dexterous manipulation of 3D soft objects for real-time deformation control, relying on Finite Element modelling. The goal is to generate proper forces on the fingertips of an anthropomorphic device during in-hand manipulation to produce desired displacements of selected control points on the object. The desired motions of the fingers are computed in real-time as an inverse solution of a Finite Element Method (FEM), the forces applied by the fingertips at the contact points being modelled by Lagrange multipliers. The elasticity parameters of the model are preliminarily estimated using a vision system and a force sensor. Experimental results were shown with an underactuated anthropomorphic hand that performs a manipulation task on a soft cylindrical object.

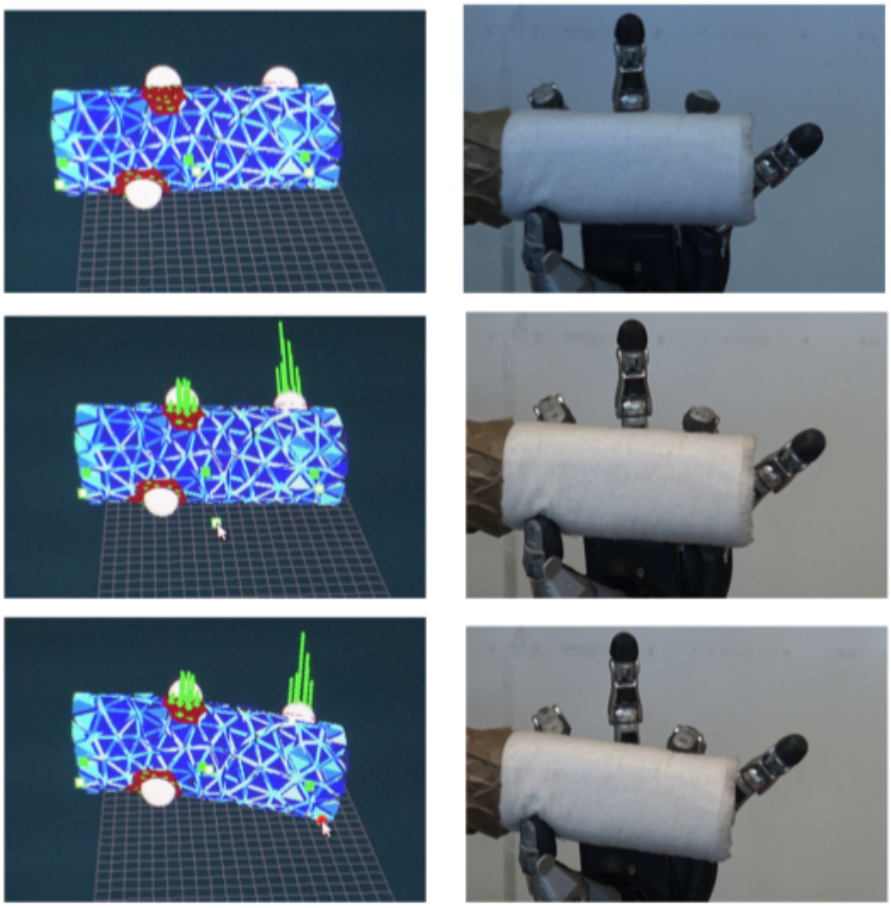


Figure 8. The manipulation of the 3D Soft Object inside the hand is driven by the inverse FEM simulation computed in real-time

FLOWERS Project-Team

7. New Results

7.1. Computational Models Of Human Learning and Development

7.1.1. Computational Models Of Information-Seeking and Curiosity-Driven Learning in Humans and Animals

Participants: Pierre-Yves Oudeyer [correspondant], William Schueller, 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 [100], [125], [135]. Therefore we hypothesize that, rather than using a single optimization process as it has been the case in most previous theoretical work [82], 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 [123] [54], [111]. 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 [98], [134], and rats show individual variation in learning styles and novelty seeking behaviors [74], 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 [130], [146], [48], [150]. 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 [98], [93] 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 performance-based or LP-based exploration in two contexts in which they could freely choose to learn about 4 competing tasks. Participants (n = 505, recruited via Amazon Mechanical Turk) were tested on a paradigm in which they could freely choose to engage with one of four different classification tasks. 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 currently analyzing the results and working on a computational models of the underlying cognitive and motivational mechanisms.

The central question going into the study was, how active learners become interested in specific learning exercises: how do they decide which task to be interested in – i.e., allocate "study time" - given that the underlying rewards or patterns are sparse and unknown? Using a family of statistical (multinomial logit), subjective-utility-based models of discrete choice behavior [109] we performed an exploratory all-subsets model selection exercise [61] to see if we can identify important and/or interesting variables that could reliably predict task choices. The initial set of variables included, among other things, various performance-based competence heuristics (e.g. current hit rate, likelihood of current hit rate). Model selection and multimodel inference pointed to a handful of variables that had *relatively* high influence on task choices (including the

likelihood of current hit rate and relative amount of time spent on a task), but their absolute effects were small, leaving most of the variation in task choices unexplained. This exercise also pointed out the potential limitations of our approach, either in operationalization of competence as a purely performance-based statistic, or in the potential lack of behavioral constraints in design of the experiment (participants may have been basing their choices on unanticipated variables). This latter limitation is tricky, since we are interested in exploratory behavior in unconstrained settings. What could have alleviated this challenge is a more diverse set of measurements that could include, for example, online records of participants' subjective feelings of interest, competence, liking, or learning potential. At this point, results concerning the LP hypothesis still have not revealed themselves, but we have gained valuable clues on how to find them. The next important step is to use cognitive models with transparent knowledge representations (e.g. Bayesian classifiers or neural networks) as an alternative way to operationalize subjective feelings of competence. The cognitive modeling approach emphasizes the idealistic assumptions made about the mind and examines their implicated behavioral outcomes. By doing that, cognitive models of learning and subjective competence can show whether our assumptions about the cognitive processes involved lead to the same behavioral patterns as the ones humans actually produce.

Although, the results of the Active Categorization study are still inconclusive, we found some interesting interim behavioral trends that are worth replicating and investigating. Participants showed preference for tasks of what we intended to be extreme complexity (i.e. too easy or too difficult) by spending more time on them (see figure 8). The group that was instructed to explore freely allocated their time more evenly, but showed a slight preference towards the easiest task where classification was based on a single dimension. The group that was instructed to try to maximize their learning during the experiment and expected a test at the end spent the majority of their time on the hardest (in fact, impossible) task to learn, where class assignment was independent of the two dimensions of variability. This suggests that active sampling strategies are subjected to extraneous constraints, and specifically, that some constraints may lead to inefficient exploration. It also potentially challenges the LP hypothesis, but it is to early to come to any strong conclusions about that, since we do not know how difficulty of the tasks was ranked subjectively by the participants.

Another puzzling observation comes from self-reported meta-cognitive judgments about the tasks. Figure 9 shows the average (min-max normalized) ratings of future learning potential and sensing the existence of a rule of each task. It is not clear why the learning potential for the hardest task (R) was reported to be high, despite the fact that it was believed to have no rule for classification. On the one hand, it is possible that while participants had not discovered the rule yet, they might have still believed there was a rule to be discovered. On the other hand, participants could really believe tha there was no rule to be discovered, but were not confident in that judgment, so high learning potential relates not to classification per se, but to discovering an interesting aspect of the task itself. There are other competing interpretations. Again, these observations compel us to better understand the contents of knowledge and knowledge-dependent processes used in the task, which we hope to achieve by applying and examining computational cognitive models of learning and meta-cognition.

7.1.2. Computational Models Of Tool Use and Speech Development: the Roles of Active Learning, Curiosity and Self-Organization

Participants: Pierre-Yves Oudeyer [correspondant], Sébastien Forestier, Rémy Portelas.

7.1.2.1. Modeling Speech and Tool Use Development in Infants

A scientific challenge in developmental and social robotics is to model how autonomous organisms can develop and learn open repertoires of skills in high-dimensional sensorimotor spaces, given limited resources of time and energy. This challenge is important both from the fundamental and application perspectives. First, recent work in robotic modeling of development has shown that it could make decisive contributions to improve our understanding of development in human children, within cognitive sciences [82]. Second, these models are key for enabling future robots to learn new skills through lifelong natural interaction with human users, for example in assistive robotics [127].

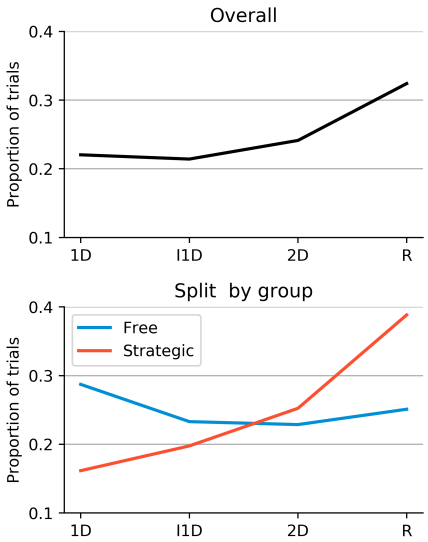


Figure 8. Proportion of trials on each task (1D, I1D, 2D, and R). 1D was the task were categorization was determined by a single variable dimension. In I1D (ignore 1D), the stimuli varied across 2 dimensions, but only one determined the stimulus category. In 2D, there were 2 variable dimensions and both 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.

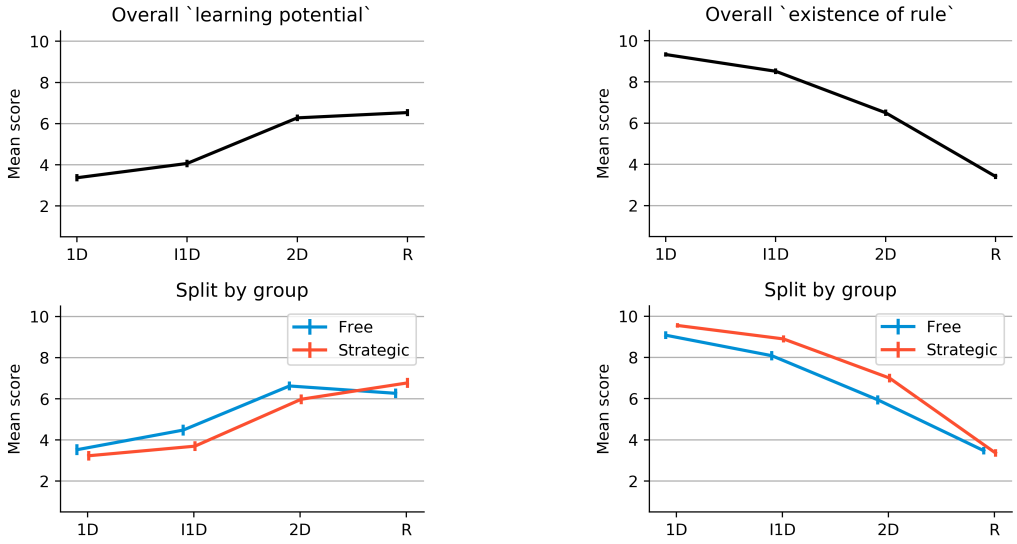


Figure 9. Average self-reported ratings of learning potential and existence of a rule for each task (1D, I1D, 2D, R; see figure 8 for disambiguation). The learning potential ratings ("Rate each monster family based on how much more you think you could learn if you had more time to play with it") were reported on a 10-point scale ([1] Definitely Could Not Learn More – [10] Definitely Could Learn More). The existence of rule ratings ("Rate each monster family based on how likely you think it had a rule for food preferences") were similarly reported ([1] Definitely No Rule – [10] Definitely a Rule. The error bars represent standard errors. The top plots show data aggregated across experimental groups, shown separately in the bottom plots.

In recent years, two strands of work have shown significant advances in the scientific community. On the one hand, algorithmic models of active learning and imitation learning combined with adequately designed properties of robotic bodies have allowed robots to learn how to control an initially unknown high-dimensional body (for example locomotion with a soft material body [53]). On the other hand, other algorithmic models have shown how several social learning mechanisms could allow robots to acquire elements of speech and language [62], allowing them to interact with humans. Yet, these two strands of models have so far mostly remained disconnected, where models of sensorimotor learning were too "low-level" to reach capabilities for language, and models of language acquisition assumed strong language specific machinery limiting their flexibility. Preliminary work has been showing that strong connections are underlying mechanisms of hierarchical sensorimotor learning, artificial curiosity, and language acquisition [128].

Recent robotic modeling work in this direction has shown how mechanisms of active curiosity-driven learning could progressively self-organize developmental stages of increasing complexity in vocal skills sharing many properties with the vocal development of infants [112]. Interestingly, these mechanisms were shown to be exactly the same as those that can allow a robot to discover other parts of its body, and how to interact with external physical objects [122].

In such current models, the vocal agents do not associate sounds to meaning, and do not link vocal production to other forms of action. In other models of language acquisition, one assumes that vocal production is mastered, and hand code the meta-knowledge that sounds should be associated to referents or actions [62]. But understanding what kind of algorithmic mechanisms can explain the smooth transition between the learning of vocal sound production and their use as tools to affect the world is still largely an open question.

The goal of this work is to elaborate and study computational models of curiosity-driven learning that allow flexible learning of skill hierarchies, in particular for learning how to use tools and how to engage in social interaction, following those presented in [122], [53], [117], [112]. The aim is to make steps towards addressing the fundamental question of how speech communication is acquired through embodied interaction, and how it is linked to tool discovery and learning.

We take two approaches to study those questions. One approach is to develop robotic models of infant development by looking at the developmental psychology literature about tool use and speech and trying to implement and test the psychologists' hypotheses about the learning mechanisms underlying infant development. Our second approach is to directly collaborate with developmental psychologists to analyze together the data of their experiments and develop other experimental setup that are well suited to answering modeling questions about the underlying exploration and learning mechanisms. We thus started last year a collaboration with Lauriane Rat-Fischer, a developmental psychologist working on the emergence of tool use in the first years of human life (now in Université Paris-Nanterre). We are currently analyzing together the behaviour of 22 month old infants in a tool use task where the infants have to retrieve a toy put in the middle of a tube by inserting sticks into the tube and pushing the toy out. We are looking at the different actions of the infant with tools and toys but also its looking behaviour, towards the tool, toys or the experimenter, and we are studying the multiple goals and exploration strategies of the babies other than the salient goal that the experimenter is pushing the baby to solve (retrieving a toy inside a tube).

In our recent robotic modeling work, we showed that the Model Babbling learning architecture allows the development of tool use in a robotic setup, through several fundamental ideas. First, goal babbling is a powerful form of exploration to produce a diversity of effects by self-generating goals in a task space. Second, the possible movements of each object define a task space in which to choose goals, and the different task spaces form an object-based representation that facilitates prediction and generalization. Also, cross-learning between tasks updates all skills while exploring one in particular. A novel insight was that early development of tool use could happen without a combinatorial action planning mechanism: modular goal babbling in itself allowed the emergence of nested tool use behaviors.

Last year we extended this architecture so that the agent can imitate caregiver's sounds in addition to exploring autonomously [78]. We hypothesized that these same algorithmic ingredients could allow a joint unified development of speech and tool use. Our learning agent is situated in a simulated environment where a vocal tract and a robotic arm are to be explored with the help of a caregiver. The environment is composed of three

toys, one stick that can be used as a tool to move toys, and a caregiver moving around. The caregiver helps in two ways. If the agent touches a toy, the caregiver produces this toy's name, but otherwise produces a distractor word as if it was talking to another adult. If the agent produces a sound close to a toy's name, the caregiver moves this toy within agent reach (see Fig. 10).

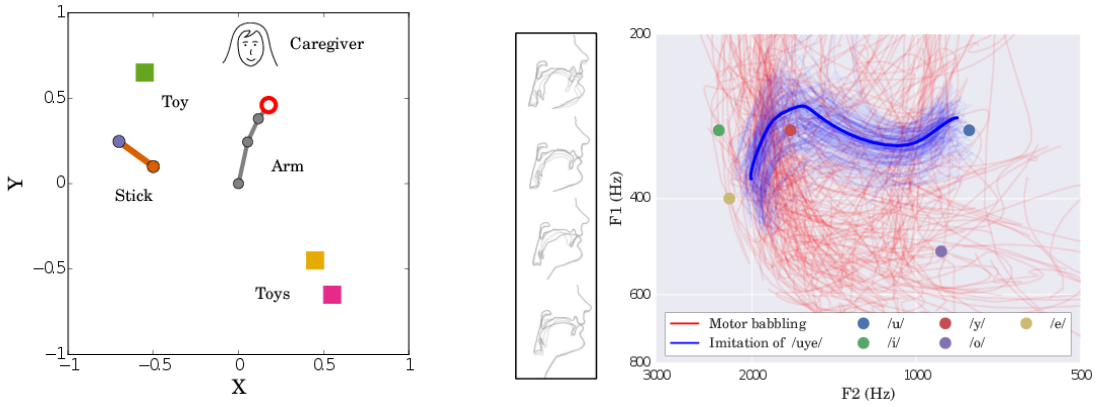


Figure 10. Agent's robotic and vocal environment. Left: Agent's 3 DOF arm, controlled with 21 parameters, grabs toys with its hand, or uses the stick to reach toys. Caregiver brings a toy within reach if the agent says its name. Right: Agent's vocal environment representing sounds as trajectories in the two first formants space. Agent's simulated vocal tract produces sounds given 28 parameters. When agent touches a toy, caregiver says toy's name. Some sounds corresponding to random parameters are plotted in red, and some sounds produced when imitating caregiver's /uye/ word in blue (best imitation in bold, error 0.3*).*

We showed that our learning architecture based on Model Babbling allows agents to learn how to 1) use the robotic arm to grab a toy or a stick, 2) use the stick as a tool to get a toy, 3) learn to produce toy names with the vocal tract, 4) use these vocal skills to get the caregiver to bring a specific toy within reach, and 5) choose the most relevant of those strategies to retrieve a toy that can be out-of-reach. Also, the grounded exploration of toys accelerates the learning of the production of accurate sounds for toy names once the caregiver is able to recognize them and react by bringing them within reach, with respect to distractor sounds without any meaning in the environment. Our model is the first to allow the study of the early development of tool use and speech in a unified framework. It predicts that infants learn to vocalize the name of toys in a natural play scenario faster than learning other words because they often choose goals related to those toys and engage caregiver's help by trying to vocalize those toys' names.

This year, we extended that model and we are currently studying on the one hand the impact of a partially contingent caregiver on agent's learning, and on the other hand the impact of attentional windows in agent's sensory perception, to see if and how an attentional window that do not match the time structure of the interaction with the caregiver could impair learning.

We also transposed this experiment to a real robotic setting during a 6-months research internship dedicated to study how IMGEP approaches scale to a real-world robotic setup. This work is related to ongoing research on simulating human babies' curiosity-driven learning mechanisms, which objectives are to test psychologists' hypotheses on human learning and to leverage these models to increase efficiency in reinforcement learning applied to robotics. Previous experiments [78] showed in simulation that intrinsically motivated reinforcement learning could be successfully applied to the early developments of speech and tool-use. The main goal of this internship was to extend this work by designing a real-world robotic experiment using a Poppy-Torso robot and a Baxter. The contributions made during this internship were 1) The design of the Poppy-Baxter

robotic playground (see figure [11] including the implementation of the communication architecture using ROS and the modeling of a 3D-printed toy, 2) Tuning of the experiment's parameters and learning process and 3) Analysis of the results in terms of exploration. Using this setup, we showed that the intrinsically motivated approach to model the early developments of speech and tool use developed in simulation can successfully scale to such a real-world experiment. Our curiosity-driven agents efficiently learned to vocalize the toy's name and to handle it in various and complex ways.

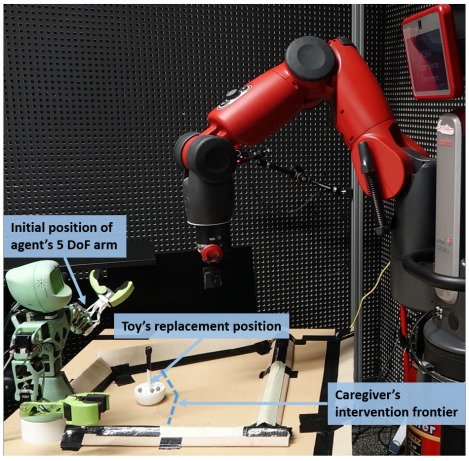


Figure 11. The POBAX Playground, in which the learning agent (Poppy-Torso) is able to interact with its arm and a simulated vocal tract. For each episode, if the agent touches the toy, the Caregiver (Baxter robot) says its name, otherwise the caregiver gives one of 3 distractor names. If the agent says the toys' name, the caregiver replaces it within the agent's arm-reach.

7.1.3. Models of Self-organization of lexical conventions: the role of Active Learning and Active Teaching in Naming Games

Participants: William Schueller [correspondant], Pierre-Yves Oudeyer.

How does language emerge, evolve and gets transmitted between individuals? What mechanisms underly the formation and evolution of linguistic conventions, and what are their dynamics? Computational linguistic studies have shown that local interactions within groups of individuals (e.g. humans or robots) can lead to self-organization of lexica associating semantic categories to words [143]. However, it still doesn't scale well to complex meaning spaces and a large number of possible word-meaning associations (or lexical conventions), suggesting high competition among those conventions.

In statistical machine learning and in developmental sciences, it has been argued that an active control of the complexity of learning situations can have a significant impact on the global dynamics of the learning process [82], [92], [101]. This approach has been mostly studied for single robotic agents learning sensorimotor affordances [123], [113]. However active learning might represent an evolutionary advantage for language formation at the population level as well [128], [145].

Naming Games are a computational framework, elaborated to simulate the self-organization of lexical conventions in the form of a multi-agent model [144]. Through repeated local interactions between random couples of agents (designated *speaker* and *hearer*), shared conventions emerge. Interactions consist of uttering a word – or an abstract signal – referring to a topic, and evaluating the success or failure of communication.

However, in existing works processes involved in these interactions are typically random choices, especially the choice of a communication topic.

The introduction of active learning algorithms in these models produces significant improvement of the convergence process towards a shared vocabulary, with the speaker [121], [140], [67] or the hearer [141] actively controlling vocabulary growth.

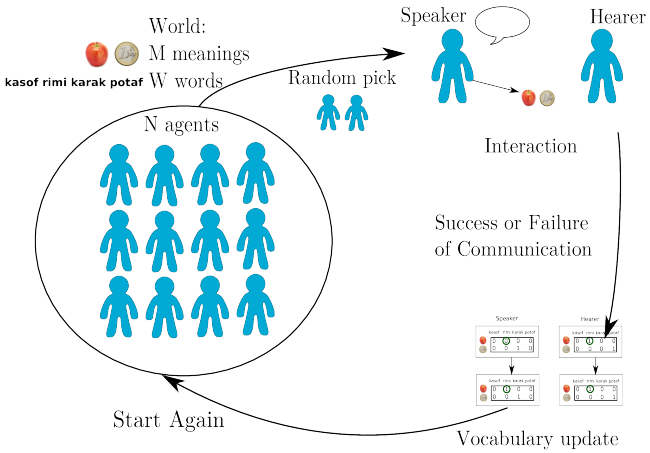


Figure 12. Illustration of the Naming Game model. Through repeated pairwise interactions, a population of individuals agrees on a shared lexicon mapping words to meanings.

7.1.3.1. Active topic choice strategies

Usually, the topic used in an interaction of the Naming Game is picked randomly. A first way of introducing active control of complexity growth is through the mechanism of topic choice: choosing it according to past memory. It allows each agent to balance reinforcement of known associations and invention of new ones, which can be seen as an exploitation vs. exploration problem. This can speed up convergence processes, and even lower significantly local and global complexity: for example in [140], [141], where heuristics based on the number of past successful interactions were used.

We defined new strategies in [31], [14] based on a maximization of an internal measure called LAPS, or Local Approximated Probability of Success. The derived strategies are called LAPSmax (exact measure but heuristical optimization algorithm) and Coherence (simplified measure but exact optimization).

Those strategies can speed up convergence the convergence process, but also diminish significantly the local complexity – i.e. the maximum number of distinct word-meaning association present in the population. See figure 13 .

7.1.3.2. Statistical lower bounds and performance measures

We showed that the time needed to converge to a shared lexicon admits a statistical lower bound [14]:

$$t_{conv} \geq N \cdot M \cdot \ln M \tag{1}$$

Where M is the number of meanings and N the population size.

Using this lower bound, we can define performance measures (between 0 and 1, best value being 1) to classify strategies and compare behavior for different values of the parameters (like population size). We distinguish in particular performance measures for convergence time, convergence speed, and maximum lexicon size. Using this, we can show that LAPSmax and Coherence yield good performance measures, which are stable with population size (cf fig. 14), and significantly better than previous strategies.

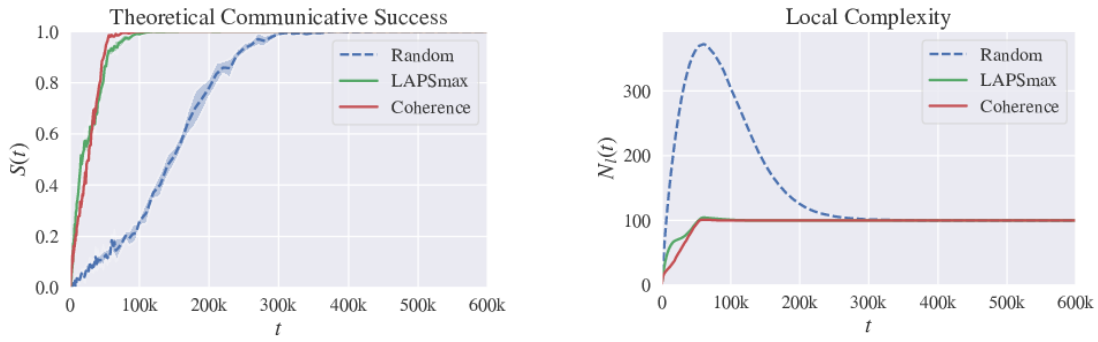


Figure 13. Evolution through time (number of interactions) of the measures of convergence (global probability of success) and global complexity (number of distinct word-meaning association present in the population) for simulations using Random Topic Choice vs. LAPSmax and Coherence Topic Choice strategies. The active topic choice strategy yields faster convergence, with less complexity. $N = 100$, $M = 100$, W *is unbounded.*

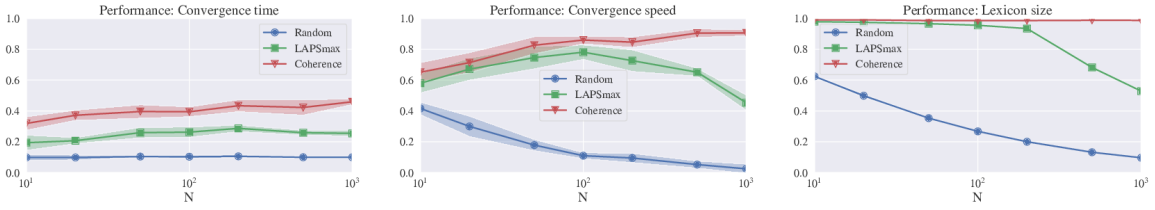


Figure 14. Performance measures for LAPSmax and Coherence strategies, compared with Random Topic Choice, for different values of N *(population size).* $M = 100$, W *is unbounded.*

7.1.3.3. Interactive application for collaborative creation of a language: Experimenting how humans actively negotiate new linguistic conventions

How do humans agree and negotiate linguistic conventions? This question is at the root of the domain of experimental semiotics [80], which is the context of our experiment/application. Typically, the experiments of this field consist in making human subjects play a game where they have to learn how to interact/collaborate through a new unknown communication medium (such as abstract symbols). In recent years, such experiments allowed to see how new conventions could be formed and evolve in population of individuals, shading light on the origins and evolution of languages [94], [79].

We consider a version of the Naming Game [152], [102], focusing on the influence of active learning/teaching mechanisms on the global dynamics. In particular, agreement is reached sooner when agents actively choose the topic of each interaction [121], [140], [141].

Through this experiment, we confront existing topic choice algorithms to actual human behavior. Participants interact through the mediation of a controlled communication system – a web application – by choosing words to refer to objects. Similar experiments have been conducted in previous work to study the agreement dynamics on a name for a single picture [63]. Here, we make several pictures or interaction topics available, and quantify the extent to which participants actively choose topics in their interactions.

Global description: Each user interacts for about 3-4 min (<30 interactions) with a brand new population of 4 simulated agents. They take the role of one designated agent, and play the Naming Game as this agent. Each time they interact as speakers, they can select the topics of conversation from a set of 5 objects, and are offered 6 possible words to refer to them. Their choices influence the global emergence of a common lexical convention, reached when communications are successful. The goal is to maximize a score based on the number of successful interactions (among the 50 in total for each run). They can see a list of the past interactions, with chosen topic, chosen word, and whether the interaction was successful or not. This experiment allows us to directly measure if there is a bias in the choice of topics, compared to random choice, based on memory of past interactions. Performance can then be compared to existing topic choice algorithms [121], [140], [141] and [31].

First version: A first version was developed for the Kreyon Conference in Rome, in September 2017. The experiment was however too close to the theoretical model, and users were not motivated to play and finish the experiment. Provided feedback was often perceived as frustrating.

Second version: A second version was developed with the help of Sandy Manolios. This second version is more entertaining, includes a motivating context, a backstory, more adapted feedback, and a more user-friendly visual interface.

Results: Users in both experiments seem to actively control their rate of invention of new conventions, by selecting more often (than random) objects that they already have a word for. See figure 16 .

7.2. Autonomous Learning in Artificial Intelligence

7.2.1. Intrinsically Motivated Goal Exploration and Multi-Task Reinforcement Learning

Participants: Sébastien Forestier, Pierre-Yves Oudeyer [correspondant], Alexandre Péré, Olivier Sigaud, Cédric Colas, Adrien Laversanne-Finot, Rémy Portelas.

7.2.1.1. Intrinsically Motivated Exploration of Spaces of Parameterized Goals and Application to Robot Tool Learning

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 [52].

Figure 15. Examples with the interface of respectively the first and the second version. Play the game here: http://naming-game.space

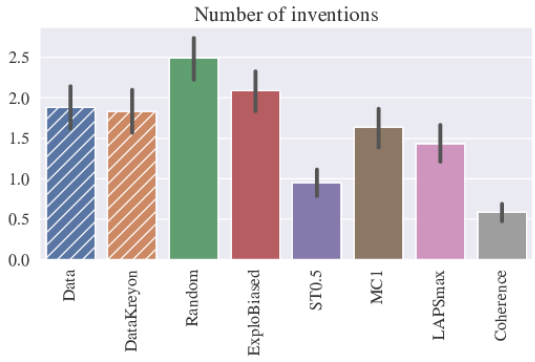


Figure 16. Average number of inventions for both experiments. DataKreyon: data collected at the Kreyon Conference in 2017 (first version); Data: data collected during summer 2018 (second version version); Other: simulated data for various strategies.

Last year we developed a formal framework called “Unsupervised Multi-Goal Reinforcement Learning”, as well as a formalization of intrinsically motivated goal exploration processes (IMGEPs), that is both more compact and more general than our previous models [76]. We experimented several implementations of these processes in a complex robotic setup with multiple objects (see Fig. 17), 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.

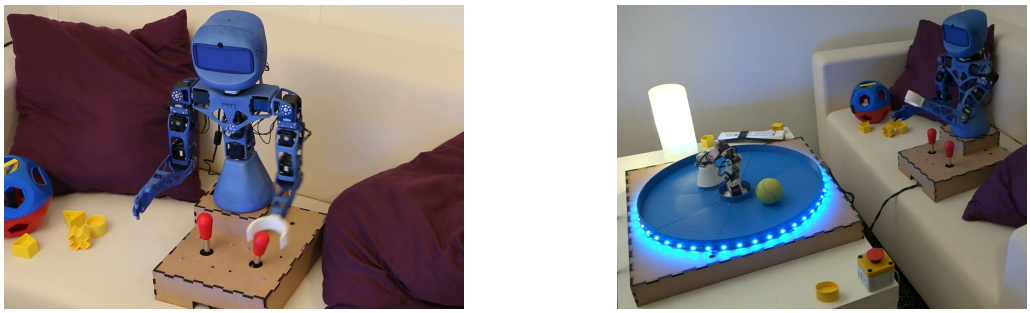


Figure 17. Robotic setup. Left: a Poppy Torso robot (the learning agent) is mounted in front of two joysticks. Right: full setup: a Poppy Ergo robot (seen as a robotic toy) is controlled by the right joystick and can hit a tennis ball in the arena which changes some lights and sounds.

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. 18). 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.

7.2.1.2. Leveraging the Malmo Minecraft platform to study IMGEP in rich simulations

In 2018 we started 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 [19]). 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 Deep Learning of 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

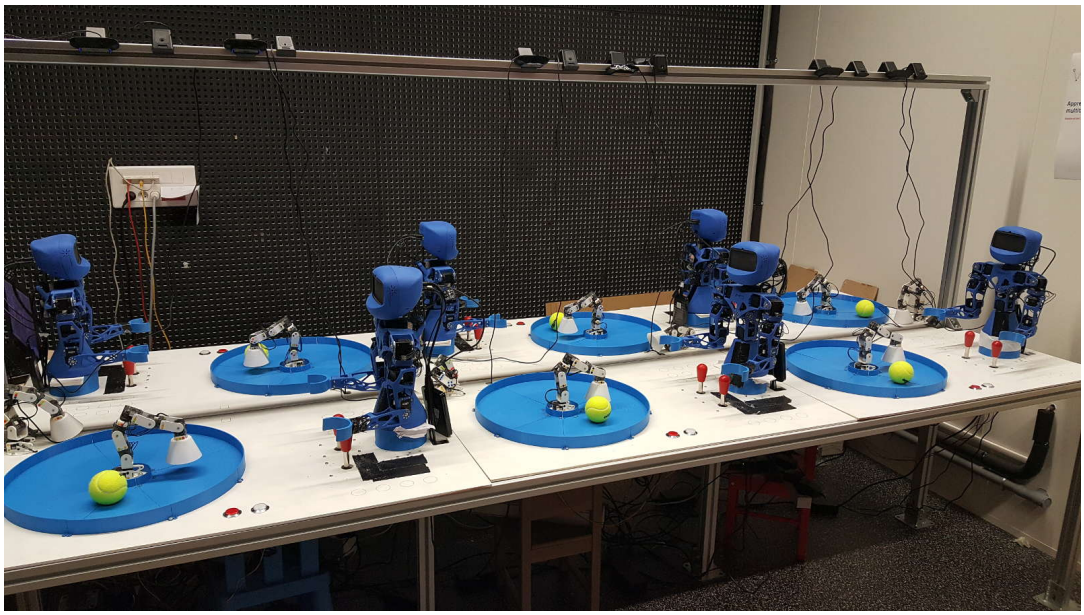


Figure 18. Platform of 6 robots with identical environment: joysticks, Poppy Ergo, ball in an arena, and a distractor. The central bar supports the 12 top cameras.

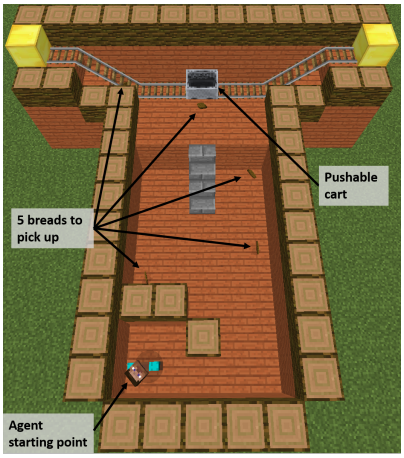


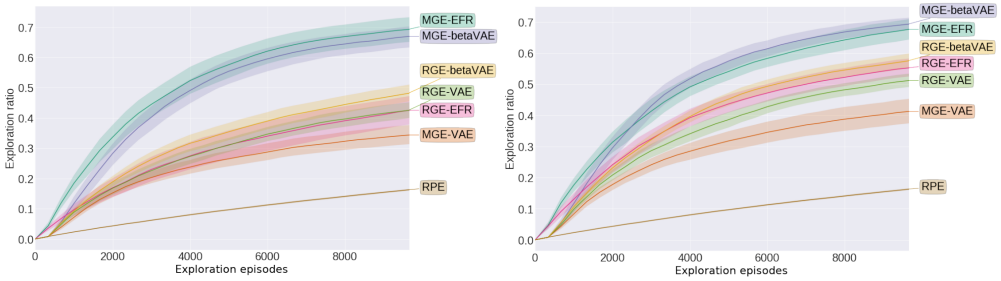
Figure 19. Malmo Mountain Cart. In this episodic environment the agent starts at the bottom left corner of the arena and is able to act on the environment using 2 continuous action commands: move and strafe. If the agent manages to get out of its starting area it may be able to collect items dispatched 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.

spaces. 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 [38].

7.2.1.4. Curiosity Driven Exploration of Learned Disentangled Goal Spaces

As described in the previous paragraph, we have shown in [38] that one can use deep representation learning algorithms to learn an adequate goal space in simple environments. 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 [28].

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.



(a) Small exploration noise ($\sigma = 0.05$) (b) Large exploration noise ($\sigma = 0.1$)

Figure 20. 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).

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 20). 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.

7.2.1.5. Combining deep reinforcement learning and curiosity-driven exploration

A major challenge of autonomous robot learning is to design efficient algorithms to learn sensorimotor skills in complex and high-dimensional continuous spaces. Deep reinforcement learning (RL) algorithms are natural candidates in this context, because they can be adapted to the problem of learning continuous control policies with low sample complexity. However, these algorithms, such as DDPG [97] suffer from exploration issues in the context of sparse or deceptive reward signals.

In this project, we investigate how to integrate deep reinforcement learning algorithms with curiosity-driven exploration methods. A key idea consists in decorrelating the exploration stage from the policy learning stage by using a memory structure used in deep RL called a replay buffer. Curiosity-driven exploration algorithms, also called Goal Exploration Processes (GEPs) are used in a first stage to efficiently explore the state and action space of the problem, and the corresponding data is stored into a replay buffer. Then a DDPG learns a control policy from the content of this replay buffer.

Last year, an internship has been dedicated to trying this methodology in practice but did not reach conclusions because of instability issues related to DDPG. The project was restarted this year, which led to an ICML publication [25]. We used an open-source implementations [72], and compared the combination GEP-PG (GEP + DDPG) to the traditional DDPG [97] as well as the former state-of-the art DDPG implementation endowed with parameter-based exploration [131]. Results were presented on the popular OpenAI Gym benchmarks Continuous Mountain Car (CMC) and Half-Cheetah (HC) [58], where state-of-the-art results were demonstrated.

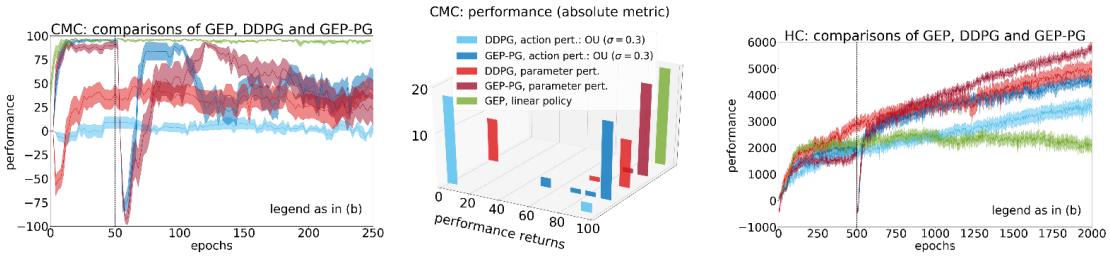


Figure 21. Left: learning curves on Continuous Mountain Car (20 seeds, mean +/- sem). Middle, best performances reached across learning on CMC. Right, learning curves on Half-Cheetah (20 seeds, mean +/- sem).

7.2.1.6. Monolithic Intrinsically Motivated Multi-Goal and Multi-Task Reinforcement Learning

In this project we merged two families of algorithms. The first family is the Intrinsically Motivated Goal Exploration Processes (IMGEP) developped in the team (see [77] for a description of the framework). 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.

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) [137] 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 [49] 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 [105] proposed to use UVFA to achieve multi-task learning with a discrete task-set.

In this project, we developed CURIOUS [43], 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 [58], see Figure 22 . 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 23 .

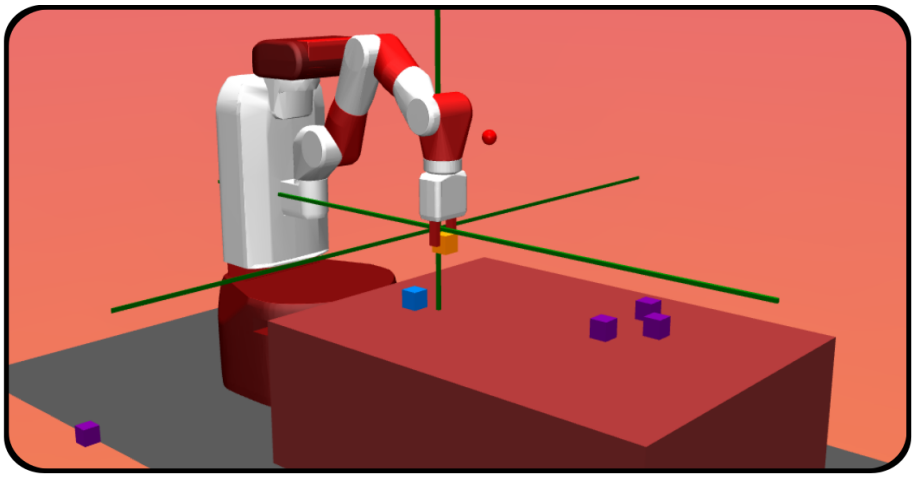


Figure 22. Custom multi-task and multi-goal environment to test the CURIOUS algorithm.

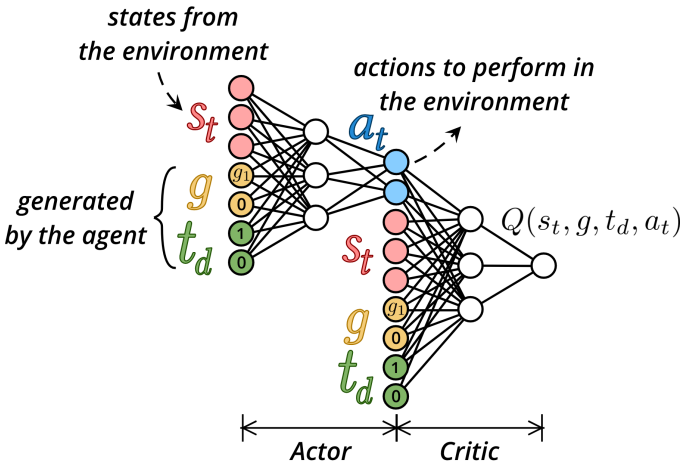


Figure 23. Architecture extended from Universal Value Function Approximators. In this example, the agent is targeting task T_1 among two tasks, each corresponding to a 1 dimension goal.

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. 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 24 .

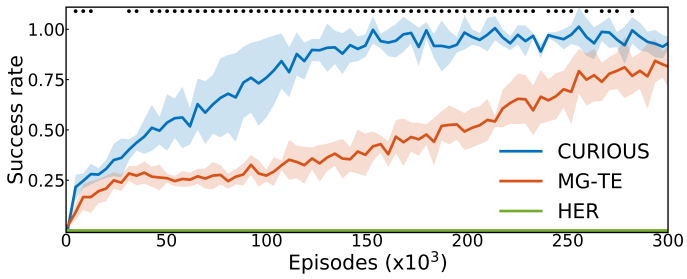


Figure 24. Comparison of CURIOUS to alternative algorithms.

7.2.2. Transfer Learning from Simulated to Real World Robotic Setups with Neural-Augmented Simulators

Participants: Florian Golemo [correspondant], Pierre-Yves Oudeyer.

This work was made in collaboration with Adrien Ali Taiga and Aaron Courville, and presented at CoRL 2018 [27]. Reinforcement learning with function approximation has demonstrated remarkable performance in recent years. Prominent examples include playing Atari games from raw pixels, learning complex policies for continuous control, or surpassing human performance on the game of Go. However most of these successes were achieved in non-physical environments (simulations, video games, etc.). Learning complex policies on physical systems remains an open challenge. Typical reinforcement learning methods require a lot of data which makes it unsuitable to learn a policy on a physical system like a robot, especially for dynamic tasks like throwing or catching a ball. One approach to this problem is to use simulation to learn control policies before applying them in the real world. This raises new problems as the discrepancies between simulation and real world environments ("reality gap") prevent policies trained in simulation from performing well when transfered to the real world. This is an instance of *domain adaption* where the input distribution of a model changes between training (in simulation) and testing (in real environment). The focus of this work is in settings where resetting the environment frequently in order to learn a policy directly in the real environment is highly impractical. In these settings the policy has to be learned entirely in simulation but is evaluated in the real environment, as *zero-shot transfer*.

In simulation there are differences in physical properties (like torques, link weights, noise, or friction) and in control of the agent, specifically joint control in robots. We propose to compensate for both of these source of issues with a generative model to bridge the gap between the source and target domain. By using data collected in the target domain through task-independent exploration we train our model to map state transitions from the source domain to state transition in the target domain. This allows us to improve the quality of our simulated robot by grounding its trajectories in realistic ones. With this learned transformation of simulated trajectories we are able to run an arbitrary RL algorithm on this augmented simulator and transfer the learned policy directly to the target task. We evaluated our approach in several OpenAI gym environments that were modified to allow for drastic torque and link length differences.

7.2.3. Curiosity-driven Learning for Automated Discovery of Physico-Chemical Structures

Participants: Chris Reinke [correspondant], Pierre-Yves Oudeyer.

Intrinsically motivated goal exploration algorithms 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 in high-dimensional continuous state and action spaces (e.g. [75], [53]). 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 (e.g. [132]). 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 (e.g. [38]), but they have focused on goals represented as static target configurations of the environment in robotics sensorimotor spaces. This project studies how these new families of machine learning algorithms can be extended and used for automated discovery of behaviours of dynamical systems in physics/chemistry.

The work on the project started in November 2018 and is currently in the state of exploring potential systems and algorithms.

7.2.4. Statistical Comparison of RL Algorithms.

In this project [42], we review the statistical tests recommended by [87] to compare RL algorithms (T-test, bootstrap test, Kolmogorov-Smirnov). Kolmogorov-test is discarded as it only allows to compare distributions and not mean or median performance. We wrote a tutorial in the form of an arxiv paper [42] to present the use of the Welch's t-test and the bootstrap test to compare algorithm performances. In the last section of that paper, initial assumptions of the test are described. In particular, two limiting points are discussed. First, the computation of the required sample size to satisfy requirements in type-II error (false negative) is highly sensitive to the estimations of the empirical standard deviations of the algorithms performance distributions. It is showed that for small sample sizes (< 20) the empirical standard deviation of a Gaussian distribution is biased negatively in average. Using an empirical standard deviation smaller than the true one results in under-estimations of the type-II error and therefore of the required sample size to meet requirement on that error. Second we propose empirical estimations of the type-I error (false positive) as a function of the sample size for the Welch's t-test and the bootstrap test for a particular example (Half-Cheetah environment from OpenAI Gym [58] using DDPG algorithm [97]). It is showed that the type-I error is largely underestimated by the bootstrap test for small sample size (x6 for $n = 2$, x2 for $n = 5$, x1.5 for $n = 20$). The Welch's t-test offers a satisfying estimation of the type-I error for all sample size. In conclusion, the bootstrap test should not be used. The Welch's t-test should be used with a sufficient number of seeds to obtain a reasonable estimation of the standard deviation so as to get an accurate measure of type-II error (N>20).

7.3. Representation Learning

7.3.1. State Representation Learning in the Context of Robotics

Participants: David Filliat [correspondant], Natalia Diaz Rodriguez, Timothee Lesort, Antonin Raffin, René Traoré, Ashley Hill.

During the DREAM project, we participated in the development of a conceptual framework of open-ended lifelong learning [20] 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 [23], and we have 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 have developed and open sourced[0] the S-RL toolbox [39] containing baseline algorithms, data generating environments, metrics and visualization tools for assessing SRL methods. The framework has been published at the NIPS workshop on Deep Reinforcement Learning 2018.

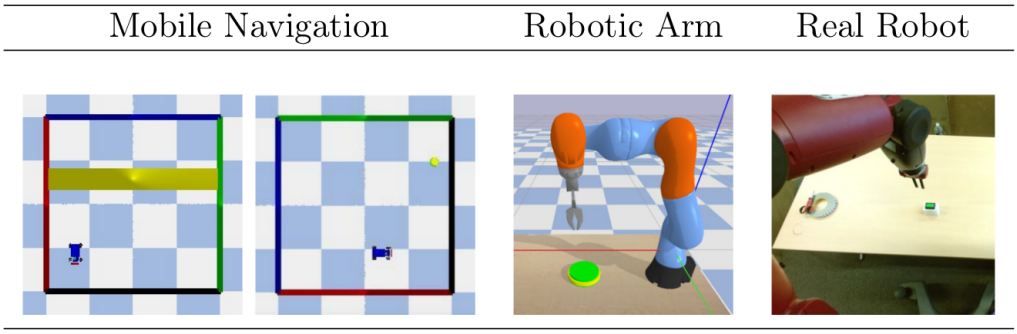


Figure 25. Environments and datasets for state representation learning.

The environments proposed in Fig. 25 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. 26 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).

We also proposed a new approach that consists in 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.

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. 27 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 [72], 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 [39] along with ground truth (GT).

[0] https://github.com/araffin/robotics-rl-srl

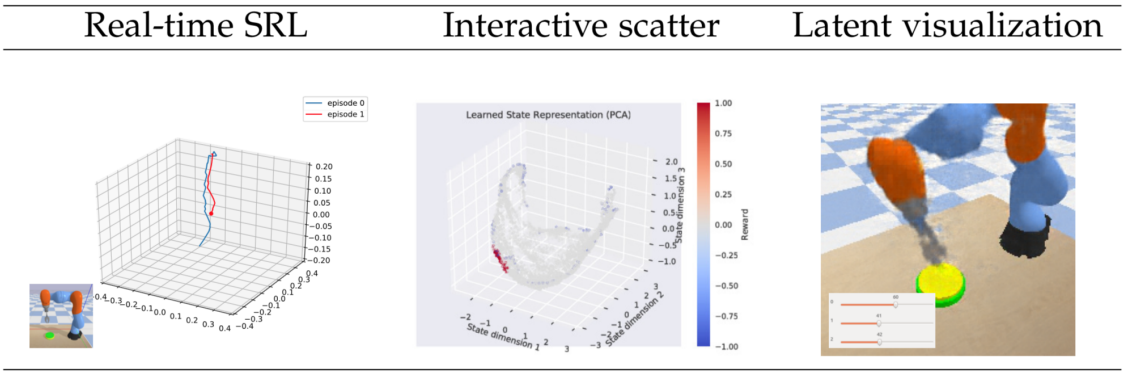


Figure 26. Visual tools for analysing SRL; Left: Live trajectory of the robot in the state space. Center: 3D scatter plot of a state space; clicking on any point displays the corresponding observation. Right: reconstruction of the point in the state space defined by the sliders.

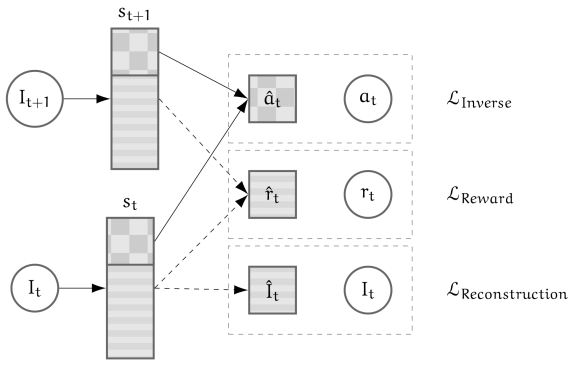


Figure 27. SRL Splits model: combines a reconstruction of an image I, a reward (r) prediction and an inverse dynamic models losses, using two splits of the state representation s. Arrows represent model learning and inference, dashed frames represent losses computation, rectangles are state representations, circles are real observed data, and squares are model predictions.

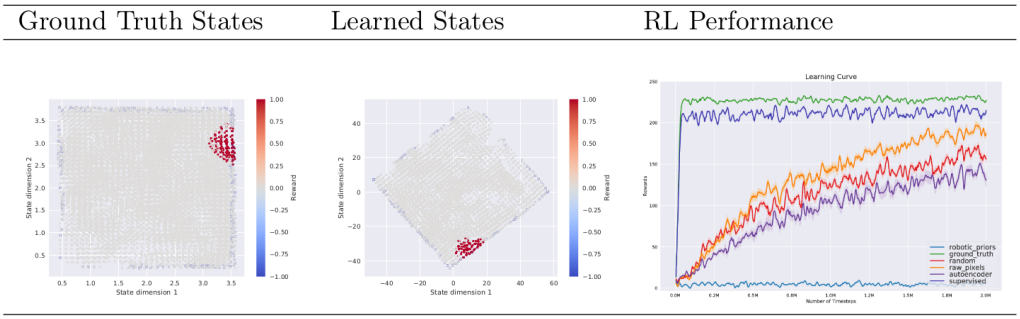


Figure 28. Ground truth states (left), states learned (Inverse and Forward) (center), and RL performance evaluation (PPO) (right) for different baselines in the mobile robot environment. Colour denotes the reward, red for positive, blue for negative and grey for null reward (left and center).

Table 28 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 [39].

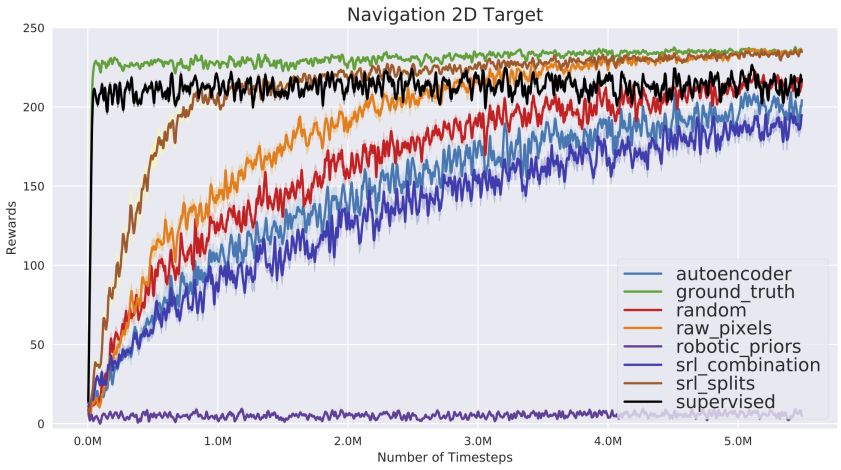


Figure 29. Performance (mean and standard error for 10 runs) for PPO algorithm for different state representations learned in Navigation 2D random target environment.

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. 29 shows that our unsupervised state representation learned with the split model even improves on the supervised case.

7.3.2. 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 [108], 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 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 [34] 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 30 .

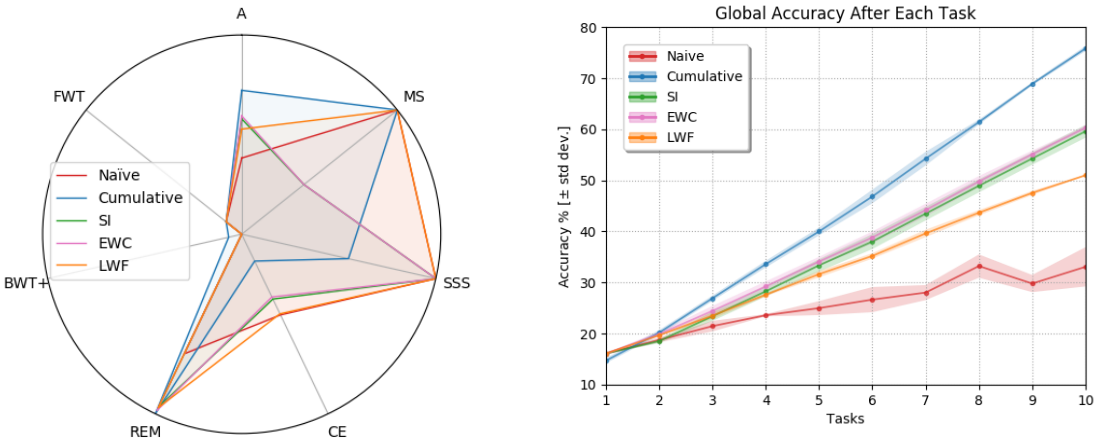


Figure 30. (left) Spider chart: CL metrics per strategy (larger area is better) and (right) Accuracy per CL strategy computed over the fixed test set.

Generative models can also be evaluated from the perspective of Continual learning.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 32) 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 [33]. 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

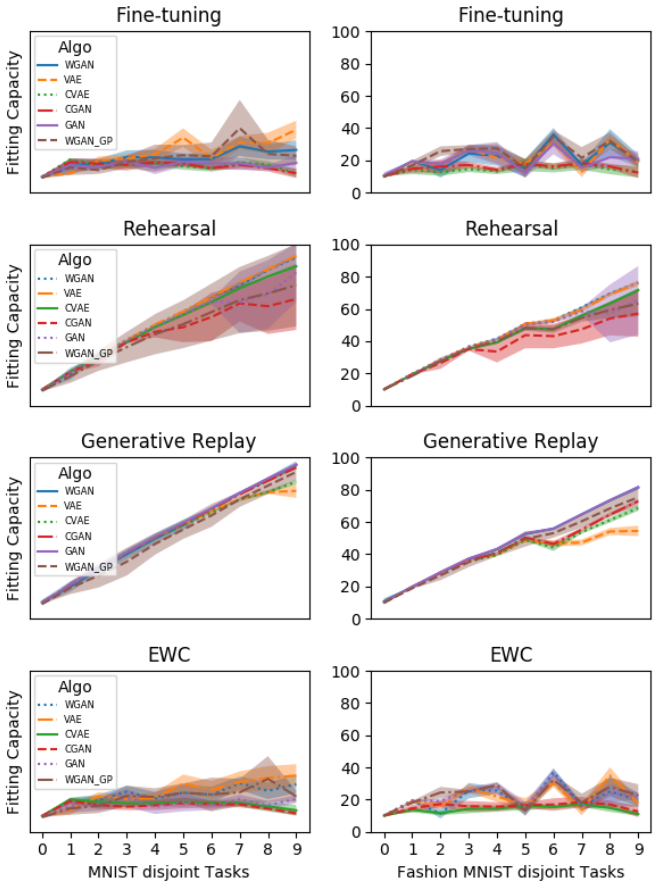


Figure 31. Means and standard deviations over 8 seeds of Fitting Capacity metric evaluation of VAE, CVAE, GAN, CGAN and WGAN. The four considered CL strategies are: Fine Tuning, Generative Replay, Rehearsal and EWC. The setting is 10 disjoint tasks on MNIST and Fashion MNIST.

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 32 . 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 [32]. 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, 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.

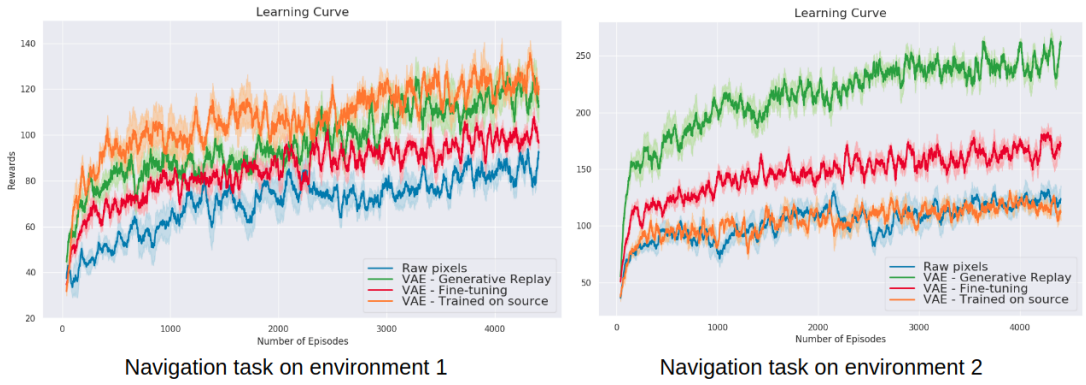


Figure 32. Mean reward and standard error over 5 runs of RL evaluation using PPO with different types of inputs. Fine-tuning and Generative Replay models are trained sequentially on the first and second environment, and then used to train a policy for both tasks. Generative Replay outperforms all other methods. It shows the need for continually learning features in State Representation Learning in settings where the environment changes.

Real life examples of applications envisioned for continual learning include learning on the edge, real time embedded systems, and applications such as the project proposal at the NeurIPS workshop on AI for Good on intelligent drone swarms for search and rescue operations at sea [36].

7.3.3. Knowledge engineering tools for neural-symbolic learning

Participant: Natalia Díaz Rodríguez [correspondant].

This section includes diverse partners distributed world wide and is result of former stablished collaborations and includes work in the context of knowledge engineering for neural-symbolic learning and reasoning systems. In [35] we presented *Datil*, a tool for learning fuzzy ontology datatypes based on clustering techniques and fuzzyDL reasoner. Ontologies for modelling healthcare data aggregation as well as knowledge graphs for modelling influence in the fashion domain are concrete ontological proposals for knowledge modelling. The former looks at wearables data interoperability for Ambient Assisted Living application development, including concepts such as height, weight, locations, activities, activity levels, activity energy expenditure, heart rate, or stress levels, among others [41]. The second proposal, considers the intrinsic subjectivity needed to effectively model subjective domains such as fashion in recommendations systems. Subjective influence networks are proposed to better quantify influence and novelty in networks. A set of use cases this approach is intended to address is discussed, as well as possible classes of prediction questions and machine learning experiments that could be executed to validate or refute the model [40].

7.4. Applications in Robotic myoelectric prostheses

Participant: Pierre-Yves Oudeyer [correspondant].

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.

7.4.1. 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.

7.5. Applications in Educational Technologies

7.5.1. Machine Learning for Adaptive Personalization in Intelligent Tutoring Systems

Participants: Benjamin Clement, Alexandra Delmas, Pierre-Yves Oudeyer [correspondant], Didier Roy, Helene Sauzeon.

7.5.1.1. 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 [65]. Systematic experiments were also presented with simulated students.

7.5.1.2. Kidlearn Experiments in 2018: Evaluating the impact of ZPDES and choice on learning efficiency and motivation

An experiment was held between mars 2018 and July 2018 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. The details of these new results, as well as the overall results of this project, were presented during the PhD defense of Benjamin Clement in decembre 2018.

7.5.1.3. Fostering Health Education With a Serious Game in Children With Asthma: Pilot Studies for Assessing Learning Efficacy and Automatized Learning Personalization

Coupled with Health Education programs, an e-learning platform—KidBreath—was participatory designed [19] and assessed in situ (Study 1) and was augmented and tested with an Intelligent Tutoring System (ITS) based on Multi-Armed Bandit Methods (Study 2). For each study, the impact of KidBreath practice was assessed in children with asthma in terms of pedagogical efficacy (knowledge of the illness), pedagogical efficiency (usability, type of motivation and level of interest elicited), and therapeutic effect (illness perception, system's expectation and judgement in disease self-management, child's implication in study). For the Study 1, asthma children aged 8 to 11 years used the tool at home without time pressure for 2 months according to a predefined learning sequence defined by the research team. Results supported pedagogical efficacy of KidBreath, with a significant increase of general knowledge about asthma after use. It also featured a greater learning gain for children knowing the least about the illness before use. Results on pedagogical efficiency revealed a great intrinsic motivation elicited by KidBreath showing a deep level of interest in the edutainment activities. Study 2 explored an augmented version of KidBreath with learning optimization algorithm (called ZPDES) after its use during 1 month. Pedagogical efficacy was less conclusive than Study 1 because less content was displayed due to algorithm parameters. However, the ITS-augmented KidBreath use showed a strong impact in pedagogical efficiency and therapeutic adherence features. Even if implementation improvements must be done in future works, this preliminary study highlighted the viability of our methods to design an ITS as serious game in health education context for all chronic diseases.

- Journée EdTech, Talence, mai 2018

7.5.2. Poppy Education: Designing and Evaluating Educational Robotics Kits

Participants: Pierre-Yves Oudeyer [correspondant], Didier Roy, Thibault Desprez, Théo Segonds, Stéphanie Noirpoudre.

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 33)

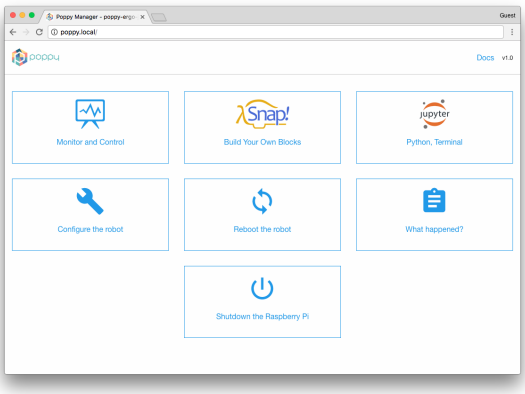


Figure 33. Home page on http://poppy.local

- 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 34), which is a variant of Scratch. Its features allow a thorough introduction of information technology. Several specific "blocks" have been developed for this.

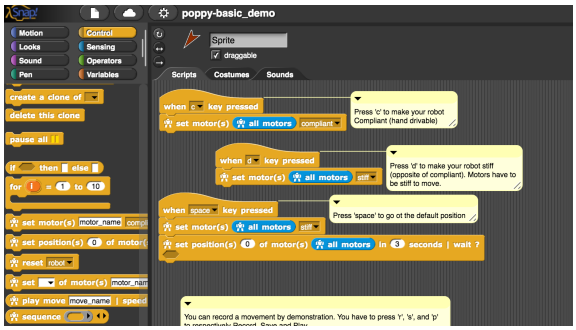


Figure 34. The visual programming system Snap

- 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 35). 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 36).

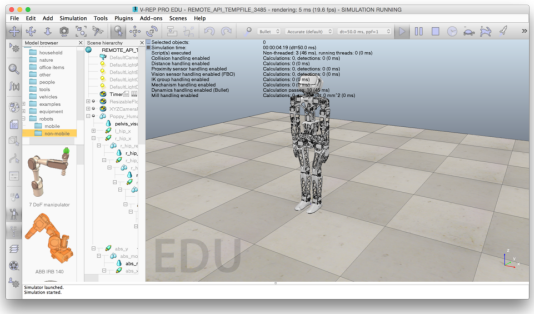


Figure 35. V-rep

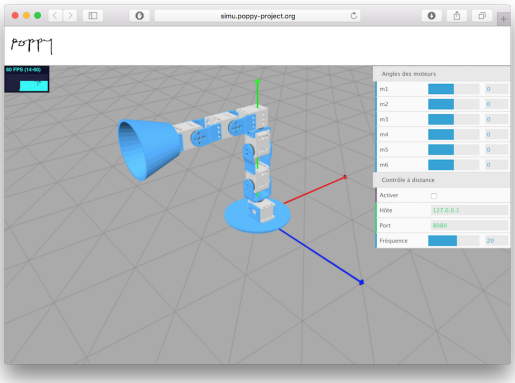


Figure 36. 3D viewer

7.5.2.1. 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 [120].

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 37. 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

7.5.2.2. 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 [119] 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

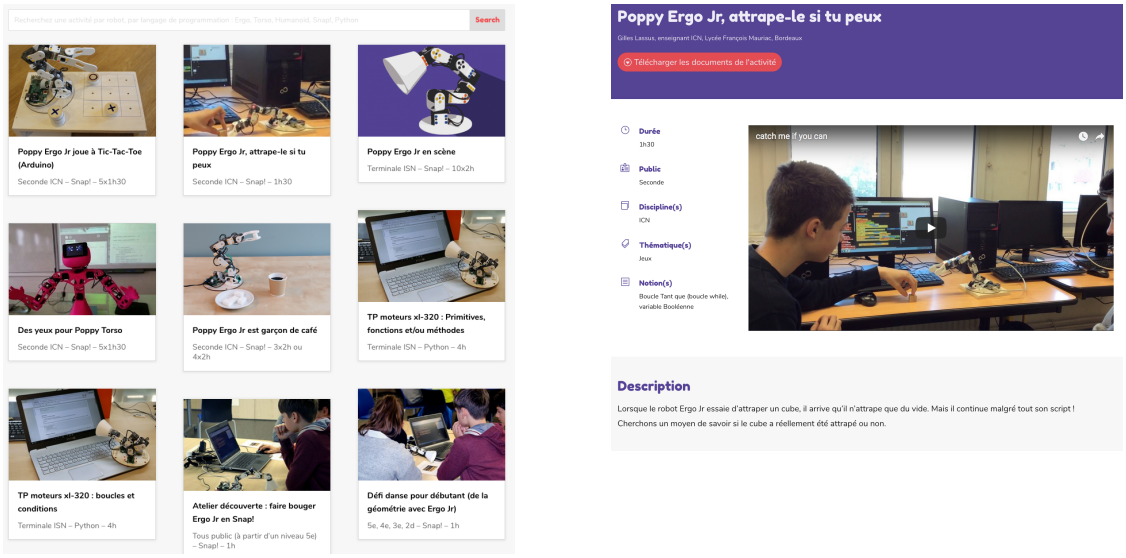


Figure 38. Open-source educational activities with Poppy robots are available on Poppy-Education.org

- 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/

7.5.2.3. 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) [59] and The AttrakDiff [96]. 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) [26] published at the conference Didapro (Lausanne Feb, 2018). Figures 39 and 40 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.

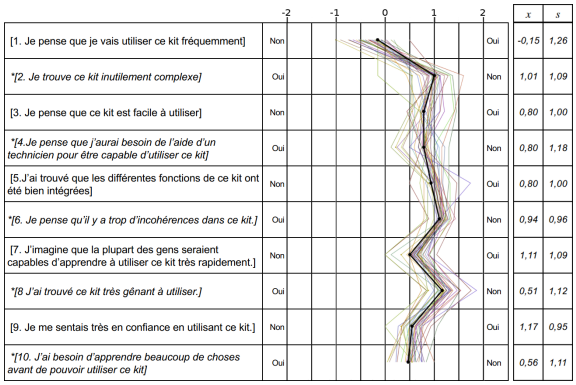


Figure 39. Result of SUS survey

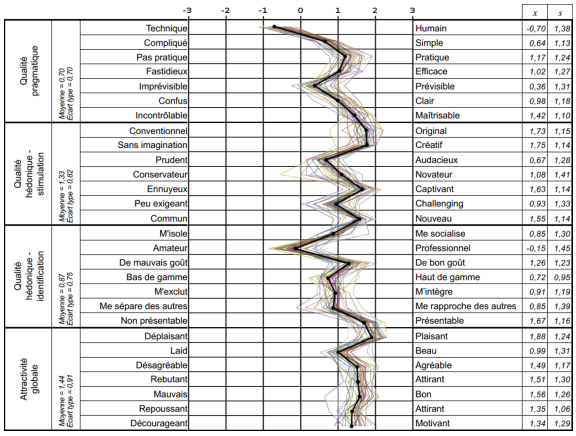


Figure 40. Result of AttrakDiff survey

- 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/

7.5.2.4. 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.

7.5.3. IniRobot: Educational Robotics in Primary Schools

Participants: Didier Roy [correspondant], Pierre-Yves Oudeyer.

Reminder : IniRobot (a project done in collaboration with EPFL/Mobsya) aims to create, evaluate and disseminate a pedagogical kit which uses Thymio robot, an open-source and low cost robot, for teaching computer science and robotics.

IniRobot Project aims to produce and diffuse a pedagogical kit for teachers and animators, to help them and to train them directly or by the way of external structures. The aim of the kit is to initiate children to computer science and robotics. The kit provides a micro-world for learning, and takes an inquiry-based educational approach, where kids are led to construct their understanding through practicing an active investigation methodology within teams. See https://dm1r.inria.fr/c/kits-pedagogiques/inirobot or http://www.inirobot.fr.

Deployment: After 4 years of activity, IniRobot is used by more than 3000 adults, 30 000 children in France. Inirobot is also used in higher education, for example in Master 2 "Neurosciences, human and animal cognition" at the Paul Sabatier University in Toulouse. Inirobot is additionally used to train the management and elected officials of the Bordeaux metropolitan area (20 people). The digital mediators of the 8 Inria centers are trained to Inirobot and use it in their activities.

7.5.3.1. Partnership

The project continues to be carried out in main collaboration with the LSRO Laboratory from EPFL (Lausanne) and others collaborations such as the French National Education/Rectorat d'Aquitaine, the Canopé Educational Network, the ESPE (teacher's school) Aquitaine, the ESPE Martinique, the ESPE Poitiers and the National Directorate of Digital Education.

7.5.3.2. Created pedagogical documents and resources

- The inirobot pedagogical kit [83]: This pedagogical booklet provides activities scenarized as missions to do. An updated version of the Inirobot pedagogical kit is available at: https://dm1r.inria.fr/uploads/default/original/1X/70037bdd5c290e48c7ec4cb4f26f0e426a4b4cf6.pdf. Another pedagogical booklet has been also created by three pedagogical advisers for primary school, with pedagogical instructions and aims, under our supervision. The new pedagogical kit,"Inirobot Scolaire, Langages et robotique", which extends Inirobot to a full primary school approach is available at http://tice33.ac-bordeaux.fr/Ecolien/ASTEP/tabid/5953/language/fr-FR/Default.aspx
- Inirobot website and forum: https://dm1r.inria.fr/c/kits-pedagogiques/inirobot or http://www.inirobot.fr On this website, teachers, animators and general public can download documents, exchange about their use of inirobot's kit.

7.5.3.3. Scientific mediation

Inirobot is very popular and often presented in events (conferences, workshops, ...) by us and others.

7.5.3.4. Spread of Inirobot activities

Inirobot activities are used by several projects: Dossier 123 codez from Main à la Pâte Fundation, Classcode project, ...

7.5.3.5. MOOC Thymio

The MOOC Thymio, released in october 2018, in collaboration with Inria Learning Lab and EPFL (Lausanne, Switzerland), on FUN platform and edX EPFL Platform), use Inirobot activities to teach how to use Thymio robot in education.

HEPHAISTOS Project-Team

7. New Results

7.1. Robotics

7.1.1. Analysis of Cable-driven parallel robots

Participants: Alain Coulbois, Artem Melnyk, 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 [19]. 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 [15] (best paper award of the Eucomes conference) 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. For that purpose it is of interest to add sensors on the robot beside the measurement of cable lengths in order to improve the solving time by using additional constraints and possibly ending up with a single solution. But these measurements are uncertain although we may assume that the measurement errors are bounded. It is necessary to determine these error bounds for a practical use of these measurement and we have conducted an experimental investigation of various additional measurements [12]: a mechanical system for measuring the angle of the cable plane with respect to a reference axis, cable angulation with accelerometers glued on the cable, a “poor man lidar” on the platform for optically determining several cables angulation, accelerometers on the platform and cable tensions with strain gauges while the pose of the platform was estimated accurately by using a metrology arm and laser range-meters. This investigation has shown that:

- the friction in the mechanical system leads to large errors for the cable plane angle (up to 30 degrees). For later measure we have bypassed this system
- even for small and medium-sized CDPRs the sagging effect cannot be neglected for estimating cable angulation
- accelerometers on the cable and the lidar system have a good accuracy (between 1 and 5 degrees)
- cable tension measurement is very approximate even with high accuracy strain gauges and cannot be used for control purposes.

We have also continued to investigate calculation of planar cross-sections of the workspace for CDPR with sagging cables, i.e. when 4 of the 6 platform pose parameters are fixed leaving only 2 free parameters. Brand new algorithms have been developed, based on a continuation approach [12],[13]. The main idea is that almost everywhere the workspace border is a one-dimensional variety so that if one of the free parameters is fixed, then a pose on the border should satisfy a square equation system constituted of the kinematic equations and the constraints equations (e.g. that a cable length is equal to a given maximum limit). Pose on the border are obtained by choosing an arbitrary pose that has an inverse kinematic solution that satisfy the constraints in the workspace and then moves incrementally along one of the free axis using a certified Newton scheme for finding the inverse kinematics solution until the constraint equations are almost satisfied in which case the certified Newton scheme is used to determine exactly (i.e with an arbitrary accuracy) a pose that lies on the border. Then a continuation scheme is used to find new poses on the border until we reach a pose at which a new set of constraints is satisfied i.e. a starting point for a new border arc. The border is then composed of several polygonal arcs that approximate the real border. The scheme is devised so that we completely master the difference between the real workspace area and the region defined by the polygonal approximation of the border. If necessary we may reduce this difference by adding new vertices on the border polygon. An important point is that the constraints define border arcs but also singularity curves (i.e. pose at which the direct kinematics equations are singular) and a specific continuation scheme has been developed to determine those arcs. Indeed the cancellation of the determinant of the jacobian of the direct kinematic equations is part of the equations that are satisfied on this type of border arc but this determinant cannot be obtained in closed-form. Consequently we have devised a certified Newton scheme that just require to evaluate the determinant and its derivatives at a given pose. A consequence of the existences of such arcs is that the workspace may have several *aspects* i.e. workspace region that can be reached only for a given inverse kinematics solution and is unreachable for the other one(s).

7.1.2. 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 this year by artists interested in mimicking the 3D additive manufacturing process on a large scale with glass micro-beads for a live art performance to be held in 2019 (www.lestanneries.fr/exposition/monuments-larmes-prince). We have been working on the design of the robotics system, namely a cables parallel robots with autonomous refilling capabilities.

7.1.3. Robotized ultrasound probe

Participant: Jean-Pierre Merlet.

In collaboration with the EPIONE project we have started investigation the development of a portable robotized cardiac ultrasound probe that may be used while performing an effort test. A first step, somewhat surprising was the necessity to instrument an existing probe in order to determine what are the forces that the doctor exert on the probe during an investigation and the maximal angulation of the probe (apparently this data has not been measured beforehand). We add an accelerometer (for measuring the angle) and a force sensor in a 3D-printed covering of the probe and recorded the data during several experiments. We were then planing to develop a small, portable 3 d.o.f. rotational parallel robot whose range of motion was within the maximum angles that has been determined experimentally and was able to sustain the force exerted by the doctor. Unfortunately there was not a general consensus between the doctors and the company manufacturing the probe on the number of d.o.f. that was requested for the robot (which clearly have a drastic influence on the mechanical design and on the dimensional synthesis of the robot) so that the project is on stand-by.

7.1.4. Parallel robot performances and uncertainties

Participants: Jean-Pierre Merlet, Hiparco Lins Vieira [correspondant].

The purpose of this study, which is the PhD subject of H. Lins Vieira, is to develop interval analysis-based algorithm for determining if some performance requirements for parallel robots (e.g. on workspace, accuracy, load lifting ability) can be guaranteed in spite of the unavoidable manufacturing and control uncertainties of the system.

7.2. Assistance

We are still going on in building a framework for customizable and modular assistive robotics including hardware, software and communication and medical monitoring. The development of our platforms shows that we are now able to identify problematic issues for end-users, helpers and the medical community and to propose appropriate hardware/software solutions. But the most time consuming part of our work is related to evaluation and therefore experimentation: this involves legal/ethical issues (for which we have contributed [5]), participation of the medical community (for evaluation and recruitment) and heavy administrative management. Clearly we are lacking of permanent staff as we have long term objectives that cannot be fulfilled only with PhD or post-doc students. We need also engineers during specific periods (for hardware development and experimentation) but over a longer time than the one or two years currently proposed by Inria.

7.2.1. Rehabilitation in an immersive environment

Participants: Artem Melnyk, Jean-Pierre Merlet, Yves Papegay [correspondant], Ting Wang.

Rehabilitation is a tedious and painful process and it is difficult to assess its trend. Using an immersive environment has shown to increase the patient motivation but is not sufficient regarding rehabilitation efficiency. First the visual feedback (event 3D) is not sufficient to provide a full immersive feeling as body motion is not involved. Controlling body motion is also very important for therapists that currently must continuously correct the patient pose so that the rehabilitation exercise is the most efficient. We propose to add motion generators in the environment to reinforce realism (thereby increasing patient motivation) but also to allow therapists to use these generators to control the body pose so that they will be able to repeat rehabilitation exercises in a controlled context. Furthermore these generators are instrumented to provide information on the body pose and additional external sensors complete these measurements for rehabilitation assessment. We have developed 3 types of motions generators: one 6 d.o.f. motion base, a CDPR that is able to lift a patient and 2 multipurpose lifting columns.

When starting this project we were planning to use Inria-Sophia immersive room , hence allowing us to focus on the rehabilitation station. Unfortunately this room is no more available. This year we have developed a 2D renderer that has been connected to a flexible software platform allowing the various agents to exchange messages. We have been able to build a first version of our rehabilitation platform using a treadmill as exercise tool and columns to animate the treadmill (figure 1). For measuring the gait pattern we are using a planar lidar for detecting the leg motions, a kinect for detecting the motion of a skeleton and a distance sensor that measure the body motion with respect to the head of the treadmill. Figures 2 and 3 show an extract of the measurements obtained during a typical walk. It may be seen that the lidar data are very clean and allows one to estimate the mean position of the leg as a function of time (from which we will be able to deduce the number of steps, velocities of the leg, ...). Kinect data are much more noisy although that a fusion with the lidar data and the distance data will allows us to detect significant trunk motion. A typical walk of 3mn provides approximately 20 Mo of data.

Note that we are not using wearable sensors (although they are available: accelerometers for the arms and legs, shoes with pressure sensor and accelerometers): this is voluntary as our contacts with the medical community have indicated that many patients will not be comfortable with wearable sensors. In the same manner we have experimented having a headset instead of the screen but it appears that visualization is very disturbing and uncomfortable. Subject safety is ensured: during the exercise the subject must keep a push button pressed and when released the treadmill stop immediately. An emergency stop button is also available for the operator. Furthermore the system has been designed to provide various supports for avoiding fall and is surrounded by soft carpets.

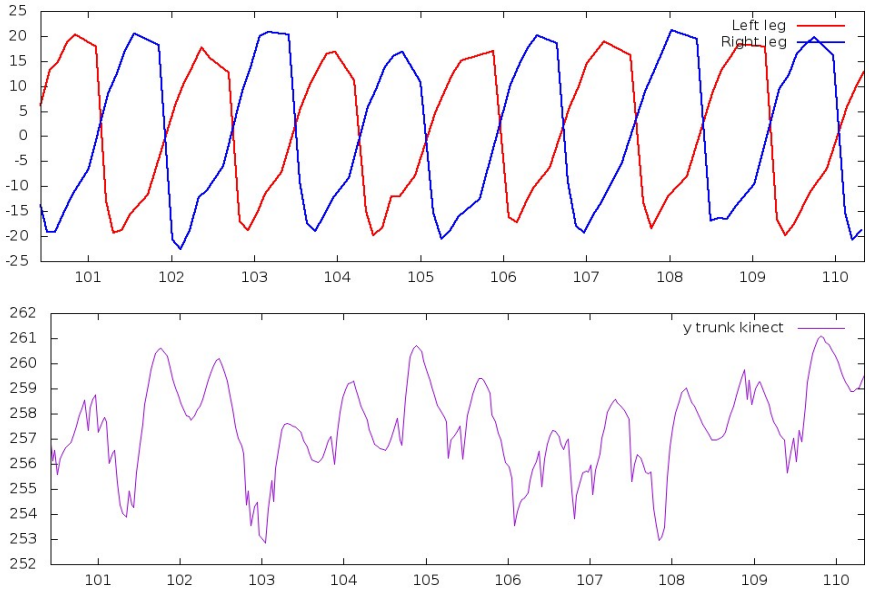


Figure 2. An extract of the legs motions in the walking direction as measured by the lidar and the trunk forward/backward motion estimated by the kinect

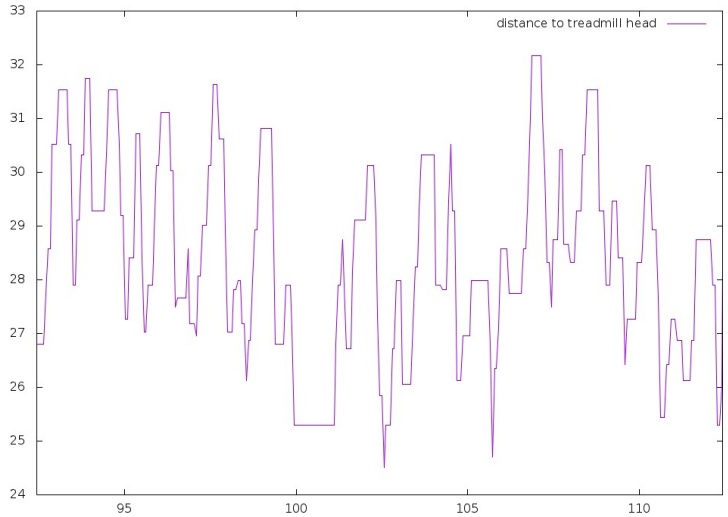


Figure 3. An extract of the trunk forward/backward motions as measured by the distance sensor in the head of the treadmill

The rehabilitation station for walking analysis on a treadmill in various walking condition is now almost fully functional and reliable. The next step will start of the beginning of 2019 with an experiment involving a cohort of voluntary subjects of Inria in order to obtain a significant amount of data. A statistical analysis of these data will then be performed in order to examine if synthetic and medically pertinent indicators (besides classical indicators such a number of steps, velocity, ...) may be obtained. The next step will involve repeating this experiment with pathological patients from Centre Héliomarin de Vallauris, most probably at the end of 2019. Meanwhile we will integrate our motion base as another element of the rehabilitation station with the purpose of equilibrium analysis, using a sea landscape as virtual environment with fans providing a realistic simulation of winds.

7.3. Smart Environment for Human Behaviour Recognition

Participants: Alain Coulbois, Aurélien Massein, 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 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.

This year we have focused our attention in several aspects : improvement of the hardware of the experimental monitoring system and tools for handling and analyzing the data gathered.

The PhD work about optimal location of sensors in a smart environments has been defended in november, defining new metrics on set of sensors and new methods [0].

7.3.1. Hardware

Two monitoring systems have been installed. The first one in the first floor of EHPAD Valrose in Nice, and a second one in Institut Claude Pompidou in Nice. Both systems are composed of multi sensors barriers that provide raw data from which we deduce the time and direction of its crossing by a person.

For the second experimental system the analysis of the first data have shown that the system was not reliable enough while the data themselves were not satisfactory because of the specificity of the building (large corridors, large waiting room, picture windows and the number of sensors installed (77). We have worked on the hardware of the system (redundant power supply, better orientation of barriers, better communication system) to improve the gathered data.

[0] Design of Instrumented Environment for Human Monitoring, defended on 12/26/2018

7.3.2. Tools for handling data and data analysis

We have developed a simulation program, written in C and using the GTK library, that generates barrier-events (i.e. crossing time, direction of crossing, speed of crossing). This program is based on Monte Carlo procedures simulating the displacement of both elderly and caregivers in the EHPAD environment equipped with movement detectors. The code can simulate up to 20 persons and randomly draws room-to-room movements according to the walking speed of each individual (caregivers walk at a faster pace than elderly), and counts the locations and time coordinates of each movement event identified by the detectors. The figure 4 gives a view of the graphic interface. Such a simulation program, and the results produced, will provide basic training data to reconstruct patient movements from the information collected by the activity detectors.

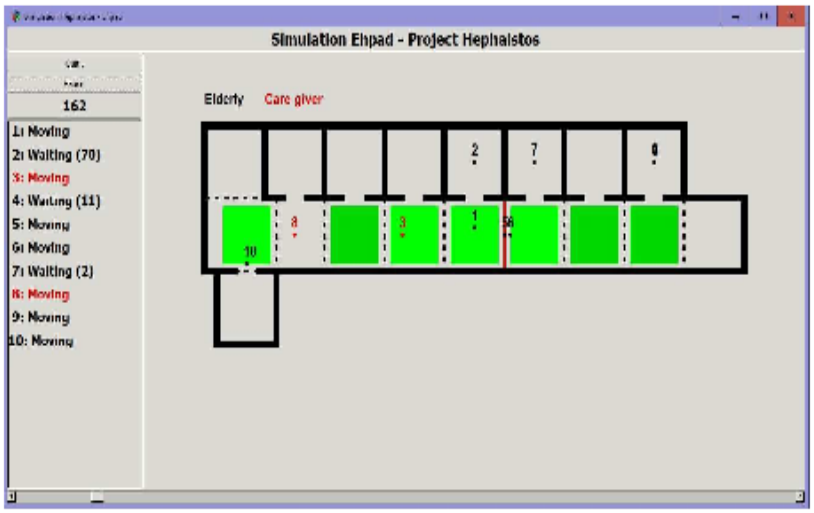


Figure 4. Simulation tool for event detection analysis

Another scientific activities were based on the development of diagnostic tools (also written in C) to visualize (and thus to check and to interpret) events identified by each detector in such equipped environment. Finally, another activity – that is still under development – is to analyze statistically gait data obtained through the event detection. In this case, the goal is to build a series of relevant statistical descriptive parameters that will be used to describe, identify and compare gait features and pathology in medically assisted environments. This last part is developed used the R software.

In the two installed system data are collected continuously during the all day and a large number of barrier crossing is observed. We are currently comparing raw and simulated data before moving on with a statistical analysis.

LARSEN Project-Team

7. New Results

7.1. Lifelong Autonomy

7.1.1. Foundations of Reinforcement Learning

7.1.1.1. ρ-POMDPs have Lipschitz-Continuous ϵ-Optimal Value Functions

Participant: Vincent Thomas.

Collaboration with Jilles Dibangoye (INSA Lyon).

Many state-of-the-art algorithms for solving Partially Observable Markov Decision Processes (POMDPs) rely on turning the problem into a "fully observable" problem—a belief MDP—and exploiting the piece-wise linearity and convexity (PWLC) of the optimal value function in this new state space (the belief simplex Δ). This approach has been extended to solving ρ-POMDPs—i.e., for information-oriented criteria-when the reward ρ is convex in Δ. General ρ-POMDPs can also be turned into "fully observable" problems, but with no means to exploit the PWLC property. In this paper, we focus on POMDPs and ρ-POMDPs with $\lambda\rho$-Lipschitz reward function, and demonstrate that, for finite horizons, the optimal value function is Lipschitz-continuous. Then, value function approximators are proposed for both upper-and lower-bounding the optimal value function, which are shown to provide uniformly improvable bounds. This allows proposing two algorithms derived from HSVI which are empirically evaluated on various benchmark problems.

Publication: [14]

7.1.1.2. Addressing Active Sensing Problem through MCTS

Participants: Vincent Thomas, Geremy Hutin.

The problem of active sensing is of paramount interest for building self awareness in robotic systems. It consists of a system to make decisions in order to gather information (measured through the entropy of the probability distribution over unknown variables) in an optimal way.

In the past, we have proposed an original formalism *ρ-POMDP* and new algorithms for representing and solving active sensing problems [33] by using point-based algorithms. This year, new approaches based on Monte-Carlo Tree Search algorithms (MCTS) and Partially Observable Monte-Carlo Planning (POMCP) [45] have been proposed to build the policies of an agent whose aim is to gather information.

7.1.2. Robot Learning

Our main objective is to design data-efficient trial-and-error learning algorithms (reinforcement learning) that can work with continuous states and continuous actions. The main use-case is robot damage recovery: a robot has to to discover new behaviors by trial-and-error without a diagnosis of the damage.

7.1.2.1. Adaptive and Resilient Soft Tensegrity Robots

Participant: Jean-Baptiste Mouret.

Collaboration with John Rieffel (Union College, USA).

Living organisms intertwine soft (e.g., muscle) and hard (e.g., bones) materials, giving them an intrinsic flexibility and resiliency often lacking in conventional rigid robots. The emerging field of soft robotics seeks to harness these same properties to create resilient machines. The nature of soft materials, however, presents considerable challenges to aspects of design, construction, and control—and up until now, the vast majority of gaits for soft robots have been hand-designed through empirical trial-and-error. In this contribution, we introduced an easy-to-assemble tensegrity-based soft robot capable of highly dynamic locomotive gaits and demonstrating structural and behavioral resilience in the face of physical damage. Enabling this is the use of a machine learning algorithm able to discover effective gaits with a minimal number of physical trials. These results lend further credence to soft-robotic approaches that seek to harness the interaction of complex material dynamics to generate a wealth of dynamical behaviors.

Publication: [10]

7.1.2.2. Bayesian Optimization with Automatic Prior Selection for Data-Efficient Direct Policy Search

Participants: Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret.

One of the most interesting features of Bayesian optimization for direct policy search is that it can leverage priors (e.g., from simulation or from previous tasks) to accelerate learning on a robot. In this contribution, we are interested in situations for which several priors exist but we do not know in advance which one fits best the current situation. We tackle this problem by introducing a novel acquisition function, called Most Likely Expected Improvement (MLEI), that combines the likelihood of the priors and the expected improvement. We evaluate this new acquisition function on a transfer learning task for a 5-DOF planar arm and on a possibly damaged, 6-legged robot that has to learn to walk on flat ground and on stairs, with priors corresponding to different stairs and different kinds of damages. Our results show that MLEI effectively identifies and exploits the priors, even when there is no obvious match between the current situations and the priors.

Publication: [23]

7.1.2.3. Multi-objective Model-based Policy Search for Data-efficient Learning with Sparse Rewards

Participants: Rituraj Kaushik, Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret.

The most data-efficient algorithms for reinforcement learning in robotics are model-based policy search algorithms, which alternate between learning a dynamical model of the robot and optimizing a policy to maximize the expected return given the model and its uncertainties. However, the current algorithms lack an effective exploration strategy to deal with sparse or misleading reward scenarios: if they do not experience any state with a positive reward during the initial random exploration, they are very unlikely to solve the problem. To address this challenge, we proposed a novel model-based policy search algorithm, Multi-DEX, that leverages a learned dynamical model to efficiently explore the task space and solve tasks with sparse rewards in a few episodes. To achieve this, we frame the policy search problem as a multi-objective, model-based policy optimization problem with three objectives: (1) generate maximally novel state trajectories, (2) maximize the cumulative reward and (3) keep the system in state-space regions for which the model is as accurate as possible. We then optimize these objectives using a Pareto-based multi-objective optimization algorithm. The experiments show that Multi-DEX is able to solve sparse reward scenarios (with a simulated robotic arm) in much lower interaction time than VIME, TRPO, GEP-PG, CMA-ES and Black-DROPS.

Publication: [18]

7.1.2.4. Using Parameterized Black-Box Priors to Scale Up Model-Based Policy Search for Robotics

Participants: Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret.

Among the few model-based policy search algorithms, the recently introduced Black-DROPS algorithm exploits a black-box optimization algorithm to achieve both high data-efficiency and good computation times when several cores are used; nevertheless, like all model-based policy search approaches, Black-DROPS does not scale to high dimensional state/action spaces. In this paper, we introduce a new model learning procedure in Black-DROPS that leverages parameterized black-box priors to (1) scale up to high-dimensional systems, and (2) be robust to large inaccuracies of the prior information. We demonstrate the effectiveness of our approach with the "pendubot" swing-up task in simulation and with a physical hexapod robot (48D state space, 18D action space) that has to walk forward as fast as possible. The results show that our new algorithm is more data-efficient than previous model-based policy search algorithms (with and without priors) and that it can allow a physical 6-legged robot to learn new gaits in only 16 to 30 seconds of interaction time.

Publication: [12]

7.1.2.5. Data-efficient Neuroevolution with Kernel-Based Surrogate Models

Participants: Adam Gaier, Jean-Baptiste Mouret.

Collaboration with Alexander Asteroth (Hochschule Bonn-Rhein-Sieg, Germany)

Surrogate-assistance approaches have long been used in computationally expensive domains to improve the data-efficiency of optimization algorithms. Neuroevolution, however, has so far resisted the application of these techniques because it requires the surrogate model to make fitness predictions based on variable topologies, instead of a vector of parameters. Our main insight is that we can sidestep this problem by using kernel-based surrogate models, which require only the definition of a distance measure between individuals. Our second insight is that the well-established Neuroevolution of Augmenting Topologies (NEAT) algorithm provides a computationally efficient distance measure between dissimilar networks in the form of "compatibility distance", initially designed to maintain topological diversity. Combining these two ideas, we introduce a surrogate-assisted neuroevolution algorithm that combines NEAT and a surrogate model built using a compatibility distance kernel. We demonstrate the data-efficiency of this new algorithm on the low dimensional cart-pole swing-up problem, as well as the higher dimensional half-cheetah running task. In both tasks the surrogate-assisted variant achieves the same or better results with several times fewer function evaluations as the original NEAT.

Publication: [17] (best paper, GECCO 2018, Complex System track)

7.1.2.6. Alternating Optimization and Quadrature for Robust Control

Participants: Konstantinos Chatzilygeroudis, Jean-Baptiste Mouret.

Collaboration with Shimon Whiteson (Oxford, UK).

Bayesian optimization has been successfully applied to a variety of reinforcement learning problems. However, the traditional approach for learning optimal policies in simulators does not utilise the opportunity to improve learning by adjusting certain environment variables — state features that are randomly determined by the environment in a physical setting but are controllable in a simulator. In this wwork, we consider the problem of finding an optimal policy while taking into account the impact of environment variables. We present alternating optimization and quadrature (ALOQ), which uses Bayesian optimization and Bayesian quadrature to address such settings. ALOQ is robust to the presence of significant rare events, which may not be observable under random sampling, but have a considerable impact on determining the optimal policy. The experimental results demonstrate that our approach learns more efficiently than existing methods.

Publication: [22]

7.1.2.7. Learning robust task priorities of QP-based whole-body torque-controllers

Participants: Marie Charbonneau, Serena Ivaldi, Valerio Modugno, Jean-Baptiste Mouret.

Generating complex whole-body movements for humanoid robots is now most often achieved with multi-task whole-body controllers based on quadratic programming. To perform on the real robot, such controllers often require a human expert to tune or optimize the many parameters of the controller related to the tasks and to the specific robot, which is generally reported as a tedious and time consuming procedure. This problem can be tackled by automatically optimizing some parameters such as task priorities or task trajectories, while ensuring constraints satisfaction, through simulation. However, this does not guarantee that parameters optimized in simulation will also be optimal for the real robot. As a solution, the present paper focuses on optimizing task priorities in a robust way, by looking for solutions which achieve desired tasks under a variety of conditions and perturbations. This approach, which can be referred to as domain randomization, can greatly facilitate the transfer of optimized solutions from simulation to a real robot. The proposed method is demonstrated using a simulation of the humanoid robot iCub for a whole-body stepping task.

Publication: [11]

7.1.3. Quality Diversity Algorithms

Quality diversity algorithms are a new kind of evolutionary algorithms that focuses on finding a large set of high-performing solutions (instead of the global optimum). We use them for design and as a step for data-efficient robot learning.

7.1.3.1. Data-Efficient Design Exploration through Surrogate-Assisted Illumination

Participants: Adam Gaier, Jean-Baptiste Mouret.

Collaboration with Alexander Asteroth (Hochschule Bonn-Rhein-Sieg, Germany)

Design optimization techniques are often used at the beginning of the design process to explore the space of possible designs. In these domains illumination algorithms, such as MAP-Elites, are promising alternatives to classic optimization algorithms because they produce diverse, high-quality solutions in a single run, instead of only a single near-optimal solution. Unfortunately, these algorithms currently require a large number of function evaluations, limiting their applicability. In this work, we introduce a new illumination algorithm, Surrogate-Assisted Illumination (SAIL), that leverages surrogate modeling techniques to create a map of the design space according to user-defined features while minimizing the number of fitness evaluations. On a 2-dimensional airfoil optimization problem SAIL produces hundreds of diverse but high-performing designs with several orders of magnitude fewer evaluations than MAP-Elites or CMA-ES. We demonstrate that SAIL is also capable of producing maps of high-performing designs in realistic 3-dimensional aerodynamic tasks with an accurate flow simulation. Data-efficient design exploration with SAIL can help designers understand what is possible, beyond what is optimal, by considering more than pure objective-based optimization.

Publication: [7]

7.1.3.2. Discovering the Elite Hypervolume by Leveraging Interspecies Correlation

Participants: Vassilis Vassiliades, Jean-Baptiste Mouret.

Evolution has produced an astonishing diversity of species, each filling a different niche. Algorithms like MAP-Elites mimic this divergent evolutionary process to find a set of behaviorally diverse but high-performing solutions, called the elites. Our key insight is that species in nature often share a surprisingly large part of their genome, in spite of occupying very different niches; similarly, the elites are likely to be concentrated in a specific "elite hypervolume" whose shape is defined by their common features. In this paper, we first introduce the elite hypervolume concept and propose two metrics to characterize it: the genotypic spread and the genotypic similarity. We then introduce a new variation operator, called “directional variation”, that exploits interspecies (or inter-elites) correlations to accelerate the MAP-Elites algorithm. We demonstrate the effectiveness of this operator in three problems (a toy function, a redundant robotic arm, and a hexapod robot).

Publication: [25]

7.1.3.3. Maintaining Diversity in Robot Swarms with Distributed Embodied Evolution

Participants: Amine Boumaza, François Charpillet.

We investigated how behavioral diversity can be maintained in evolving robot swarms by using distributed Embodied Evolution. In these approaches, each robot in the swarm runs a separate evolutionary algorithm, and populations on each robot are built through local communication when robots meet; therefore, genome survival results not only from fitness-based selection but also from spatial spread. To better understand how diversity is maintained in distributed embodied evolution, we propose a postanalysis diversity measure — global diversity (over the swarm), and local diversity (on each robot) —, on two swarm robotic tasks — navigation and item collection —, with different intensities of selection pressure, and compare the results of distributed embodied evolution to a centralized case. We conclude that distributed evolution intrinsically maintains a larger behavioral diversity when compared to centralized evolution, which allows for the search algorithm to reach higher performances, especially in the more challenging collection task.

Publication: [16]

7.2. Natural Interaction with Robotics Systems

7.2.1. Control of Interaction

Because of the AnDy project, we are currently focused on interaction in industrial contexts, in particular to encourage ergonomic motions.

7.2.1.1. Robust Real-time Whole-Body Motion Retargeting from Human to Humanoid

Participants: Serena Ivaldi, Luigi Penco, Brice Clement, Jean-Baptiste Mouret.

Transferring the motion from a human operator to a humanoid robot is a crucial step to enable robots to learn from and replicate human movements. The ability to retarget in real-time whole-body motions that are challenging for the humanoid balance is critical to enable human to humanoid teleoperation. In this work, we design a retargeting framework that allows the robot to replicate the motion of the human operator, acquired by a wearable motion capture suit, while maintaining the whole-body balance. We introduce some dynamic filter in the retargeting to forbid dangerous motions that can make the robot fall. We validate our approach through several experiments on the iCub robot, which has a significantly different body structure and size from the one of the human operator.

Publication: [24]

7.2.1.2. Prediction of Human Whole-Body Movements with AE-ProMPs

Participants: Serena Ivaldi, Oriane Dermy, Francis Colas, François Charpillet.

The ability to predict intended movements is crucial for collaborative robots to anticipate the human actions and for assistive technologies to alert if a particular movement is non-ergonomic and potentially dangerous for humans. In this paper, we address the problem of predicting the future human whole-body movements given early observations. We propose to predict the continuation of the high-dimensional trajectories mapped into a reduced latent space, using autoencoders (AE). The prediction is based on a probabilistic description of the movement primitives (ProMPs) in the latent space, which notably reduces the computational time for the prediction to occur, and hence enables to use the method in real-time applications. We evaluate our method, named AE-ProMPs, for predicting future movements belonging to a dataset of 7 different actions performed by a human, recorded by a wearable motion tracking suit.

Publication: [13]

Publications: [28],

7.2.2. Generating Assistive Humanoid Motions for Co-Manipulation Tasks with a Multi-Robot Quadratic Program Controller

Participants: Karim Bouyarmane, Serena Ivaldi.

Human-humanoid collaborative tasks require that the robot takes into account the goals of the task, interaction forces with the human, and its own balance. We present a formulation for a real-time humanoid controller which allows the robot to keep itself stable, while also assisting the human in achieving their shared objectives. This is achieved with a multi-robot quadratic program controller, which solves for human motion reconstruction and optimal robot controls in a single optimization problem. Our experiments on a simulated robot platform demonstrate the ability to generate interactions motions and forces that are similar to what a human collaborator would produce.

Publication: [21]

7.2.2.1. Activity Recognition With Multiple Wearable Sensors for Industrial Applications

Participants: Francis Colas, Serena Ivaldi, Adrien Malaisé, Pauline Maurice, François Charpillet.

We address the problem of recognizing the current activity performed by a human operator, providing an information useful for automatic ergonomic evaluation for industrial applications. While the majority of research in activity recognition relies on cameras observing the human, here we explore the use of wearable sensors, which are more suitable in industrial environments. We use a wearable motion tracking suit and a sensorized glove. We describe our approach for activity recognition with a probabilistic model based on Hidden Markov Models, applied to the problem of recognizing elementary activities during a pick-and-place task inspired by a manufacturing scenario. We show that our model is able to correctly recognize the activities with 96% of precision if both sensors are used.

Publication: [19],[19]

7.2.2.2. Activity Recognition for monitoring eldely people at home

Participants: Yassine El Khadiri, François Charpillet.

Early detection of frailty signs is important for senior people who prefer to keep living in their homes instead of moving to a nursing home. Sleep quality is a good predictor for frailty monitoring. Thus we are interested in tracking sleep parameters like sleep wake patterns to predict and detect potential sleep disturbances of the monitored senior residents. We use an unsupervised inference method based on actigraphy data generated by ambient motion sensors scattered around the senior's apartment. This enables our monitoring solution to be flexible and robust to the different types of housings it can equip while still attaining accuracy of 0.94 for sleep period estimates.

Publication: [15]

7.2.3. Ethics

7.2.3.1. Ethical and Social Considerations for the Introduction of Human-Centered Technologies at Work

Participants: Serena Ivaldi, Adrien Malaisé, Pauline Maurice, Ludivine Allienne.

Human-centered technologies such as collaborative robots, exoskeletons, and wearable sensors are rapidly spreading in industry and manufacturing because of their intrinsic potential at assisting workers and improving their working conditions. The deployment of these technologies, albeit inevitable, poses several ethical and societal issues. Guidelines for ethically aligned design of autonomous and intelligent systems do exist, however we argue that ethical recommendations must necessarily be complemented by an analysis of the social impact of these technologies.

In a recent paper[20], we report on our preliminary studies on the opinion of factory workers and of people outside this environment on human-centered technologies at work. In light of these studies, we discuss ethical and social considerations for deploying these technologies in a way that improves acceptance.

Publication: [20]

PERVASIVE Project-Team

7. New Results

7.1. Using Attention to Address Human-Robot Motion

Participants: Thierry Fraichard, Rémi Paulin, Patrick Reignier.

To capture the specificity of robot motion among people, we choose the term **Human-Robot Motion** (HRM) [0], to denote the study of how robots should move among people. HRM is about designing robots whose motions are deemed socially **acceptable** from a human point of view while remaining **safe**.

After 15 years of research on HRM, the main concept that has emerged is that of *social spaces*, *i.e.* regions of the environment that people consider as psychologically theirs [33], any intrusion in their social space will be a source of discomfort. Such social spaces are characterized by the position of the person, *i.e.* "Personal Space", or the activity they are currently engaged in, *i.e.* "Interaction Space" and "Activity Space". The most common approach in HRM is to define costmaps on such social spaces: the higher the cost, the less desirable it is to be there. The costmaps are then used for navigation purposes, *e.g.* [37] and [36].

Social spaces are of course relevant to HRM but they have limitations. First, it is not straightforward to define them; what is their shape or size, especially in cluttered environments? Second, it seems obvious that there is more to acceptability than geometry only: the appearance of a robot and its velocity will also influence the way it is perceived by people. Finally, social spaces can be conflicting because when a robot needs to interact with a person, it is very likely that it will have to penetrate a social space.

To complement social spaces, we have started to explore whether human attention could be useful to address HRM vis-à-vis the acceptability aspect. Why attention? The answer is straightforward: the acceptability of a robot motion is directly related to the way it is perceived by a person hence our interest in human attention. For a person, attention is a cognitive mechanism for filtering the person's sensory information (to avoid an overwhelming amount of information) [35]. It controls where and to what the person's attentional resources are allocated.

In 2014, we introduced the concept of **attention field**, *i.e.* a predictor of the amount of attention that a person allocates to the robot when the robot is in a given state. In [32], the attention field was computed thanks to a computational model of attention proposed in [34] in the context of ambient applications and pervasive systems. In this model, attentional resources are focused on a single specific area of the person's visual space (as per the zoom lens model [31]). Later studies have demonstrated that the situation is more complex and that attentional resources can be distributed over multiple objects in the visual space [35].

In 2018, we have developed a novel **computational model of attention** that takes this property into account. This model is used to compute the attention field for a robot. The attention field is then used to define different **attentional properties** for the robot's motions such as distraction or surprise. The relevance of the attentional properties for HRM have been demonstrated on a proof-of-concept **acceptable motion planner** on various case studies where a robot is assigned different tasks. The multi-criteria nature of motion planning in the context of HRM led to the design of an acceptable motion planner based upon a state-of-the-art many-objective optimization algorithm. It shows how to compute acceptable motions that are non-distracting and non-surprising, but also motions that convey the robot's intention to interact with a person. All these contributions have been presented in the PhD of Rémi Paulin [6] and the conference article [26].

7.2. Simulating Haptic Sensations

Participants: Jingtao Chen, Sabine Coquillart

[0] In reference to Human-Robot Interaction (HRI), *i.e.* the study of the interactions, in the broad sense of the word, between people and robots.

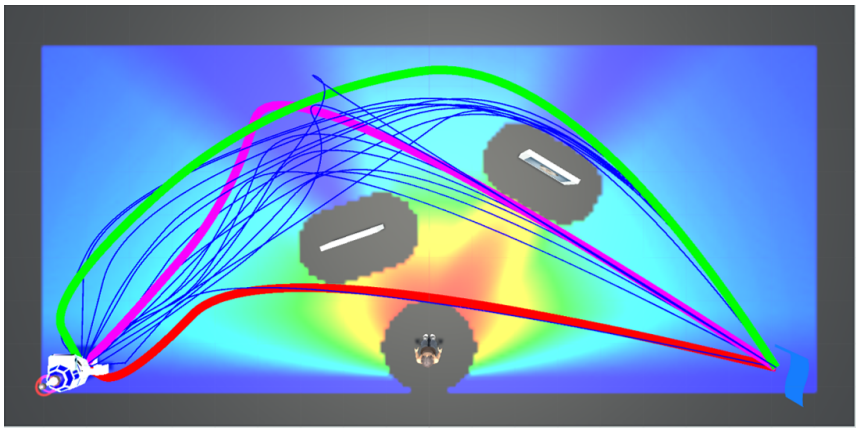


Figure 1. Motions with different attentional properties in a scenario featuring a person watching at paintings in a museum and a robot which is to travel from left to right: less distracting (green) vs. shortest (red) motions are depicted. The purple motion is a trade-off solution.

Partners: Inria GRA, LIG, GIPSA, G-SCOP

Pseudo-haptic feedback is a technique aiming to simulate haptic sensations without active haptic feedback devices. Peudo-haptic techniques have been used to simulate various haptic feedbacks such as stiffness, torques, and mass. In the framework of the Persyval project, a novel pseudo-haptic experiment has been set up. The aim of this experiment is to study the force and EMG signals during a pseudo-haptic task. A stiffness discrimination task similar to the one published in Lecuyer's PhD thesis has been chosen. The experimental set-up has been developed, as well as the software controlling the experiment. Pre-tests have been conducted. They have been followed by formal tests with subjects.

7.3. Observing and Modeling Awareness and Expertise During Problem Solving

Participants: Thomas Guntz, Dominique Vaufreydaz, James Crowley, Philippe Dessus, Raffaella Balzarini

7.3.1. Observing and Modelling Competence and Awareness from Eye-gaze and Emotion

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 [Guntz et al 18a]. 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 [Damasio 91].

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.

In 2019 we will conduct an additional experiment designed confirm and explore this hypothesis. Results will be reported in a journal paper (under preparation) as well as in the doctoral thesis of Thomas Guntz, to be defended in late 2019.

7.3.2. Bibliography

[Damasio 91] Damasio, A., Somatic Markers and the Guidance of Behavior. New York: Oxford University Press. pp. 217–299, 1991.

[Guntz et al. 18a] T. Guntz, R. Balzarini, D. Vaufreydaz, and J.L. Crowley, "Multimodal Observation and Classification of People Engaged in Problem Solving: Application to Chess Players". Multimodal Technologies and Interaction, Vol 2 No. 2, p11, 2018.

[Guntz et al. 18b] T. Guntz, J.L. Crowley, D. Vaufreydaz, R. Balzarini, P. Dessus, The Role of Emotion in Problem Solving: first results from observing chess, Workshop on Modeling Cognitive Processes from Multimodal Data, at the 2018 ACM International Conference on Multimodal Interaction, ICMI 2018, Oct 2018.

7.4. Learning Routine Patterns of Activity in the Home

Participants: Julien Cumin, James Crowley

Other Partners: Fano Ramparany, Greg Lefevre (Orange Labs)

During the month of February 2017, we have collected 4 weeks of data on daily activities within the Amiqual4Home Smart Home Living lab apartment. This dataset was presented at the international Conference on Ubiquitous Computing and Ambient Intelligence, UCAmI 2017, at Bethlehem PA, in Nov 2017 and is currently available for download from the Amiqual4Home web server (http://amiqual4home.inria.fr/en/orange4home/)

The objective of this research action is to develop a scalable approach to learning routine patterns of activity in a home using situation models. Information about user actions is used to construct situation models in which key elements are semantic time, place, social role, and actions. Activities are encoded as sequences of situations. Recurrent activities are detected as sequences of activities that occur at a specific time and place

each day. Recurrent activities provide routines that can be used to predict future actions and anticipate needs and services. An early demonstration has been to construct an intelligent assistant that can respond to and filter inter-personal communications.

7.5. Bayesian Reasoning

Participants: Emmanuel Mazer, Raphael Frisch, Marvin Faix, Augustin Lux, Didier Piau, Jeremy Belot.

To overcome the ever growing needs in computing power, alternative computing paradigms have been developed such as stochastic architectures. These latter have found substantial interests for energy efficient implementations in artificial intelligence. In particular, mixing stochastic computing with Bayesian models makes a promising paradigm for non-conventional computational architectures dedicated to Bayesian inference. The ability to deal with uncertainty and adapt its computational accuracy is some of the advantages of these computing approaches.

During 2018 we have designed a first hardware prototype to localize a sound source with a stochastic machine. The goal of this project was to provide a proof of concept of stochastic machines by implementing an autonomous platform of sound source localization. It includes an sound acquisition module, a pre-processing circuit, and the stochastic machine. The platform has been implemented on an Altera Cyclone V FPGA and validated functionally with digital simulations. Several optimization to improve size and power consumption have been proposed. Results in terms of computation time, power and used FPGA resources allowed to assess their impact on future design. The same architecture of stochastic machine was also analyzed in simulation to provide design guidelines for our next design [25].

Further, we have proposed a way to reduce the memory needs of our architecture by sharing a memory between the processing units (in collaboration with TIMA and C2M -Université Paris Sud). This optimization reduces the area and the cost of our architecture. However, its impact on power consumption is not obvious. Therefore, we designed an integrated circuit (ASIC) with our original and optimized proposals. We synthesized the VHDL description of the circuit in the FDSOI 28nm technology from STMicroelectronics. Notice that the memory has been implemented thanks to a SRAM memory compiler. The results highlight that the optimized machine significantly reduces both the circuit area (by 30%) and the power consumption (by 35%). Nevertheless, the simulations showed that, in the optimized version, the memory represents nearly 60 % of our circuit area and more than 55% of the power consumption. According to the latest literature, the Magnetic Random Access Memory (MRAM) technology provides some promising features and would approximately reduce by a factor of 20 the memory area. Moreover, this feature should drastically impact the power consumption. Thus, our future works will focus on the implementation of Bayesian machines using MRAM instead of SRAM. A poster describing this work was presented at the International Conference on rebooting Computing.

We have proposed (in collaboration with ISIR - Université Paris Sorbonne)a new way to localize several sound sources using a Bayesian model. This multi-source localization algorithm is fast and can readily be implemented on our stochastic machine (Paper submitted at ICASSP 2019). The Figure 2 shows the location of the source and of the microphones in the simulated environment. The Figure 3 shows the posterior distribution of the location of one source using a short frame and the Figure 4 shows the result using fifty frames. As the frame are very short the localization of the two sources is readily obtained and it is used as a bootstrap for the source separation algorithm .

We devised and successfully tested a Bayesian model for the source separation problem. The model assumes the localization of the sources are known. The inference - retrieving the sound emitted by each source from the mixed signals obtained with several microphones - takes place in a very high dimensional space. Nevertheless, the Gibbs algorithm is well suited to solve the problem when the location of the sources are known. A very efficient implementation of this algorithm was tested with a realistic sound simulator using human voices. The algorithm can be implemented on a sampling machine and the corresponding stochastic architecture has been devised. It is currently implemented on an FPGA.

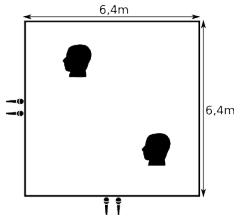
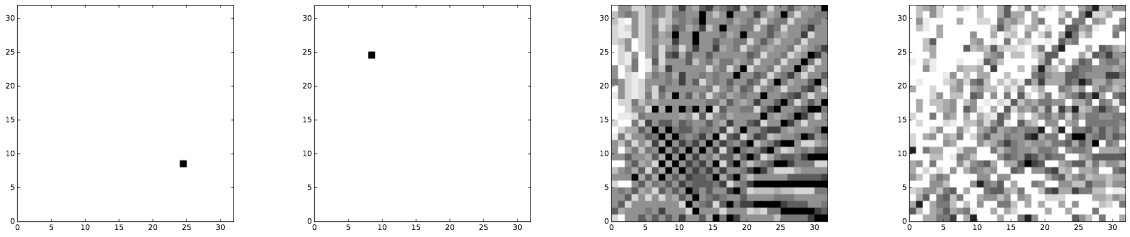


Figure 2. Simulated room setup.



(a) Frame 1 (b) Frame 3 (c) Frame 29 (d) Frame 46

Figure 3. Posterior distribution maps for a single source obtained for 4 very short time-frames of a given 50-frame bloc.

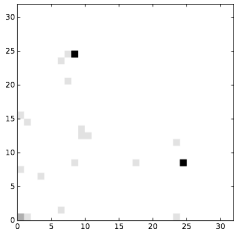


Figure 4. Final distribution map after fusion over 50 frames. The two black squares correspond to the actual positions of the two sources.

RAINBOW Project-Team

7. New Results

7.1. Optimal and Uncertainty-Aware Sensing

7.1.1. Visual Tracking for Motion Capture and virtual reality

Participants: Guillaume Cortes [Hybrid], Eric Marchand.

Considering the visual tracking system for motion proposed last year, we studied a novel approach for Mobile Spatial Augmented Reality on Tangible objects [14]. MoSART is dedicated to mobile interaction with tangible objects in single or collaborative situations. It is based on a novel 'all-in-one' Head-Mounted Display (AMD) including a projector (for the SAR display) and cameras (for the scene registration). Equipped with the HMD the user is able to move freely around tangible objects and manipulate them at will. The system tracks the position and orientation of the tangible 3D objects and projects virtual content over them. The tracking is a feature-based stereo optical tracking providing high accuracy and low latency. A projection mapping technique is used for the projection on the tangible objects which can have a complex 3D geometry. Several interaction tools have also been designed to interact with the tangible and augmented content, such as a control panel and a pointer metaphor, which can benefit as well from the MoSART projection mapping and

7.1.2. Deformable Object 3D Tracking based on Depth Information and Physical Model

Participants: Agniva Sengupta, Eric Marchand, Alexandre Krupa.

In the context of the iProcess project (see Section 9.3.3.2), we have developed a method for tracking rigid objects of complex shapes. This year, we started to elaborate a method to track deformable objects using a depth camera (RGB-D sensor). This method is based on the assumption that a coarse mesh representing the model of the object is known and that a simple volumetric tetrahedral mesh has been computed offline, representing the internal physical model of the object. To take into account the deformation of the object, a corotational Finite Element Method (FEM) is considered as the physical model. Given the sequential pointcloud of the object undergoing deformation, we have developed an algorithm that fits the deformable model to the observed pointcloud. The FEM simulation is done using the SOFA library and our approach was tested for the tracking of simulated deformation of objects. For the moment, the method succeeds to accurately track the object deformation, given that we know the point of application of force (causing the deformation) and the force direction vector. Online estimation of the direction vector of this force is currently a work in progress.

7.1.3. General Model-based Tracker

Participants: Souriya Trinh, Fabien Spindler, Eric Marchand, François Chaumette.

We have extended our model-based visual tracking method by considering as new potential measurement the depth map provided by a RGB-D sensor [75]. The method has been adapted to be fully modular and can combine edge, texture, and depth features. It has been released in the new version of ViSP.

7.1.4. Reflectance and Illumination Estimation for Realistic Augmented Reality

Participants: Salma Jiddi, Eric Marchand.

Photometric registration consists in blending real and virtual scenes in a visually coherent way. To achieve this goal, both reflectance and illumination properties must be estimated. These estimates are then used, within a rendering pipeline, to virtually simulate the real lighting interaction with the scene.

We have been interested in indoor scenes where light bounces off of objects with different reflective properties (diffuse and/or specular). In these scenarios, existing solutions often assume distant lighting or limit the analysis to a single specular object [63]. We address scenes with various objects captured by a moving RGB-D camera and estimate the 3D position of light sources. Furthermore, using spatio-temporal data, our algorithm recovers dense diffuse and specular reflectance maps. Finally, using our estimates, we demonstrate photo-realistic augmentations of real scenes (virtual shadows, specular occlusions) as well as virtual specular reflections on real world surfaces.

We also consider the problem of estimating the 3D position and intensity of multiple light sources using an approach based on cast shadows on textured real surfaces [62], [86]. We separate albedo/texture and illumination using lightness ratios between pairs of points with the same reflectance property but subject to different lighting conditions. Our selection algorithm is robust in presence of challenging textured surfaces. Then, estimated illumination ratios are integrated, at each frame, within an iterative process to recover position and intensity of light sources responsible of cast shadows.

7.1.5. Multi-Layered Image Representation for Robust SLAM

Participant: Eric Marchand.

Robustness of indirect SLAM techniques to light changing conditions remains a central issue in the robotics community. With the change in the illumination of a scene, feature points are either not extracted properly due to low contrasts, or not matched due to large differences in descriptors. We proposed a multi-layered image representation (MLI) that computes and stores different contrast-enhanced versions of an original image [76]. Keypoint detection is performed on each layer, yielding better robustness to light changes. An optimization technique is also proposed to compute the best contrast enhancements to apply in each layer in order to improve detection and matching. We extend the MLI approach [77] and we show how Mutual Information can be used to compute dynamic contrast enhancements on each layer. We demonstrate how this approach dramatically improves the robustness in dynamic light changing conditions on both synthetic and real environments compared to default ORB-SLAM. This work focuses on the specific case of SLAM relocalization in which a first pass on a reference video constructs a map, and a second pass with a light changed condition relocalizes the camera in the map.

7.1.6. Trajectory Generation for Optimal State Estimation

Participants: Marco Cognetti, Marco Ferro, 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. Indeed, the latter is far from being negligible for several robotic applications (a prominent example are aerial vehicles). Last year we developed a general framework for solving *online* the active sensing problem by continuously replanning an optimal trajectory that maximize a suitable norm of the Constructibility Gramian (CG), while also coping with a number of constraints including limited energy and feasibility. This approach, however, did not consider the presence of process noise which, as explained, can have a significant effect in many robotic systems of interest (e.g., UAVs). This year we have then extended this work to the case of a non-negligible process noise in [56], where we showed how to generate optimal trajectories able to still maximize the amount of information collected while moving, but by properly weighting (and attenuating) the negative effects of process noise in the execution of the planned trajectory. 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 manipulators), and to the mutual localization problem for a multi-robot group.

7.1.7. 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. Uncertainty in the landmarks positions have to be considered, but this can lead to pessimism in the computed solution. Hence we derived a quantifier-free expression of the pose solution-set to improve the vision-based position domain computation [66]. Image and range based cooperative localization of UAVs has been studied, first in the case of two robots sharing their measurements [65]. Then, scaling to the case of multiple robots as also been addressed, by sharing the computed position domains [64], [67].

7.2. Advanced Sensor-Based Control

7.2.1. Model Predictive Control for Visual Servoing of a UAV

Participants: Bryan Penin, François Chaumette, Paolo Robuffo Giordano.

Visual servoing is a well-known class of techniques meant to control the pose of a robot from visual input by considering an error function directly defined in the image (sensor) space. These techniques are particularly appealing since they do not require, in general, a full state reconstruction, thus granting more robustness and lower computational loads. However, because of the quadrotor underactuation and inherent sensor limitations (mainly limited camera field of view), extending the classical visual servoing framework to the quadrotor flight control is not straightforward. For instance, for realizing a horizontal displacement the quadrotor needs to tilt in the desired direction. This tilting, however, will cause any downlooking camera to point in the opposite direction with, e.g., possible loss of feature tracking because of the limited camera field of view.

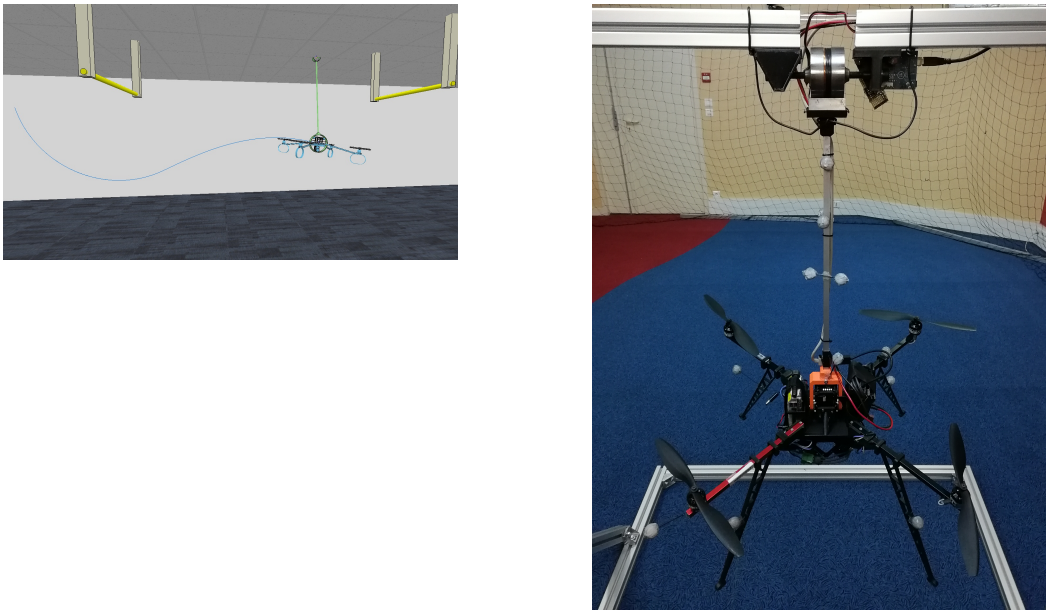
In order to cope with these difficulties and achieve a high-performance visual servoing of quadrotor UAVs, we have developed a series of online trajectory re-planning (MPC-like) schemes for explicitly dealing with this kind of constraints during flight. In particular, in [33], the problem of aggressive flight when tracking a target has been considered, with the additional (and complex) constraint of avoiding occlusions w.r.t. obstacles in the scene. A suitable optimization framework has been devised to be solved online during flight for continuously replanning the future UAV trajectory subject to the mentioned sensing constraints as well as actuation constraints. An experimental validation with the quadrotor UAVs available in the team has also been provided. In [34], we have instead considered the problem of planning a trajectory from a start to a goal location for a UAV equipped with an onboard camera, by assuming that measurements of environment landmarks (needed to recover the UAV state from visual input) may be intermittent due to occlusions by obstacles. The goal is then to plan a trajectory that can minimize the negative effects of "missing measurements" by keeping the state uncertainty limited despite the temporary loss of measurements. This planning problems has been solved by exploiting a bi-directional RRT algorithm for joining the start and goal locations, and an experimental validation has also been performed.

7.2.2. 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 in [17] 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. 9 (a). We are currently finalizing a real prototype (Fig. 9 (b)) for obtaining an experimental validation of the whole approach.



(a) (b)

Figure 9. UAVs in Physical Interaction with the Environment. a) The simulated MonkeyRotor performing a hook-to-hook maneuver. b) The prototype currently under finalization.

7.2.3. Trajectory Generation for Minimum Closed-Loop State Sensitivity

Participants: 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. In [70], we have explored this novel idea by showing how to properly define and evaluate the *state sensitivity matrix* and its gradient w.r.t. the desired trajectory parameters. This then allows setting up an optimization problem in which the desired trajectory is optimized so as to minimize a suitable norm of the state sensitivity. The machinery has been applied to two case studies involving a unicycle and a planar quadrotor with successful results (monte-carlo statistical analysis). We are currently considering extensions of this initial idea (e.g., by also considering a notion of *input sensitivity*), as well as an experimental validation of the approach.

7.2.4. 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 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 extracted 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. We are now in the process of evaluating our motion control technique and exploring relevant applications, as well as preparing a publication summarizing this work.

7.2.5. 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. In this activity, we are first interested in collective pedestrian dynamics, i.e., how humans move in crowds, how they interact locally and how this results into the emergence of specific patterns at larger scales [52]. Virtual Reality is one main experimental tools in our approach, so as to control and reproduce easily situations we expose participants to, as well as to explore the nature of the visual cues human use to control their locomotion [22], [68]. We are also 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 [54], [26]. We directly exploit our experimental result to propose relevant navigation control techniques for robot to make them more adapted to move among humans [41], [58].

7.2.6. Direct Visual Servoing

Participants: Quentin Bateux, Eric Marchand.

We proposed a deep neural network-based method to perform high-precision, robust and real-time 6 DOF positioning tasks by visual servoing [53]. A convolutional neural network is fine-tuned to estimate the relative pose between the current and desired images and a pose-based visual servoing control law is considered to reach the desired pose. This approach efficiently and automatically creates a dataset used to train the network. We show that this enables the robust handling of various perturbations (occlusions and lighting variations). We then propose the training of a scene-agnostic network by providing both the desired and current images to a deep network for generating the camera motion. The method is validated on a 6 DOF robot.

7.2.7. Visual Servoing using Wavelet and Shearlet Transforms

Participants: Lesley-Ann Duflot, Alexandre Krupa.

We pursued our work on the elaboration of a direct visual servoing method in which the signal control inputs are the coefficients of a multiscale image representation [4]. In particular, we considered the use of multiscale image representations that are based on discrete wavelet and shearlet transforms. This year, we succeeded to derive an analytical formulation of the interaction matrix related to the wavelet and shearlet coefficients and experimentally demonstrated the performances of the proposed visual servoing approaches [18]. We also considered this control framework in a medical application which consists in automatically moving a biological sample carried by a parallel micro-robotic platform using Optical Coherence Tomography (OCT) as visual feedback. The objective of this application was to automatically retrieve the region of the sample that corresponds to an initial optical biopsy for diagnosis purpose. Experimental results demonstrated the efficiency of our approach that uses the wavelet coefficients of the OCT image as input of the control law to perform this task [61].

7.2.8. Visual Servoing from the Trifocal Tensor

Participants: Kaixiang Zhang, François Chaumette.

In visual servoing, three images are usually available at each iteration of the control loop: the very first one, the current one, and the desired one. That is why the trifocal tensor defined from this set of images is a potential candidate for providing visual features to be used as inputs of the control scheme. We have first modeled the interaction matrix related to the components of the trifocal tensor. We have then designed a set of reduced visual features with good decoupling properties, from which a thorough Lyapunov-based stability analysis has been developed [78].

7.2.9. 3D Steering of Flexible Needle by Ultrasound Visual Servoing

Participants: Jason Chevrie, Marie Babel, Alexandre Krupa.

Needle insertion procedures under ultrasound guidance are commonly used for diagnosis and therapy. However, it is often critical to accurately reach a targeted region due to the deflection of the flexible needle and the presence of intra-operative tissue motions. Therefore this year we improved our robotic framework dedicated to 3D steering of flexible needle that is based on ultrasound visual servoing. We developed a new control approach that both steers the flexible needle toward a desired target and compensates the tissue self-motion during the needle insertion. In our approach, the target to be reached by the needle is tracked in 2D ultrasound images and the needle tip position and orientation are measured by an electromagnetic tracker. Tissue motion compensation is performed using force feedback to reduce targeting error and forces applied to the tissue. The method also uses a mechanics-based interaction model that is updated online to provide the current shape of the deformable needle. In addition, a novel control law using task functions was proposed to fuse motion compensation, needle steering via manipulation of its base and steering of the needle tip in order to reach the target. Validation of the tracking and steering algorithms were performed in gelatin phantom and bovine liver on which periodical perturbation motions (magnitude of 15 mm) were applied to simulate physiological motions. Experimental results demonstrated that our approach can reach a moving target with an average targeting error of 1.2 mm and 2.5 mm in resp. gelatin and liver, which is accurate enough for common needle insertion procedures [12].

7.2.10. Robotic Assistance for Ultrasound Elastography by Visual Servoing, Force Control and Teleoperation

Participants: Pedro Alfonso Patlan Rosales, Alexandre Krupa.

Ultrasound elastography is an image modality that unveils elastic parameters of a tissue, which are commonly related with certain pathologies. It is performed by applying continuous stress variation on the tissue in order to estimate a strain map from successive ultrasound images. Usually, this stress variation is performed manually by the user through the manipulation of an ultrasound probe and it results therefore in an user-dependent quality of the strain map. To improve the ultrasound elastography imaging and provide quantitative measurements, we developed an assistant robotic palpation system that automatically applies the motion to a 2D or 3D ultrasound probe that is needed to generate in real-time the elastography images during teleoperation [5]. This year, we have extended our robotic framework by developing a method that provides to the user the capability to physically feel the stiffness of the observed tissue of interest via a haptic device. This work has been submitted to the ICRA 2019 conference.

7.2.11. Deformation Servoing of Soft Objects

Participant: Alexandre Krupa.

This year, we started a new research activity whose objective is to provide robotic control approaches that improve the dexterity of robots interacting with deformable objects. The goal is to control one or several robots interacting with a soft object in such a way to reach a desired configuration of object deformation. Nowadays, most of the existing deformation control methods require accurate models of the object and/or environment in order to perform such tasks. Contrarily to these methods, we want to propose model-free methods that rely only on visual observation provided by a RGB-D sensor to control the deformation of soft objects without a priori knowledge of their material mechanical parameters and without a priori knowledge of their environment. In a preliminary study, we compared the model-based method based on physics simulation (Finite Element Model) and the model-free method of the state of the art. We also developed a first approach

based on visual servoing that uses in the robot control law an online estimation of the interaction matrix that links the variation of the object deformation to the velocity of the robot end-effector. These different approaches have been implemented in simulation and are currently tested on a robotic arm (Adept Viper 650) interacting with a soft object (sponge). The first results are encouraging since they showed that our model-free visual servoing approach based on online estimation of the interaction matrix provides similar results than the model-based approach based on physics simulation.

7.2.12. Multi-Robot Formation Control

Participants: Paolo Robuffo Giordano, Fabrizio Schiano.

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 our previous works we have addressed the problem of coordinating a team of quadrotor UAVs equipped with onboard cameras from which one could extract "relative bearings" (unit vectors in 3D) w.r.t. the neighboring UAVs in visibility. This problem is known as bearing-based formation control and localization. In [71], we considered the localization problem for multi-robots (that is, the problem of reconstructing the relative poses from the available bearing measurements), by recasting it as a nonlinear observability problem: this rigorous analysis led us to introduce the notion of *Dynamic Bearing Observability Matrix*, which in a sense extends the classical Bearing Rigidity Matrix to explicitly account for the robot motion. It was then possible to show that the scale factor of the formation is, indeed, observable by processing the bearing measurements and (known) agent motion, a result confirmed experimentally by employing a EKF on a group of quadrotor UAVs. This and more results on bearing-based formation control and localization for quadrotor UAVs are summarized in [7].

7.2.13. Coupling Force and Vision for Controlling Robot Manipulators

Participants: François Chaumette, Paolo Robuffo Giordano, Alexander Oliva.

The goal of this recent 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. 6.6.4), 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. Our aim is therefore to advance the state-of-the-art in the field of torque-controlled manipulator arms by also including in the loop in an explicit way the use of a vision sensor. We will probably rely on estimation strategies for coping with the different rates of the two sensing modalities, and to online trajectory replanning strategies for dealing with constraints of the system (e.g., limited fov of the camera, of the fact that visibility of the target object is lost when closing in for grasping).

7.3. Haptic Cueing for Robotic Applications

7.3.1. Haptic Guidance of a Biopsy Needle

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 9.1.6). At the difference of our work presented in Section 7.2.9 where a robotic system is used to autonomously actuate the needle, we propose in this study another way of assistance for needle insertion. The principle is to provide haptic cue feedback to the clinician in order to help him during his manual gesture by the application of repulsive or attractive forces. The proposed solution is based on a shared robotic control, where the clinician and a haptic device, both holding the base of the needle, cooperate together. In a

preliminary study, 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. From this pre-operative information and intra-operative measurements of the location of the needle, haptic cues are generated to guide the clinician toward the desired needle position and orientation. These 5 different haptic guides were recently tested by 2 physicians, both experts in needle manipulation and compared to the reference gesture performed without assistance. The results have been submitted to the IPCAI 2019 conference. Future work will consist in evaluating the different haptic guides from an user-experience study involving more participants.

7.3.2. Wearable Haptics

Participants: Marco Aggravi, Claudio Pacchierotti, Paolo Robuffo Giordano.

We worked on a wearable haptic device for the forearm and its application in robotic teleoperation [8]. The device is able to provide skin stretch, pressure, and vibrotactile stimul, see Fig. 10 . Two servo motors, housed in a 3D printed lightweight platform, actuate an elastic fabric belt, wrapped around the arm. When the two servo motors rotate in opposite directions, the belt is tightened (or loosened), thereby compressing (or decompressing) the arm. On the other hand, when the two motors rotate in the same direction, the belt applies a shear force to the arm skin. Moreover, the belt houses four vibrotactile motors, positioned evenly around the arm at 90 degrees from each other. The device weights 220 g for 115×122×50 mm of dimensions, making it wearable and unobtrusive. We carried out a perceptual characterization of the device as well as two human-subjects teleoperation experiments in a virtual environment, employing a total of 34 subjects.

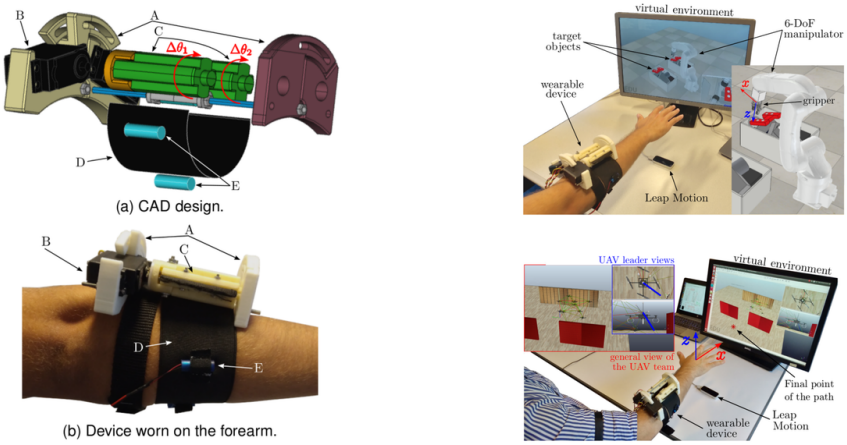


Figure 10. The proposed wearable device for the arm and its evaluation. The device consists of a static platform (A) that accommodates two servomotors (B) and two pulleys (C), a fabric belt (D), and four vibrotactile motors (E).

In the first experiment, participants were asked to control the motion of a robotic manipulator for grasping an object; in the second experiment, participants were asked to teleoperate the motion of a quadrotor fleet along a given path. In both scenarios, the wearable haptic device provided feedback information about the status of the slave robot(s) and of the given task. Results showed the effectiveness of the proposed device. Performance on completion time, length trajectory, and perceived effectiveness when using the wearable device improved of 19.8%, 25.1%, and 149.1% than when wearing no device, respectively. Finally, all subjects but three preferred the conditions including wearable haptics.

7.3.3. 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, 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. Moreover, we want to use robotic manipulators to enlarge the workspace of mid-air haptic systems, using depth cameras and visual servoing techniques to follow the motion of the user's hand.

7.3.4. Haptic Cueing in Telemanipulation

Participants: Firas Abi Farraj, Paolo Robuffo Giordano, Claudio Pacchierotti.

Robotic telemanipulators are already widely used in nuclear decommissioning sites for handling radioactive waste. However, currently employed systems are still extremely primitive, making the handling of these materials prohibitively slow and ineffective. As the estimated cost for the decommissioning and clean-up of nuclear sites keeps rising, it is clear that one would need faster and more effective approaches. Towards this goal, we presented the user evaluation of a recently proposed haptic-enabled shared-control architecture for telemanipulation [51]. An autonomous algorithm regulates a subset of the slave manipulator degrees of freedom (DoF) in order to help the human operator in grasping an object of interest. The human operator can then steer the manipulator along the remaining null-space directions with respect to the main task by acting on a grounded haptic interface. The haptic cues provided to the operator are designed in order to inform about the feasibility of the user's commands with respect to possible constraints of the robotic system. This work compared this shared-control architecture against a classical 6-DOF teleoperation approach in a real scenario by running experiments with 10 subjects. The results clearly show that the proposed shared-control approach is a viable and effective solution for improving currently-available teleoperation systems in remote telemanipulation tasks.

7.3.5. Haptic Feedback for an Augmented Wheelchair Driving Experience

Participants: Louise Devigne, Marie Babel, François Pasteau.

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 have then proposed a framework in which feedback is provided by sending forces through the wheelchair controller as the user steers the wheelchair. This solution is based on a low complex optimization framework able to perform smooth trajectory correction and to provide obstacle avoidance. The impact of the proposed haptic guidance solution on user driving performance was assessed during this pilot study for validation purposes through an experiment with 4 able-bodied participants. Results of this pilot study showed that the number of collisions significantly decreased while force feedback was activated, thus validating the proposed framework [60].

7.3.6. Virtual Shadows to Improve Self Perception in CAVE

Participants: Guillaume Cortes [Hybrid], Eric Marchand.

In immersive projection systems (IPS), the presence of the user's real body limits the possibility to elicit a virtual body ownership illusion. But, is it still possible to embody someone else in an IPS even though the users are aware of their real body ? In order to study this question, we propose to consider using a virtual shadow in the IPS, which can be similar or different from the real user's morphology. We have conducted an experiment ($N = 27$) to study the users' sense of embodiment whenever a virtual shadow was or was not present. Participants had to perform a 3D positioning task in which accuracy was the main requirement. The results showed that users widely accepted their virtual shadow (agency and ownership) and felt more comfortable when interacting with it (compare to no virtual shadow). Yet, due to the awareness of their real

body, the users have less acceptance of the virtual shadow whenever the shadow gender differs from their own. Furthermore, the results showed that virtual shadows increase the users' spatial perception of the virtual environment by decreasing the inter-penetrations between the user and the virtual objects. Taken together, our results promote the use of dynamic and realistic virtual shadows in IPS and pave the way for further studies on "virtual shadow ownership" illusion.

7.4. Shared Control Architectures

7.4.1. Shared Control for Remote Manipulation

Participants: Firas Abi Farraj, Paolo Robuffo Giordano, Claudio Pacchierotti, Rahaf Rahal, Mario Selvaggio.

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, we worked at different approaches:

- We proposed novel haptic guidance methods for a dual-arm telerobotic manipulation system [36] which are able to deal with several different constraints, such as collisions, joint limits, and singularities. We combined the haptic guidance with shared-control algorithms for autonomous orientation control and collision avoidance meant to further simplify the execution of grasping tasks. In addition, a human subject study was carried out to assess the effectiveness and applicability of the proposed control approaches both in simulated and real scenarios. Results showed that the proposed haptic-enabled shared-control methods significantly improve the performance of grasping tasks with respect to the use of classic teleoperation with neither haptic guidance nor shared control. Live demos of some of these approaches have been shown to the general public at the Maker Faire 2018 in Rome.
- In the framework of the RoMANS H2020 project, we worked together with CEA to implement an intuitive and effective shared-control teleoperation system with haptic feedback using the CEA robotic hand at the slave side and the Haption glove at the master side. The system was tested at CEA in an object grasping, manipulation, and sorting scenario. A video of the experiment is available at: https://youtu.be/M-tpVP9Fakc.
- Finally, in [38] we reported the results of a collaborative project involving LAAS-CNRS as leader, where we implemented an aerial-ground comanipulator system, denoted *Tele-MAGMaS*, where a fixed-based manipulator arm cooperates with a UAV equipped with an onboard gripper for carrying together (and manipulating) a long bar. The system has been demonstrated live during the Hannover Fair in 2017.

7.4.2. Shared Control for Mobile Robot Navigation

Participant: Paolo Robuffo Giordano.

Besides manipulators, we also considered shared control algorithms for mobile robot navigation. In [25], we have presented (and experimentally validated) an online trajectory planning approach that allows a human operator to act on the trajectory to be tracked by a mobile robot (a quadrotor UAV in the experiments) in conjunction with the robot autonomy which can locally modify the planned trajectory for avoiding obstacles of staying close to points of interest. This "shared planning approach" is quite general and its application to other robotic systems is under investigation.

7.4.3. 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 encountering *negative* obstacles, dealing with uneven grounds, etc. Indeed, differences of elevation have been reported to be one of the most challenging environmental barrier to negotiate while driving a wheelchair with tipping and falling being is the most common accidents power wheelchair users encounter. It is thus our actual 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 first shared-control algorithm which provides assistance while navigating with a wheelchair in an environment consisting of negative obstacles [80].

7.4.4. Wheelchair Kinematics and Dynamics Modeling for Shared Control

Participants: Aline Baudry, Marie Babel.

The driving experience of an electric powered wheelchair can be disturbed by the dynamic and kinematic effects of the passive caster wheels, particularly during maneuvers in narrow rooms and direction changes. In order to prevent their nasty behaviour, we proposed a caster wheel behavior model based on experimental measurements. The study has been realised for the three existing types of wheelchair, which present different kinematic behaviors, i.e. front caster type, rear caster type and mid-wheel drive. The orientation of the caster wheels has been measured experimentally for different initial orientations, velocities and user mass, according to a predefined experimental design. The repeatability of the motions has been studied, and from these measurements, their behavior has been modeled. By using this model with the wheelchair kinematic expressions, we were able to calculate the real trajectory of the wheelchair to enhance an existing driving assistance for powered wheelchair [79].

7.4.5. Wheelchair Autonomous Navigation for Fall Prevention

Participants: Solenne Fortun, Marie Babel.

The Prisme project (see Section 9.1.7) 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 designed the related control framework based on sensor-based servoing principles and validated it in simulation. Next step will consist in realizing tests within the Rehabilitation Center of Pôle Saint Hélier.

7.4.6. Robot-Human Interactions during Locomotion

Participants: Julien Legros, Javad Amirian, Fabien Grzeskowiak, Ceilidh Hoffmann, Marie Babel, Jean Bernard Hayet, 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. We are currently developing these ideas, which should bring good results in the near future.

RITS Project-Team

7. New Results

7.1. Deep Reinforcement Learning for end-to-end driving

Participants: Raoul de Charette, Maximilian Jaritz, Fawzi Nashashibi.

Following the work initiated in 2017, we continued the work on end-to-end driving using with asynchronous reinforcement learning directly. The network learns to map low level control directly with RGB images. To continue previous works initiated, we have applied recent domain adaptation and evaluated our reinforcement learning (learn in a realistic car game) in open-loop on real video footage, showing promising adaptation results. New outcome also include tests on real data (web footage). This led to a publication in ICRA [25]. This research was partially funded by Valeo.

7.2. Convolutional neural networks for Semantic and Completion with Sparse and Dense Data

Participants: Raoul de Charette, Maximilian Jaritz, Fawzi Nashashibi.

Deep convolutional networks have outperform all previous techniques on most vision tasks. This is because they are able to utilize dense data and extract relationship between local information such as gradients, or high level features. However, convolutional neural networks (CNNs) require dense data and are known to fail when data is sparse. Here, we address the research problem and proposed a solution. Instead of using a sparse convolution methodology, we show that using the right architecture with a proper training strategy the network can learn sparsity invariant feature while remaining stable when dense data are present. Our architecture uses an encoder-decoder version of Mobile NasNet with skip connections. The results show that we can accomplish both data completion or semantic segmentation changing only the last layer of the network. Performance obtained were published on Kitti Benchmark and ranks among the first ones, and the methodology was published in 3DV [26]. This research was partially funded by Valeo.

7.3. Realistic Weather Augmentation for Evaluation of Bad Weather in Computer Vision

Participants: Raoul de Charette, Shirsendu Halder.

Computer vision is evaluated on extensive databases that include large number of examples and allow the ranking of algorithms. However, all databases are acquired in clear weather conditions, where the atmosphere is a transparent medium. In rain/snow/fog, when the atmosphere is filled with particles the light is refracted/reflected/diffracted and the appearance is altered. Here we propose a new research that augment existing databases with new weather or arbitrary amount. We applied it on Kitti and Cityscapes. Our approach uses an accurate understanding of physical and optics models to generate realistic rain/fog and augment existing images or sequences. This allows us to evaluate state-of-the-art vision algorithms for both object detection and semantics and quantitatively measure the effect of rain or fog on them. This research was conducted in collaboration with Jean-Francois Lalonde from Université Laval and was supported by Samuel de Champlain Quebec-France collaboration program.

7.4. Perception for Cooperative Driving

Participants: Pierre Bourre, Raoul de Charette, Carlos Flores, Renaud Poncelet, Luis Roldao, Dinh-Van Nguyen.

In the context of multiple autonomous vehicles, sharing the perception of each other allows an enriched perception of the environment. For the PACV2x FUI project, we propose a mix of vision sensors and communication exchanges is used for merging, overtaking, and other risky situations that benefit from multi perception. A speed planning algorithm as well as low level control and lidar data clustering were developed to allow a small fleet of two to three vehicles to handle such scenarios. The vehicles use communication and GPS coordinates to closely follow a planned trajectory.

7.5. A Statistical Update of Grid Representations from Range Sensors

Participants: Luis Roldao, Raoul de Charette, Anne Verroust-Blondet.

An accurate 3D model of the surrounding environment is a fundamental feature for autonomous vehicles to perform different tasks such as obstacle detection, localization and mapping. While continuous representations are widely used in the literature, we prefer to use a three dimensional discrete grid representation in this work in order to reduce memory an computational complexity. In this case, each grid cell represents the occupancy state of a portion of the environment in a probabilistic manner.

By definition, a discretized representation inhibits a completely accurate reconstruction. Therefore, grid models are unable to create a perfect model of the surroundings. In the literature, it is usually considered that within a single scan, the state of each cell is binary (free or occupied). Hence, a cell is set occupied if at least one impact occurred within, and free if it has been traversed by any ray. The problem of such an approach is that the complete state of the cell is updated from a single partial observation, neglecting the contribution of multiple measurements and their validity. Moreover, the traversed distance of the rays within each cell is usually ignored.

Towards the goal of achieving a more accurate representation, we propose a different way to update the occupancy probability of each cell according to the observations; considering the traversed distance of the rays within each cell (ray-path information), the contribution of the complete set of observations within the cell, and the density of observations that can be obtained at such cell according to its distance from the sensor. Proposed method was evaluated in both simulated and real data. Reconstruction results show an improvement on the representation of the surroundings with less occupancy state errors in the cells of the grid. Future works will include the comparison against a continuous representation to test the accuracy along with the time and computation needs for both representations.
More details can be found in [38] and [30]. This research is partially funded by AKKA Technology.

7.6. Recognizing Pedestrians using Cross-Modal Convolutional Networks

Participants: Danut-Ovidiu Pop, Fawzi Nashashibi.

This year, we have continued our research, which is based on multi-modal image fusion schemes with deep learning classification methods. We propose four different learning patterns based on Cross-Modality deep learning of Convolutional Neural Networks:

(1) a Particular Cross-Modality Learning;

(2) a Separate Cross-Modality Learning;

(3) a Correlated Cross-Modality Learning and

(4) an Incremental CrossModality Learning model.
Moreover, we also design a new variation of a Lenet architecture, which improves the classification performance. Finally, we propose to learn this model with the incremental cross-modality approach using optimal learning settings, obtained with a K-fold Cross Validation pattern. This method outperforms the state-of-the-art classifier provided with Daimler datasets on both non-occluded and partially-occluded pedestrian tasks.

7.7. Vehicle Trajectory Prediction

Participants: Kaouther Messaoud, Itheri Yahiaoui, Anne Verroust-Blondet, Fawzi Nashashibi.

In order to enhance the road safety, the first and the most important step for an effective autonomous navigation is the environment perception and surrounding objects recognition. So, advanced sensing systems are mounted in vehicles to monitor the on-road environment. One of the most challenging tasks is to understand, analyze the driving situations and make a reasonable and safe navigation decisions accordingly. Human drivers make decisions while implicitly reasoning about how neighboring drivers will move in the future. In this context, we aim to predict the motion of drivers neighboring an autonomous vehicle based on data captured using deployed sensors.
This year, we studied the state of the art approaches for trajectory and maneuver prediction. We focused on general trajectory prediction representation while considering interactions between the neighboring drivers using different types of neural networks such as recurrent and convolutional neural networks.

7.8. WiFi Fingerprinting Localization for Intelligent Vehicles in Car Park

Participants: Dinh-Van Nguyen, Raoul de Charette, Fawzi Nashashibi.

A novel method of WiFi fingerprinting for localizing intelligent vehicles in GPS-denied area, such as car parks, has been proposed. Although the method itself is a popular approach for indoor localization application, adapting it to the speed of vehicles requires different treatment. By deploying an ensemble neural network for fingerprinting classification, the method shows a reasonable localization precision at car park speed. Furthermore, a Gaussian Mixture Model (GMM) Particle Filter is applied to increase localization frequency as well as accuracy. Experiments show promising results with average localization error of 0.6m (cf. [29]).

A more complete study on the use of Wifi fingerprinting for solving the localization problem for autonomous vehicles in GPS-denied environments is presented in the thesis manuscript entitled "Wireless Sensors Networks for Indoor Mapping and Accurate Localization for Low Speed Navigation in Smart Cities" (cf. [11]).

7.9. Enhancing the Accuracy of SLAM-based Localization Systems for Autonomous Driving

Participants: Zayed Alsayed, Anne Verroust-Blondet, Fawzi Nashashibi.

Computing a reliable and accurate pose for a vehicle in any situation is one of the challenges for Simultaneous Localization And Mapping methods (SLAM) methods. Based on the probabilistic form of the SLAM solution, SLAM methods suffer from systematic errors related to the linearization of the solution models. The accuracy of the SLAM method can be improved by estimating a correction to be applied to the SLAM output based on relevant information available from the SLAM algorithm. In [20] two approaches predicting corrections to be applied to SLAM estimations are proposed:

1) The first approach is designed for 2D SLAM methods, i.e. independently of the underlying SLAM process and sensor used, where we aim to reduce the errors due to the dynamical modeling during specific maneuvers.

2) The second method is designed to handle errors related to the probabilistic formulation of Maximum Likelihood SLAM approaches, and thus it is suitable for 2D Maximum Likelihood SLAM methods (i.e. no assumptions on the sensor used).
The validity of both approaches was proved through two experiments using different evaluation metrics and using different sensor characteristics.
More detail can be found in the thesis manuscript of Zayed Alsayed entitled "Characterizing the Robustness and Enhancing the Accuracy of SLAM-based Localization Systems for Autonomous Driving" (cf. [7]).

7.10. LIDAR-based lane marking detection for vehicle localization

Participants: Farouk Ghallabi, Fawzi Nashashibi.

Accurate self-vehicle localization is an important task for autonomous driving and ADAS. Current GNSS-based solutions do not provide better than 2-3 m in open-sky environments. In order to achieve lane-level accuracy, a lane marking detection system using a multilayer LIDAR (velodyne) and a map matching algorithm has been introduced. The perception system includes three different steps: road segmentation, image construction and line detection. Our road segmentation method purely relies on geometric analysis of each layer returns. Detected lane markings are matched to a prototype third party map which was built with absolute accuracy = 5cm. The map matching algorithm is a particle filtering process that achieves lane-level accuracy (20 cm). More details are in [23]. This work has been partially funded by Renault.

7.11. Motion Planning among Highly Dynamic Obstacles

Participants: Pierre de Beaucorps, Anne Verroust-Blondet, Renaud Poncelet, Fawzi Nashashibi.

Motion planning in a dynamic environment is of great importance in many robotics applications. In the context of an autonomous mobile robot, it requires computing a collision-free path from a start to a goal among moving and static obstacles. We have introduced a framework to integrate into a motion planning method the interaction zones of a moving robot with its future surroundings, the reachable interaction sets (RIS). It can handle highly dynamic scenarios when combined with path planning methods optimized for quasi-static environments. It has been integrated with an artificial potential field reactive method and with a Bézier curve path planning. Experimental evaluations show that this approach significantly improves dynamic path planning methods, especially when the speeds of the obstacles are higher than the one of the robot (cf. [32] for more detail). This work has been partially funded by Valeo.

7.12. Control Architecture for Adaptive and Cooperative Car-Following

Participants: Carlos Flores, Fawzi Nashashibi.

The general scope of this work deals with three open challenges in the state-of-the-art of cooperative car-following systems:

1) Deal with the impact of not only communication links delays, but also heterogeneity between vehicles' dynamics in the same string. This should be targeted ensuring the gap-regulation robustness without degrading the expected performance to keep car-following benefits (individual and string stability). In particular, when a heterogeneous string is formed, the differences between vehicles dynamics introduce disturbances in the closed loop system affecting the string stability. In [22] we presented an online Cooperative Adaptive Cruise Control (CACC) feedforward adaptation with a fractional-order feedback controller for stable heterogeneous strings of vehicles. Simulations demonstrate the advantages over conventional homogeneous structures as well as system's capability to both enhance stability and guarantee string stability regardless the vehicles distribution.

2) Design a modular architecture that permits to introduce cooperative string driving in urban environments, where interaction with vulnerable road users is highly probable. In this context, a cooperative car-following/emergency braking system with prediction-based pedestrian avoidance capabilities using vehicle-to-vehicle and vehicle-to-pedestrian communication links has been proposed in [14] and validated with RITS platforms.

3) Further extend the benefits of Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC) applications on traffic flow and safety, having strict $\mathcal{L}_2$ string stability as a hard constraint, employing different calculus techniques for the control design task. A fractional-order-based control algorithm is employed to enhance the car-following and string stability performance for both ACC and CACC vehicle strings, including communication temporal delay effects has been presented in [15]. Simulation and real experiments have been conducted for validating the approach.

The aforementioned contributions have been developed in the framework of the VALET project ANR-15-CE22-0013. They have been also implemented in the vehicle platforms of RITS team, for the sake of validation and further demonstration of the final VALET system.

This scientific work can be found as well in the thesis manuscript of Carlos Flores entitled "Control Architecture for Adaptive and Cooperative Car-Following" (cf. [8]).

7.13. Stability analysis for controller switching in autonomous vehicles

Participants: Francisco Navas, Imane Mahtout, Fawzi Nashashibi.

This work investigates the Youla-Kucera (YK) parameterization to provide stable responses for autonomous vehicles when dynamics or environmental changes occur. This work explores the use of the YK parameterization in dynamics systems such as vehicles, with special emphasis on stability when some dynamic change or the traffic situation demands controller reconfiguration:

- YK parameterization provides all stabilizing controllers for a given plant. This is used in order to perform stable controller reconfiguration. Different YK-based control structures are obtained for dealing with problems such order complexity, plant disconnection or matrix inversability. Stability properties are preserved even if different structures are employed, but transient behavior between controllers changes depending on the employed YK-based structure. One of the structures presents the best transient behavior without oscillations, a lower order controller complexity and no need to disconnect the initial controller, which would be important if the system shutdown is very expensive, or the initial controller is part of a safety circuit [28]. This structure is used together with CACC applications improving CACC state-of-the-art. An hybrid behavior between two CACC controllers with different time gaps is explored by means of the YK parameterization, in order to avoid ACC degradation when communication link with preceding vehicle is lost. The proposed system uses YK parameterization and communication with a vehicle ahead (different from the preceding one) providing stable responses and, more interestingly, reducing intervehicle distances in comparison with an ACC degradation. A similar idea of hybrid behavior between CACC controller with different time gap is developed for entering/exiting vehicles in the string. In that case, YK parameterization is able to ensure stability of these merging/splitting maneuvers.
- Dual YK parameterization provides all the plants stabilized by a controller. This is employed for solving CL identification problems, or adaptive control solutions, which integrate identification and controller reconfiguration processes. YK-based CL identification uses classical OL identification algorithms, providing better results than if it is used alone. Results in a CACC-equipped vehicle prove how CL nature of the data affects a classical OL identification algorithm, and how dual YK parameterization helps to mitigate these effects. Finally, an adaptive control application is developed by using MMAC. Longitudinal dynamics of two vehicles in a CACC string are estimated within a model set, so the good CACC sytem can be chosen even if a heterogeneous string of vehicles is considered. Dynamics estimation results much more faster than other estimation processes in the literature.
- Different types of controllers and structures are used throughout Francisco Navas thesis ([10]), proving the adaptability of the YK parameterization to any type of controller. Simulation and experimental results demonstrate real implementation of stable controller reconfiguration, CL identification and adaptive control solutions dealing with dynamics changes or different traffic situations. The author thinks that YK is a suitable control framework able to ensure responses in autonomous driving.
- In [27] a design and implementation of a novel lateral control approach is proposed within Imane Mahtout thesis work. The control strategy is based on Youla-Kucera parametrization to switch gradually between controllers that are designed separately for big and small lateral errors. The presented approach studies the critical problem of initial lateral error in line following. It ensures smooth and stable transitions between controllers and provides a smooth vehicle response regardless of the lateral error. For an initial validation the work was tested in simulation, implementing a dynamic bicycle model. It has also been tested in real platforms implementing an electric Renault ZOE, with good results when activating the system at different lateral errors. Current work is focused on using YK-parametrization in estimating lateral vehicle dynamics.

7.14. Belief propagation inference for traffic prediction

Participant: Jean-Marc Lasgouttes.

This work [45], [44], 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.

This year, continuing our collaboration with PTV Sistema, we improved our techniques and 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 [35].

7.15. Large scale simulation interfacing

Participant: Jean-Marc Lasgouttes.

The SINETIC FUI project aims to build a complete simulation environment handling both mobility and communication. We are interested here in a so-called system-level view, focusing on simulating all the components of the system (vehicle, infrastructure, management center, etc.) and its realities (roads, traffic conditions, risk of accidents, etc.). The objective is to validate the reference scenarios that take place on a geographic area where a large number of vehicles exchange messages using the IEEE 802.11p protocol. This simulation tool is done by coupling the SUMO microscopic simulator and the ns-3 network simulator thanks to the simulation platform iTETRIS.

We have focused in this part of the project on how to reduce the execution time of large scale simulations. To this end, we designed a new simulation technique called Restricted Simulation Zone which consists on defining a set of vehicles responsible of sending the message and an area of interest around them in which the vehicles receive the packets [31].

7.16. Platoons Formation for autonomous vehicles redistribution

Participants: Mohamed Hadded, Jean-Marc Lasgouttes, Fawzi Nashashibi, Ilias Xydias.

In this paper, we consider the problem of vehicle collection assisted by a fleet manager where parked vehicles are collected and guided by fleet managers. Each platoon follows a calculated and optimized route to collect and guide the parked vehicles to their final destinations. The Platoon Route Optimization for Picking up Automated Vehicles problem, called PROPAV, consists in minimizing the collection duration, the number of platoons and the total energy required by the platoon leaders. We propose a formal definition of PROPAV as an integer linear programming problem, and then we show how to use the Non-dominated Sorting Genetic Algorithm II (NSGA-II), to deal with this multi-criteria optimization problem. Results in various configurations are presented to demonstrate the capabilities of NSGA-II to provide well-distributed Pareto-front solutions.

This work has been presented at ITSC 2018 conference [24].

7.17. Prediction-based handover between VLC and IEEE 802.11p for vehicular environment

Participants: Mohammad Abualhoul, Fawzi Nashashibi.

Despite years of development and deployment, the standardized IEEE 802.11p communication for vehicular networks can be pushed toward insatiable performance demands for wireless network data access, with a remarkable increase of both latency and channel congestion levels when subjected to scenarios with a very high vehicle density.

In specific hard safety applications such as convoys, IEEE 802.11p could seriously fail to meet the fundamental vehicular safety requirements. On the other hand, the advent of LED technologies has opened up the possibility of leveraging the more robust Visible Light Communication (VLC) technology to assist IEEE 802.11p and provide seamless connectivity in dense vehicular scenarios.

In this particular research, we proposed and validated a Prediction-based Vertical handover (PVHO) between VLC and IEEE 802.11p meant to afford seamless switching and ensure the autonomous driving safety requirements [19].

Algorithm validation and platoon system performance were evaluated using a specially implemented IEEE 802.11p-VLC module in the NS3 Network Simulator. The simulation results showed a speed-based dynamic redundancy before and after VLC interruptions with seamless switching. Moreover, the deployment of VLC for platoon intra-communication can achieve a 10-25% PDR gain in high-density vehicular scenarios, where the work was published in the IEEE International Conference on Intelligent Transportation Systems 2018.

7.18. Lane-Centering to Ensure the Visible Light Communication (VLC) Connectivity for a Platoon of Autonomous Vehicles

Participants: Mohammad Abualhoul, Fawzi Nashashibi.

VLC technology limitations were defined and supported by different solutions proposals to enhance the crucial alignment and mobility limitations. In this research [17], we proposed the incorporation of the VLC technology and a Lane-Centering (LC) technique to assure the VLC-connectivity by keeping the autonomous vehicle aligned to the lane center using vision-based lane detection in a convoy-based formation. Such combination can ensure the optical communication connectivity. This contribution by RITS-Team won the best paper award during the ICVES conference.

The system performance and evaluation showed that ss soon as the road lanes are detectable, the evaluated results showed stable behavior independently from the inter-vehicle distances and without the need for any exchanged information of the remote vehicles. Further investigations are to be carried-out in this direction.

7.19. Cyberphysical Constructs for Next-Gen Vehicles and Autonomic Vehicular Networks

Participant: Gérard Le Lann.

Behaviors of Connected Automated Vehicles (CAVs) rest on robotics capabilities (sensors, motion control laws, actuators) and wireless radio communications. Reduction of non-harmful crashes and fatalities despite higher vehicular density (safety and efficiency properties) is a fundamental objective, whatever the SAE automated driving levels considered (use cases).

Based on "hard sciences", onboard robotics capabilities designed so far are satisfactory for numerous settings, to the exception of non-line-of-sight scenarios. That is the rationale for wireless radio communications. Over the years, a growing fraction of the scientific community has been questioning the adequacy of current IEEE and ETSI standards aimed at automotive wireless communications, herein referred to as wave protocols (wireless access in vehicular environment) for convenience.

Analyses based on well-known results in various areas such as life/safety-critical systems, distributed algorithms, dependable real-time computing, ad hoc mobile networking, and cyber-physics (to name a few) come to the conclusion that wave protocols do not meet essential requirements regarding safety, efficiency, privacy or cybersecurity (SPEC). These conclusions are based on scientific demonstrations. Notably, wave protocols rest on intuitive designs (no proofs, only simulations or experimental testing) that violate well-known impossibility results in asynchronous or synchronous systems. It follows that future vehicles shall be commanded and controlled by onboard robotics supplemented with wireless communication capabilities other than wave protocols. These vehicles are referred to as Next-Gen Vehicles (NGVs) in order to avoid confusion with CAVs.

That wave solutions are far from being convincing is at the core of the recommendations issued at the EU level (the latest WG29 resolution). Moreover, the important question of how to instantiate the EU GDPR directive in future CAVs is left unanswered, despite the fact that it is possible (proofs provided) to achieve safety and privacy jointly. Preliminary results for NGVs have appeared in [34].

The work reported herein, started in 2017 along with international researchers, aims at specifying solutions to the SPEC problem, considering self-organizing and self-healing Autonomic Vehicular Networks (AVNs) of NGVs. Parallel to this, risks of privacy breaches and cyberattacks proper to wave solutions have been exposed to the public via invited interventions and presentations.

An issue not very well addressed so far is to which extent robotics and computer science supplement each other. The cyber-physical perspective is essential to formulate a coherent vision. In cyber space and in physical space, safety has to do with resource sharing. Deadlock-free and fair resource sharing in systems of concurrent processes has been a major topic in computer science for more than 50 years. Asphalt (2D systems), asphalt and air space (3D systems) are the shared resources of interest in the physical space.

As is well known, there are three classes of algorithmic solutions: detection-and-recovery, prevention, avoidance. The former class is inapplicable (one cannot "roll back an accident"). Prevention is aimed at prohibiting the emergence of hazardous (no safety) or deadlock-prone (no safety, no efficiency) conditions. Solutions are the province of distributed algorithms (computer science). Avoidance is relied on for maintaining non-hazardous conditions while making progress (also, in case some of the assumptions that underlie prevention schemes would be violated). Solutions are the province of automation control (linear/non-linear dynamics).

Prevention and avoidance schemes are needed, put in action as follows: NGVs run (cyber) distributed agreement algorithms in order to preclude the emergence of hazardous conditions, prior to executing physical motions (collision-free trajectories), which motions are made feasible thanks to prevention schemes. This is how computer science and robotics can be "married" consistently: with prevention schemes, one achieves proactive safety, and with avoidance schemes, one achieves reactive safety (both types are needed).

NGVs and AVNs are life/safety-critical cyber-physical systems. Consequently, correct solutions to the SPEC problem are based on cyber-physical constructs endowed with appropriate intrinsic properties. We have devised the cell and the cohort constructs, which rest on the obvious observation according to which only vehicles sufficiently close to each other may experience a collision. Time-bounded ultra-fast message-passing and inter-vehicular coordination can be achieved within these constructs thanks to very short-range radio and optical communications, as well as deterministic protocols (MAC protocols in particular) and distributed algorithms (dissemination, approximate agreement, and consensus). Analytical expressions of upper bounds for message-passing and inter-vehicular coordination are established for worst-case conditions, such as contention and failures, message losses in particular. We have shown that these solutions can sustain message loss frequencies an order of magnitude higher than frequencies beyond which none of the wave protocols could work.

We have defined the concept of cyberphysical levels, which are orthogonal to SAE automated driving levels. Joining a cohort longitudinally or laterally (which implies a lane change) is conditioned on a number of criteria, such as cyberphysical levels, NGV sizes, and proof of authentication (requestor's name must be a certified pseudonym).

Naming raises open problems in spontaneous mobile open systems, such as AVNs. Privacy-preserving naming is even more complex. The "longitudinal privacy-preserving naming" problem is solved with the cohort construct. The "lateral privacy-preserving naming" problem which arises with NGVs members of a cell or that circulate in adjacent cohorts has solutions based on combined optical and radio communications.

Novel deterministic time-bounded MAC protocols at the core of distributed coordination algorithms are needed to solve the open problem of safe entrances into unsignalized intersections of arbitrary topologies (any number of arterials, any number of lanes per arterial) in the presence of noisy radio channels. This problem has been solved with CSMA-CD/DCR (deterministic collision resolution) MAC protocols.

7.20. Functional equations

Participant: Guy Fayolle.

The article [13] presents functional equations (involving one or two complex variables) as an Important analytic method in stochastic modelling and in combinatorics.

7.21. 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, in general, 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, recalling that the speed of convergence is exponential, a classical property deduced from the general theory of Markov chains.
3. The problem of generating rare events by mixing Gibbs samplers, Large Deviation Techniques (LDT) and cross-entropy method 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).

7.22. 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. Among other things, an important quantity, the so-called *group of the walk*, allows to deduce theoretical features about the nature of the solutions. In particular, when the *order* of the group is finite and the underlying algebraic curve is of genus 0 or 1, necessary and sufficient conditions have been given for the solution to be rational, algebraic or D-finite (i.e. solution of a linear differential equation). 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 of future research deals with the connections between simple product-form stochastic networks (so-called *Jackson networks*) and explicit solutions of functional equations for counting lattice walks.

LINKMEDIA Project-Team

6. New Results

6.1. Low-level content description and indexing

6.1.1. Scalability of the NV-tree: Three Experiments

Participants: Laurent Amsaleg, Björn Þór Jónsson [Univ. Copenhagen], Herwig Lejsek [Videntifier Tech.].

The NV-tree is a scalable approximate high-dimensional indexing method specifically designed for large-scale visual instance search. We report in [10] on three experiments designed to evaluate the performance of the NV-tree. Two of these experiments embed standard benchmarks within collections of up to 28.5 billion features, representing the largest single-server collection ever reported in the literature. The results show that indeed the NV-tree performs very well for visual instance search applications over large-scale collections.

6.1.2. Prototyping a Web-Scale Multimedia Retrieval Service Using Spark

Participants: Laurent Amsaleg, Gylfi Þór Gudmundsson [School of Computer Science, Reykjavik], Björn Þór Jónsson [Univ. Copenhagen], Michael Franklin [Computer Science Division, Berkeley].

The world has experienced phenomenal growth in data production and storage in recent years, much of which has taken the form of media files. At the same time, computing power has become abundant with multi-core machines, grids, and clouds. Yet it remains a challenge to harness the available power and move toward gracefully searching and retrieving from web-scale media collections. Several researchers have experimented with using automatically distributed computing frameworks, notably Hadoop and Spark, for processing multimedia material, but mostly using small collections on small computing clusters. In [3] we describe a prototype of a (near) web-scale throughput-oriented MM retrieval service using the Spark framework running on the AWS cloud service. We present retrieval results using up to 43 billion SIFT feature vectors from the public YFCC 100M collection, making this the largest high-dimensional feature vector collection reported in the literature. We also present a publicly available demonstration retrieval system, running on our own servers, where the implementation of the Spark pipelines can be observed in practice using standard image benchmarks, and downloaded for research purposes. Finally, we describe a method to evaluate retrieval quality of the ever-growing high-dimensional index of the prototype, without actually indexing a web-scale media collection.

6.1.3. Extreme-value-theoretic estimation of local intrinsic dimensionality

Participants: Laurent Amsaleg, Teddy Furon, Oussama Chelly [National Institute of Informatics], Stéphane Girard [MISTIS, Inria Grenoble], Michael Houle [National Institute of Informatics], Ken-Ichi Kawarabayashi [National Institute of Informatics], Michael Nett [Google].

This work is concerned with the estimation of a local measure of intrinsic dimensionality (ID) recently proposed by Houle. The local model can be regarded as an extension of Karger and Ruhl's expansion dimension to a statistical setting in which the distribution of distances to a query point is modeled in terms of a continuous random variable. This form of intrinsic dimensionality can be particularly useful in search, classification, outlier detection, and other contexts in machine learning, databases, and data mining, as it has been shown to be equivalent to a measure of the discriminative power of similarity functions. Several estimators of local ID are proposed and analyzed based on extreme value theory, using maximum likelihood estimation, the method of moments, probability weighted moments, and regularly varying functions, see [2]. An experimental evaluation is also provided, using both real and artificial data.

6.1.4. Intrinsic dimensionality for Information Retrieval

Participant: Vincent Claveau.

Examining the properties of representation spaces for documents or words in Information Retrieval (IR) brings precious insights to help the retrieval process. Following the work presented in the previous paragraph, it has been shown that intrinsic dimensionality is chiefly tied with the notion of indiscriminateness among neighbors of a query point in the vector space. In this work [13], we revisit this notion in the specific case of IR. More precisely, we show how to estimate indiscriminateness from IR similarities in order to use it in representation spaces used for documents and words. We show that indiscriminateness may be used to characterize difficult queries; moreover we show that this notion, applied to word embeddings, can help to choose terms to use for query expansion.

6.1.5. Heat Map Based Feature Ranker

Participants: Christian Raymond, Carlos Huertas [Autonomous University of Baja California, Mexico], Reyes Uarez-Ramirez [Autonomous University of Baja California, Mexico].

In [6], we present Heat Map Based Feature Ranker, an algorithm to estimate feature importance purely based on its interaction with other variables. A compression mechanism reduces evaluation space up to 66% without compromising efficacy. Our experiments show that our proposal is very competitive against popular algorithms, producing stable results across different types of data. We also show how noise reduction through feature selection aids data visualization using emergent self-organizing maps.

6.1.6. Time series retrieval and indexing using DTW-preserving shapelets

Participants: Laurent Amsaleg, Ricardo Carlini Sperandio, Simon Malinowski, Romain Tavenard [Univ. Rennes 2].

Dynamic Time Warping (DTW) is a very popular similarity measure used for time series classification, retrieval or clustering. DTW is, however, a costly measure, and its application on numerous and/or very long time series is difficult in practice. We have proposed a new approach for time series retrieval: time series are embedded into another space where the search procedure is less computationally demanding, while still accurate. This approach is based on transforming time series into high-dimensional vectors using DTW-preserving shapelets. That transform is such that the relative distance between the vectors in the Euclidean transformed space well reflects the corresponding DTW measurements in the original space. We have also proposed in [12] strategies for selecting a subset of shapelets in the transformed space, resulting in a trade-off between the complexity of the transformation and the accuracy of the retrieval. Experimental results using the well known time series datasets demonstrate the importance of this trade-off. This transformation can then be used to build efficient time series indexing schemes.

6.1.7. Fast Spectral Ranking for Similarity Search

Participants: Yannis Avrithis, Teddy Furon, Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Ondra Chum [Univ. Prague].

Despite the success of deep learning on representing images for particular object retrieval, recent studies show that the learned representations still lie on manifolds in a high dimensional space. This makes the Euclidean nearest neighbor search biased for this task. Exploring the manifolds online remains expensive even if a nearest neighbor graph has been computed offline. This work introduces an explicit embedding reducing manifold search to Euclidean search followed by dot product similarity search. This is equivalent to linear graph filtering of a sparse signal in the frequency domain. To speed up online search, we compute an approximate Fourier basis of the graph offline. We improve the state of art on particular object retrieval datasets including the challenging Instre dataset containing small objects. At a scale of 10^5 images, the offline cost is only a few hours, while query time is comparable to standard similarity search [15].

6.1.8. Mining on Manifolds: Metric Learning without Labels

Participants: Yannis Avrithis, Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Ondra Chum [Univ. Prague].

In this work we present a novel unsupervised framework for hard training example mining [17]. The only input to the method is a collection of images relevant to the target application and a meaningful initial representation, provided e.g. by pre-trained CNN. Positive examples are distant points on a single manifold, while negative examples are nearby points on different manifolds. Both types of examples are revealed by disagreements between Euclidean and manifold similarities. The discovered examples can be used in training with any discriminative loss. The method is applied to unsupervised fine-tuning of pre-trained networks for fine-grained classification and particular object retrieval. Our models are on par or are outperforming prior models that are fully or partially supervised.

6.1.9. Hybrid Diffusion: Spectral-Temporal Graph Filtering for Manifold Ranking

Participants: Yannis Avrithis, Teddy Furon, Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Ondra Chum [Univ. Prague].

State of the art image retrieval performance is achieved with CNN features and manifold ranking using a k-NN similarity graph that is pre-computed off-line. The two most successful existing approaches are temporal filtering, where manifold ranking amounts to solving a sparse linear system online, and spectral filtering, where eigen-decomposition of the adjacency matrix is performed off-line and then manifold ranking amounts to dot-product search online. The former suffers from expensive queries and the latter from significant space overhead. Here we introduce a novel, theoretically well-founded hybrid filtering approach allowing full control of the space-time trade-off between these two extremes. Experimentally, we verify that our hybrid method delivers results on par with the state of the art, with lower memory demands compared to spectral filtering approaches and faster compared to temporal filtering [16].

6.1.10. Transactional Support for Visual Instance Search

Participants: Laurent Amsaleg, Björn Þór Jónsson [Univ. Copenhagen], Herwig Lejsek [Videntifier Tech.].

This work addresses the issue of dynamicity and durability for scalable indexing of very large and rapidly growing collections of local features for visual instance retrieval. By extending the NV-tree, a scalable disk-based high-dimensional index, we show how to implement the ACID properties of transactions which ensure both dynamicity and durability. We present a detailed performance evaluation of the transactional NV-tree, showing that the insertion throughput is excellent despite the effort to enforce the ACID properties [20].

6.1.11. Time-series prediction for capacity planning

Participants: Simon Malinowski, Colin Leverger [Orange Labs], Thomas Guyet [AgroCampus Ouest], Vincent Lemaire [Orange Labs].

In a collaboration with Orange Labs, we have worked on KPI time series prediction in order to improve capacity planning. A software has been develloped. This software is detailed in [32]. It aims at visualizing and comparing different time series prediction techniques on user-defined input data. We have also developed a novel prediction algorithm that focuses on time series for with a seasonality [21]. It uses the combination of a clustering algorithm and Markov Models to produce day-ahead forecasts. Our experiments on real datasets show that in the case study, our method outperforms classical approaches (AR, Holt-Winters).

6.1.12. Scale-adaptive CNN for Crowd counting

Participants: Miaojing Shi, Lu Zhang [Fudan Univ.], Qiaobo Chen [Shanghai Jiaotong Univ.].

The task of crowd counting is to automatically estimate the pedestrian number in crowd images. To cope with the scale and perspective changes that commonly exist in crowd images, this work proposes a scale-adaptive CNN (SaCNN) architecture with a backbone of fixed small receptive fields. We extract feature maps from multiple layers and adapt them to have the same output size; we combine them to produce the final density map. The number of people is computed by integrating the density map. We also introduce a relative count loss along with the density map loss to improve the network generalization on crowd scenes with few pedestrians, where most representative approaches perform poorly on. We conduct extensive experiments and demonstrate significant improvements of SaCNN over the state-of-the-art [31].

6.1.13. Revisiting Perspective information for Efficient Crowd counting

Participants: Miaojing Shi, Zhaohui Yang [Peking Univ.], Chao Xu [Peking Univ.], Qijun Chen [Tongji Univ.].

A major challenge of crowd counting 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. 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 to demonstrate the effectiveness and efficiency of PACNN over the state-of-the-art [42].

6.1.14. Phone-Level Embeddings for Unit Selection Speech Synthesis

Participants: Laurent Amsaleg, Antoine Perquin [EXPRESSION team, IRISA], Gwénolé Lecorvé [EXPRESSION team, IRISA], Damien Lolive [EXPRESSION team, IRISA].

Deep neural networks have become the state of the art in speech synthesis. They have been used to directly predict signal parameters or provide unsupervised speech segment descriptions through embeddings. In [25] we present four models with two of them enabling us to extract phone-level embeddings for unit selection speech synthesis. Three of the models rely on a feed-forward DNN, the last one on an LSTM. The resulting embeddings enable replacing usual expert-based target costs by an euclidean distance in the embedding space. This work is conducted on a French corpus of an 11 hours audiobook. Perceptual tests show the produced speech is preferred over a unit selection method where the target cost is defined by an expert. They also show that the embeddings are general enough to be used for different speech styles without quality loss. Furthermore, objective measures and a perceptual test on statistical parametric speech synthesis show that our models perform comparably to state-of-the-art models for parametric signal generation, in spite of necessary simplifications, namely late time integration and information compression.

6.1.15. Disfluency Insertion for Spontaneous TTS: Formalization and Proof of Concept

Participants: Pascale Sébillot, Raheel Qader [EXPRESSION team, IRISA], Gwénolé Lecorvé [EXPRESSION team, IRISA], Damien Lolive [EXPRESSION team, IRISA].

This is an exploratory work to automatically insert disfluencies in text-to-speech (TTS) systems. The objective is to make TTS more spontaneous and expressive. To achieve this, we propose to focus on the linguistic level of speech through the insertion of pauses, repetitions and revisions. We formalize the problem as a theoretical process, where transformations are iteratively composed. This is a novel contribution since most of the previous work either focus on the detection or cleaning of linguistic disfluencies in speech transcripts, or solely concentrate on acoustic phenomena in TTS, especially pauses. We present a first implementation of the proposed process using conditional random fields and language models. The objective and perceptual evaluation conducted on an English corpus of spontaneous speech show that our proposition is effective to generate disfluencies, and highlights perspectives for future improvements [26]

6.1.16. Bi-directional Recurrent End-to-End Neural Network Classifier for Spoken Arab Digit Recognition

Participants: Christian Raymond, Naima Zerari [University of Batna 2, Algeria], Hassen Bouzgou [University of Batna 2, Algeria].

In [30], we propose a general end-to-end approach to sequence learning that uses Long Short-Term Memory (LSTM) to deal with the non-uniform sequence length of the speech utterances. The neural architecture can recognize the Arabic spoken digit spelling of an isolated Arabic word using a classification methodology, with the aim to enable natural human-machine interaction. The proposed system consists to, first, extract the relevant features from the input speech signal using Mel Frequency Cepstral Coefficients (MFCC) and then these features are processed by a deep neural network able to deal with the non uniformity of the sequences length. A recurrent LSTM or GRU architecture is used to encode sequences of MFCC features as a fixed size.

6.1.17. Are Deep Neural Networks good for blind image watermarking?

Participants: Teddy Furon, Vedran Vukotić [Lamark, France], Vivien Chappelier [Lamark, France].

Image watermarking is usually decomposed into three steps: i) some features are extracted from an image, ii) they are modified to embed the watermark, iii) and they are projected back into the image space while avoiding the creation of visual artefacts. The feature extraction is usually based on a classical image representation given by the Discrete Wavelet Transform or the Discrete Cosine Transform for instance. These transformations need a very accurate synchronisation and usually rely on various registration mechanisms for that purpose. This paper investigates a new family of transformation based on Deep Learning networks. Motivations come from the Computer Vision literature which has demonstrated the robustness of these features against light geometric distortions. Also, adversarial sample literature provides means to implement the inverse transform needed in the third step. This work [29] shows that this approach is feasible as it yields a good quality of the watermarked images and an intrinsic robustness.

6.2. Description and structuring

6.2.1. Automatic classification of radiological reports for clinical care

Participants: Anne-Lyse Minard, Alfonso Gerevini [Università degli Studi di Brescia, Italy], Alberto Lavelli [Fondazione Bruno Kessler, Italy], Alessandro Maffi [Università degli Studi di Brescia, Italy], Roberto Maroldi [Università degli Studi di Brescia, Italy, Azienda Socio Sanitaria Territoriale Spedali Civili di Brescia, Italy], Ivan Serina [Università degli Studi di Brescia, Italy], Guido Squassina [Azienda Socio Sanitaria Territoriale Spedali Civili di Brescia, Italy].

Radiological reporting generates a large amount of free-text clinical narratives, a potentially valuable source of information for improving clinical care and supporting research. The use of automatic techniques to analyze such reports is necessary to make their content effectively available to radiologists in an aggregated form. In this paper we focus on the classification of chest computed tomography reports according to a classification schema proposed for this task by radiologists of the Italian hospital ASST Spedali Civili di Brescia. The proposed system is built exploiting a training data set containing reports annotated by radiologists. Each report is classified according to the schema developed by radiologists and textual evidences are marked in the report. The annotations are then used to train different machine learning based classifiers. We present in this paper a method based on a cascade of classifiers which make use of a set of syntactic and semantic features. The resulting system is a novel hierarchical classification system for the given task, that we have experimentally evaluated [5].

6.2.2. Revisiting the medial axis for planar shape decomposition

Participants: Yannis Avrithis, N. Papanelopoulos [NTU Athens], S. Kollias [Univ. Lincoln].

We introduce 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 [7]. 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 psychophysical experiments. With our single computational model, we outperform even an ensemble method on several other competing models.

6.2.3. Is ATIS too shallow to go deeper for benchmarking Spoken Language Understanding models?

Participants: Christian Raymond, Frédéric Béchet [Aix Marseille University].

We started a collaboration about benchmarking scientific benchmarks. We started in [11] by the ATIS (Air Travel Information Service) corpus, that will be soon celebrating its 30th birthday. Designed originally to benchmark spoken language systems, it still represents the most well-known corpus for benchmarking Spoken Language Understanding (SLU) systems. In 2010, in a paper titled "What is left to be understood in ATIS?", Tur et al. discussed the relevance of this corpus after more than 10 years of research on statistical models for performing SLU tasks. Nowadays, in the Deep Neural Network (DNN) era, ATIS is still used as the main benchmark corpus for evaluating all kinds of DNN models, leading to further improvements, although rather limited, in SLU accuracy compared to previous state-of-the-art models. We propose in this paper to investigate these results obtained on ATIS from a qualitative point of view rather than just a quantitative point of view and answer the two following questions: what kind of qualitative improvement brought DNN models to SLU on the ATIS corpus? Is there anything left, from a qualitative point of view, in the remaining 5% of errors made by current state-of-the-art models?

6.2.4. KRAUTS: A German Temporally Annotated News Corpus

Participants: Anne-Lyse Minard, Strötgen Jannik [Max Planck Institute for Informatics, Germany], Lukas Lange [Max Planck Institute for Informatics, Germany], Manuela Speranza [Fondazione Bruno Kessler, Italy], Bernardo Magnini [Fondazione Bruno Kessler, Italy].

In recent years, temporal tagging, i.e., the extraction and normalization of temporal expressions, has become a vibrant research area. Several tools have been made available, and new strategies have been developed. Due to domain-specific challenges, evaluations of new methods should be performed on diverse text types. Despite significant efforts towards multilinguality in the context of temporal tagging, for all languages except English, annotated corpora exist only for a single domain. In the case of German, for example, only a narrative style corpus has been manually annotated so far, thus no evaluations of German temporal tagging performance on news articles can be made. In this paper, we present KRAUTS, a new German temporally annotated corpus containing two subsets of news documents: articles from the daily newspaper DOLOMITEN and from the weekly newspaper DIE ZEIT. Overall, the corpus contains 192 documents with 1,140 annotated temporal expressions, and has been made publicly available to further boost research in temporal tagging citejannik:hal-01844834.

6.2.5. Active learning to assist annotation of aerial Images in environmental surveys

Participants: Ewa Kijak, Mathieu Laroze [OBELIX team, IRISA], Romain Dambreville [OBELIX team, IRISA], Chloe Friguet [OBELIX team, IRISA], Sébastien Lefèvre [OBELIX team, IRISA].

Remote sensing technologies greatly ease environmental assessment over large study areas using aerial images, e.g. for monitoring and counting animals or ships. Such data are most often analyzed by a manual operator, leading to costly and non scalable solutions. If object detection algorithms are used to fasten and automate the counting processes, these algorithms need to have prior ground truth available, which is a time-consuming and tedious process for field experts or engineers. We introduced a method to assist the annotation process in aerial images by introducing an active learning algorithm, allowing interaction with the expert such as class confirmation or correction at the labeling stage, and querying the expert with groups of samples taken from the same image to ease user annotation. Usual active learning algorithms perform instance selection from the whole set of input data. In this work, the selection of the queried instances is constrained by requiring that they belong to a group, (a part of) an image in our case, to ease the annotator task as the queried instances are proposed in their comprehensive context. We defined a score to rank the images and identify the one

that should be annotated at each iteration, based on both uncertainty and true positives. The main objective is to reduce the number of human interactions on the overall process, starting from a first annotated image, rather than reaching the maximum final accuracy. Therefore, the annotation cost is measured through the gain in interactions (corrections of the classifier decisions by the annotator) with respect to a labeling task from scratch. At each iteration, the classifier is retrained according to a specific subset of data. Several strategies have been compared and their performances regarding the interaction gain have been discussed [19], [36].

6.3. Search, linking and navigation

6.3.1. Detecting fake news and tampered images in social networks

Participants: Cédric Maigrot, Ewa Kijak, Vincent Claveau.

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. This is the issue addressed by the participants to the Verification Multimedia Use task of Mediaeval 2016. They used several approaches and clues from different modalities (text, image, social information).

One promising approach is to examine if the image (if any) has been doctored. In recent work [23], we study context-aware methods to localize tamperings in images from social media. The problem is defined as a comparison between image pairs: an near-duplicate image retrieved from the network and a tampered version. We propose a method based on local features matching, followed by a kernel density estimation, that we compare to recent similar approaches. The proposed approaches are evaluated on two dedicated datasets containing a variety of representative tamperings in images from social media, with difficult examples. Context-aware methods are proven to be better than blind image forensics approach. However, the evaluation allows to analyze the strengths and weaknesses of the contextual-based methods on realistic datasets.

In further work [9], [22], we explore the interest of combining and merging these approaches in order to evaluate the predictive power of each modality and to make the most of their potential complementarity.

6.3.2. A Crossmodal Approach to Multimodal Fusion in Video Hyperlinking

Participants: Christian Raymond, Guillaume Gravier, Vedran Vukotić.

With the recent resurgence of neural networks and the proliferation of massive amounts of unlabeled data, unsupervised learning algorithms became very popular for organizing and retrieving large video collections in a task defined as video hyperlinking. Information stored as videos typically contain two modalities, namely an audio and a visual one, that are used conjointly in multimodal systems by undergoing fusion. Multimodal autoencoders have been long used for performing multimodal fusion. In this work, we start by evaluating different initial, single-modal representations for automatic speech transcripts and for video keyframes. We progress to evaluating different autoencoding methods of performing multimodal fusion in an offline setup. The best performing setup is then evaluated in a live setup at TRECVID's 2016 video hyperlinking task. As in offline evaluations, we show that focusing on crossmodal translations as a way of performing multimodal fusion yields improved multimodal representations and that our simple system, trained in an unsupervised manner, with no external information information, defines the new state of the art in a live video hyperlinking setup. We conclude by performing an analysis on data gathered after the live evaluations at TRECVID 2016 and express our thoughts on the overall performance of our proposed system [8].

6.3.3. A study on multimodal video hyperlinking with visual aggregation

Participants: Mateusz Budnik, Mikail Demirdelen, Guillaume Gravier.

Video hyperlinking offers a way to explore a video collection, making use of links that connect segments having related content. Hyperlinking systems thus seek to automatically create links by connecting given anchor segments to relevant targets within the collection. In 2018, we pursued our long-term research effort towards multimodal representations of video segments in a hyperlinking system based on bidirectional deep neural networks, which achieved state-of-the-art results in the TRECVid 2016 evaluation. A systematic study of different input representations was done with a focus on the aggregation of the representation of multiple keyframes. This includes, in particular, the use of memory vectors as a novel aggregation technique, which provides a significant improvement over other aggregation methods on the final hyperlinking task. Additionally, the use of metadata was investigated leading to increased performance and lower computational requirements for the system [35].

6.3.4. Opinion mining in social networks

Participants: Anne-Lyse Minard, Christian Raymond, Vincent Claveau.

As part of the DeFT text-mining challenge, we participated in the elaboration of a task on fine-grained opinion mining in tweets [34] and to the analysis of the participants' results. We have also proposed systems [33] for each sub-task: (i) tweet classification according to the topic of the tweet, (ii) tweet classification according to their polarity, (iii) detection of the polarity markers and target of opinion in tweets. For the two first tasks, the approaches we proposed rely on a combination of boosting, decision trees and Recurrent Neural Networks. For the last task, we experimented with RNN coupled with a CRF layer. All of these systems performed very well and ranked in the best performing systems for each of the task.

6.3.5. Biomedical Information Extraction in social networks

Participants: Anne-Lyse Minard, Christian Raymond, Vincent Claveau.

This year, we participated in SMM4H challenge about extracting medical information from social networks. Faour tasks were proposed: (i) detection of posts mentioning a drug name, (ii) classification of posts describing medication intake, (iii) classification of adverse drug reaction mentioning posts, (iv) Automatic detection of posts mentioning vaccination behavior. In [24], we presented the systems developed by IRISA to participate to these four tasks. For these tweet classification tasks, we adopt a common approach based on recurrent neural networks (BiLSTM). Our main contributions are the use of certain features, the use of Bagging in order to deal with unbalanced datasets, and on the automatic selection of difficult examples. These techniques allow us to reach 91.4, 46.5, 47.8, 85.0 as F1-scores for Tasks 1 to 4, ranking us among the 3 first participants for each task.

6.3.6. Information Extraction in the biomedical domain

Participants: Clément Dalloux, Vincent Claveau, N. Grabar [STL-CNRS].

Automatic detection of negated content is often a pre-requisite in information extraction systems, especially in the biomedical domain. Following last year work, we propose two main contributions in this field [43]. We first introduced a new corpora built with excerpts from clinical trial protocols in French and Brazilian Portuguese, describing the inclusion criteria for patient recruitment. The corpora are manually annotated for marking up the negation cues and their scope. Secondly, two supervised learning approaches are been proposed for the automatic detection of negation. Besides, one of the approaches is validated on English data from the state of the art: the approach shows very good results and outperforms existing approaches, and it also yields comparable results on the French data.

We also have developed other data-sets (annotated corpora). Indeed, textual corpora are extremely important for various NLP applications as they provide information necessary for creating, setting and testing these applications and the corresponding tools. They are also crucial for designing reliable methods and reproducible results. Yet, in some areas, such as the medical area, due to confidentiality or to ethical reasons, it is complicated and even impossible to access textual data representative of those produced in these areas. We propose the CAS corpus [14] built with clinical cases, such as they are reported in the published scientific literature in French. We describe this corpus, containing over 397,000 word occurrences, and its current annotations (PoS, lemmas, negation, uncertainty).

As part of this work, we also developed software available as web-services on http://allgo.inria.fr (see the Software section).

6.3.7. Revisiting Oxford and Paris: Large-Scale Image Retrieval Benchmarking

Participants: Yannis Avrithis, F. Radenovic [Univ. Prague], Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Ondra Chum [Univ. Prague].

In this work [27] we address issues with image retrieval benchmarking on standard and popular Oxford 5k and Paris 6k datasets. In particular, annotation errors, the size of the dataset, and the level of challenge are addressed: new annotation for both datasets is created with an extra attention to the reliability of the ground truth. Three new protocols of varying difficulty are introduced. The protocols allow fair comparison between different methods, including those using a dataset pre-processing stage. For each dataset, 15 new challenging queries are introduced. Finally, a new set of 1M hard, semi-automatically cleaned distractors is selected. An extensive comparison of the state-of-the-art methods is performed on the new benchmark. Different types of methods are evaluated, ranging from local-feature-based to modern CNN based methods. The best results are achieved by taking the best of the two worlds. Most importantly, image retrieval appears far from being solved.

6.3.8. Unsupervised object discovery for instance recognition

Participants: Oriane Siméoni, Yannis Avrithis, Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], 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 descriptors derived from the salient regions improve particular object retrieval, most noticeably in a large collections containing small objects [28].

MAGRIT Project-Team

7. New Results

7.1. Matching and localization

Participants: Marie-Odile Berger, Vincent Gaudilliere, Antoine Fond, Gilles Simon.

Vanishing point detection

Accurate detection of vanishing points (VPs) is a prerequisite for many computer vision problems such as camera self-calibration, single view structure recovery, video compass, robot navigation and augmented reality, among many others. More specifically, knowing three orthogonal VPs aligned with the buildings of a scene (the Manhattan directions) allows computing the intrinsic parameters of the camera as well as warped images where the buildings' facades are orthorectified, facilitating their detection and registration. VPs are also used in our work on epipolar geometry estimation to help matching line segments, a particularly difficult task in low-textured environments.

We introduced an *a-contrario* method to solve this problem. Our key contribution was to show that, as soon as the horizon line (HL) is inside the image boundaries, this line can usually be detected as an alignment of oriented line segments. This comes from a simple geometric property, that any horizontal line segment at the height of the camera's optical center projects to the HL regardless of its 3-D direction. This property generally yields statistically meaningful events, detectable from *a-contrario* analysis. Additional candidate HLs are sampled around these events using a Gaussian Mixture Model (GMM), and scored according to the strongest of the VPs hypothesized along them. VP hypotheses are also obtained from an *a-contrario* method, using integral geometry to accurately model the background noise. Experiments made on three urban datasets showed that our method, not only achieves state-of-the-art performance w.r.t. computation times and accuracy of the HL, but also yields much less spurious VPs than the previous top-ranked methods. This work was published at ECCV'2018 [23] and an article is in preparation for submission in a peer-reviewed journal. In this article, we show that our method also outperforms state-of-the-art methods on a new industrial dataset that we built and will make publicly available. We also establish a relation between the Number of False Alarms (NFA) obtained for the meaningful events and the spreads of the GMM. In addition, the Matlab code implementing our method has been made publicly available.

Urban AR

Urban localization plays a major role in many applications including navigation aid, labeling of local touristic landmarks, and robot localization. The outdoor accuracy of mobile phone GPS is only 12.5 meters and can be worse in urban areas where the street is flanked by buildings on both sides. By contrast, buildings' facades are meaningful landmarks to rely on for large-scale localization. Last year, we proposed a method to automatically detect facades in an image, based on image cues that measure facade characteristics such as shape, color, contours, semantic structure and symmetry. Matching the detected facade with a facade database using a metric learned through a siamese neural network allowed us to estimate a first initialization of the registration parameters by solving the least-square problem that maps the four transformed corners of the reference to the four corners of the detection.

This year, we attempted to rely on semantic segmentation to improve the accuracy of that initial registration [11]. Simultaneously, we aimed to iteratively improve the quality of the semantic segmentation through registration. Registration and semantic segmentation were jointly solved in a Expectation-Maximization framework. We especially introduced a Bayesian model that uses prior semantic segmentation as well as geometric structure of the facade reference modeled by Generalized Gaussian Mixtures. We showed the advantages of our method in terms of robustness to clutter and change of illumination on urban images from various databases. We currently are assessing the relevance of the method using the large scale dataset SFM Aachen, in order to compare it with state-of-the-art SFM-based localization.

AR in industrial environments
Industrial environments are normally inundated with textureless objects, specular surfaces, repetitive objects and artificial lights, etc. which may fail traditional 2D/3D matching-based approaches. Line segments are numerous in industrial environments, but contrary to what happens in urban scenes, matching is a tough issue since most segments are silhouette contours whose appearance is viewpoint dependant. The combinatory of segment matches is thus very high, making impossible in practice the use of RANSAC algorithms for pose computation.

Within V. Gaudilliere's PhD thesis [21], [25], we took advantage of global properties of the environment, both geometric - such as the presence of numerous vertical planes - and contextual to guide matching. First, sub-image correspondences based on high level ConvNet features are used as prior for vertical planes detection and matching. Then, local homographies are detected between matched regions. To ensure efficient estimations, we have developed a dedicated RANSAC framework in which model hypotheses are first generated based on vanishing point and visual keypoint correspondences, and then validated on key points and line segments. This potential set of matched features are finally filtered with a robust fundamental matrix estimation. That scheme enables us to circumvent problems encountered in poorly-textured images (sparsity of visual keypoints and difficulties to match segments) while taking advantage of the abundance of segments and vanishing points characteristic of industrial environments

7.2. Handling non-rigid deformations

Participants: Marie-Odile Berger, Jaime Garcia Guevara, Daryna Panicheva, Pierre-Frédéric Villard.

Elastic multi-modal registration
Our previous works about 3D tracking for deformable objects [1] are template-based methods and thus need to carefully register the model onto the image in the first image. In practice this task is performed manually and is especially difficult for deformable organs. Within J. Guevara's PhD thesis, we have proposed an automatic method for registering pre and per-operative imagery which exploits the matching of the vascular trees, visible in most pre and intra-operative images. Although methods dedicated to non-rigid graph registration exist, they are not efficient when large intra-operative deformations of tissues occur. Our contribution is an extension of the graph-matching algorithm based on Gaussian process regression (GPR) proposed in [28]. Our idea is to combine GPR with a biomechanical model of the organ. Indeed, GPR allows for rigorous and fast error propagation but is extremely versatile and may thus produce non physically coherent registration, while biomechanical transformations are slower to compute but are capable of handling non linear deformation while preserving their physical nature. They thus allow earlier incoherent matching hypotheses to be removed. Integrating the two approaches allows us to significantly improve the quality of the registration while reducing computation times. These contributions have been published in the IPCAI conference [20] and in the International Journal of Computer Assisted Radiology and Surgery [13].

Individual-specific heart valve modeling
We first finished up a feasibility study aiming at providing fast image-based mitral valve mechanical simulation from individualized geometry [19]. The method was demonstrated on one dataset which was interactively segmented. In order to extend the pipeline to any data, robust methods to segment the valve components are required. Within the context of D. Panicheva's PhD thesis, we are currently working on means allowing to automatically segment the chordae. 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. In this contribution, we exploit the fact that there are no other generalized cylinders than the chordae in the CT scan and we propose 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.

INVIVE: The Individual Virtual Ventilator: 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 propose to use a meshfree RBF method. During this year, we have worked on building an implicit representation of the surface of the diaphragm based on the topology coming from last year researches. It has been associated with a linear elasticity model and the boundary conditions have been measured on landmark points extracted by medical experts.

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, which may have an impact on the clinical outcome of the procedure. This question is considered as a single view 3D curve reconstruction problem. We follow a constrained non-rigid shape from motion approach, using a physics-based model as a prior for the object shape. The navigation model is implemented through interactive simulation that provides a prediction of the device, taking into account non-linear effects such as friction during contacts.

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 of the model with image data (opaque markers placed along the device). Progress has been made this year towards an effective formulation of the filter. In particular, a good estimate is recovered for the device shape in the case of ambiguous views (overlapping anatomy, bifurcations).

The method has been implemented in Sofa simulation software platform. Validation has been performed on porcine in-vivo data, acquired in accordance with UE norms, in collaboration with Pr. Mario Gimenez and Dr. Alain Garcia from IHU-Strasbourg.

In-vivo procedures will be performed under ethical approval of MSER (reference to ethic protocol *APAFIS #15433-2018060815283960*).

Markerless similarity metrics were investigated during Juan Rocha's Master's thesis. The update equations of the filter were generalized to tackle curve to image similarity metrics, traditionally used in multi-view reconstruction methods.

7.3. Image processing

Participants: Gilles Simon, Fabien Pierre, Frédéric Sur.

Variational methods for image processing
In the previous decade, variational methods in image processing have been widely used with a huge number of applications. The convex hypothesis generally makes the proof of convergence easier, whereas it is not fulfilled by the most interesting problems in imaging. The non-convexity may appear in some applications such as image colorization with multiple candidates selection [27] or in the case of M-estimators computation, in particular with an assumption of Cauchy noise [16]. These two points of view of the non-convex variational methods bring two different mathematical challenges to ensure the convergence of the numerical scheme. The choice of one candidate among a collection of possible ones implies bi-convex functions (functions with multiple variables, convex with respect to each ones). The computation of M-estimators with Cauchy noise hypothesis implies smooth but non-convex functions.

Our contributions concern both types of non-convexity. For bi-convex functions, we have demonstrated in [27] the convergence of alternate gradient descent numerical scheme with inertial relaxation of the iterates. Moreover, an application to image colorization has been proposed. In [16], a fixed-point algorithm has been studied to solve the problem of the Myriad filters. The particularity of this work is the convergence of the numerical scheme to a local minimum with probability 1, which is, up to our best knowledge, a novelty in the optimization community.

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. This year's outcomes of our collaboration with Institut Pascal (Clermont-Ferrand) focus on three areas.

We have proposed an algorithm to render synthetic speckle images deformed under a predetermined deformation fixed by the user [17]. The goal is to generate ground truth datasets in order to assess the performance of the numerous variants of DIC and also the influence of extrinsic factors such as the noise or the marking pattern. It is required to carefully design the rendering algorithm in order to ensure that any measurement bias is caused by DIC estimation and not by the rendering algorithm itself. We have proposed to render speckle images based on a Boolean model, a standard model of stochastic geometry, a Monte Carlo estimation giving the gray level at any pixel. A software library and datasets are publicly available.

We have also investigated the optimization of the pattern marking the specimen [15], which is the topic of various 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 lighting conditions, the noise level in displacement and strain maps obtained with checkerboards is lower than that obtained with classic 2D grids.

Another scientific contribution concerns the restoration of displacement and strain maps. DIC and LSA both provide displacement fields equal to the actual one convolved 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 [14]. 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 the spatial resolution.

Cartoon-texture image 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 [18] is shown to achieve a good compromise between computation time and accuracy. Matlab code is publicly available.

MORPHEO Project-Team

7. New Results

7.1. Surface Motion Capture Animation Synthesis

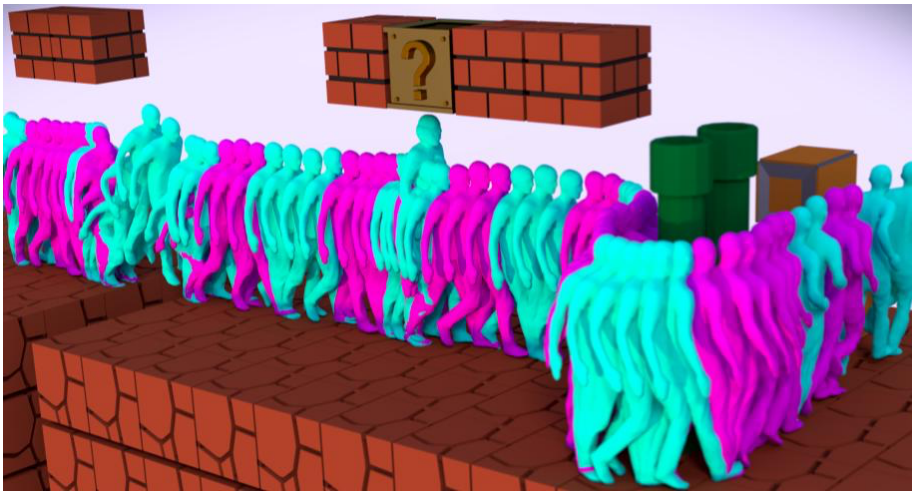


Figure 3.

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 3).

This result was published in a prominent computer graphics journal, IEEE Transactions on Visualization and Computer Graphics [2].

7.2. A Multilinear Tongue Model Derived from Speech Related MRI Data of the Human Vocal Tract

We present a multilinear statistical model of the human tongue that captures anatomical and tongue pose related shape variations separately. The model is derived from 3D magnetic resonance imaging data of 11 speakers sustaining speech related vocal tract configurations. To extract model parameters, we use a minimally supervised method based on an image segmentation approach and a template fitting technique. Furthermore,

we use image denoising to deal with possibly corrupt data, palate surface information reconstruction to handle palatal tongue contacts, and a bootstrap strategy to refine the obtained shapes. Our evaluation shows that, by limiting the degrees of freedom for the anatomical and speech related variations, to 5 and 4, respectively, we obtain a model that can reliably register unknown data while avoiding overfitting effects. Furthermore, we show that it can be used to generate plausible tongue animation by tracking sparse motion capture data.

This result was published in Computer Speech and Language 51 [3].

7.3. CBCT of a Moving Sample from X-rays and Multiple Videos

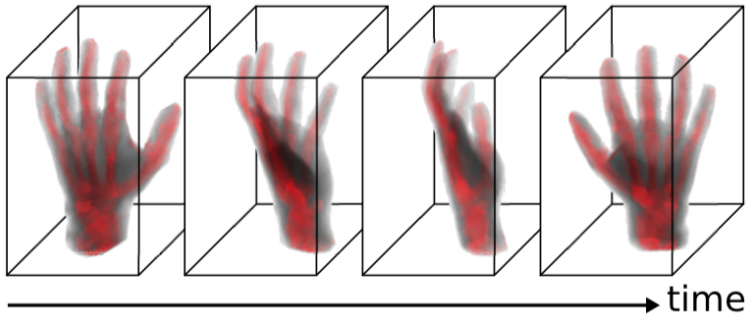


Figure 4. Dense volumetric attenuation reconstruction from a rigidly moving sample captured by a single planar X-ray imaging device and a surface motion capture system. Higher attenuation (here bone structure) is highlighted in red.

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 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 4 .

This result was published in a prominent medical journal, IEEE Transactions on Medical Imaging [4].

7.4. Automatic camera calibration using multiple sets of pairwise correspondences

We propose a new method to add an uncalibrated node into a network of calibrated cameras using only pairwise point correspondences (see figure 5). While previous methods perform this task using triple correspondences, these are often difficult to establish when there is limited overlap between different views. In such challenging cases we must rely on pairwise correspondences and our solution becomes more advantageous. Our method includes an 11-point minimal solution for the intrinsic and extrinsic calibration of a camera from pairwise correspondences with other two calibrated cameras, and a new inlier selection framework that extends the

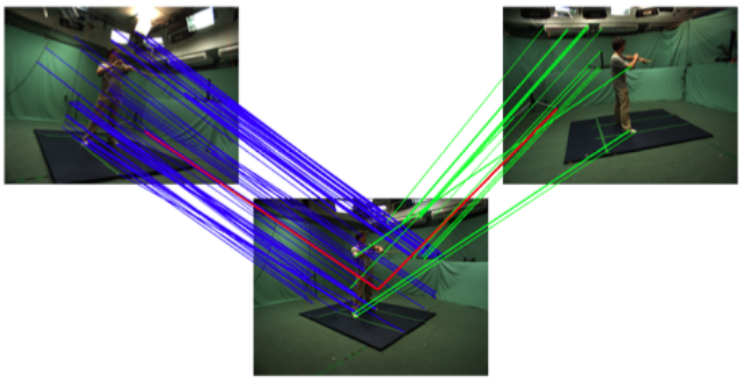


Figure 5. Correspondences extracted from SIFT features. Given the wide baseline between the views there is a single reliable triple correspondence (red) while there are many reliable pairwise correspondences (blue and green).

traditional RANSAC family of algorithms to sampling across multiple datasets. Our method is validated on different application scenarios where a lack of triple correspondences might occur: addition of a new node to a camera network; calibration and motion estimation of a moving camera inside a camera network; and addition of views with limited overlap to a Structure-from-Motion model.

This result was published in a prominent medical journal, IEEE Transactions on Pattern Analysis and Machine Intelligence [5].

7.5. Multilinear Autoencoder for 3D Face Model Learning

Figure 6. Shape variations caused by different expressions of the same subject.

Generative models have proved to be useful tools to represent 3D human faces and their statistical variations (see figure 6). With the increase of 3D scan databases available for training, a growing challenge lies in the ability to learn generative face models that effectively encode shape variations with respect to desired attributes, such as identity and expression, given datasets that can be diverse. This paper addresses this challenge by proposing a framework that learns a generative 3D face model using an autoencoder architecture, allowing hence for weakly supervised training. The main contribution is to combine a convolutional neural network-based en-coder with a multilinear model-based decoder, taking therefore advantage of both the convolutional network robustness to corrupted and incomplete data, and of the multilinear model capacity to effectively model and decouple shape variations. Given a set of 3D face scans with annotation labels for

the desired attributes, e.g. identities and expressions, our method learns an expressive multilinear model that decouples shape changes due to the different factors. Experimental results demonstrate that the proposed method outperforms recent approaches when learning multilinear face models from incomplete training data, particularly in terms of space decoupling, and that it is capable of learning from an order of magnitude more data than previous methods.

This result was published in IEEE Winter Conference on Applications of Computer Vision [6].

7.6. Spatiotemporal Modeling for Efficient Registration of Dynamic 3D Faces

We consider the registration of temporal sequences of 3D face scans. Face registration plays a central role in face analysis applications, for instance recognition or transfer tasks, among others. We propose an automatic approach that can register large sets of dynamic face scans without the need for landmarks or highly specialized acquisition setups. This allows for extended versatility among registered face shapes and deformations by enabling to leverage multiple datasets, a fundamental property when e.g. building statistical face models. Our approach is built upon a regression-based static registration method, which is improved by spatiotemporal modeling to exploit redundancies over both space and time. We experimentally demonstrate that accurate registrations can be obtained for varying data robustly and efficiently by applying our method to three standard dynamic face datasets.

This work has been published in 3D Vision 2018 [7].

7.7. Shape Reconstruction Using Volume Sweeping and Learned Photoconsistency

Figure 7. Challenging scene captured with Kinovis. (left) one input image, (center) reconstructions obtained with our previous work based on classical 2D features, (right) proposed solution. Our results validate the key improvement of a CNN-learned disparity to MVS for performance capture scenarios. Results particularly improve in noisy, very low contrast and low textured regions such as the arm, the leg or even the black skirt folds.

The rise of virtual and augmented reality fuels an increased need for contents suitable to these new technologies including 3D contents obtained from real scenes (see figure 7). We consider in this paper the problem of 3D shape reconstruction from multi-view RGB images. We investigate the ability of learning-based strategies to effectively benefit the reconstruction of arbitrary shapes with improved precision and robustness. We especially target real life performance capture, containing complex surface details that are difficult to recover with existing approaches. A key step in the multi-view reconstruction pipeline lies in the search for matching features between viewpoints in order to infer depth information. We propose to cast the matching on a 3D receptive field along viewing lines and to learn a multi-view photoconsistency measure for that purpose. The intuition is that deep networks have the ability to learn local photometric configurations in a broad way, even with respect to different orientations along various viewing lines of the same surface point. Our results demonstrate this ability, showing that a CNN, trained on a standard static dataset, can help recover surface details on dynamic scenes that are not perceived by traditional 2D feature based methods. Our evaluation also shows that our solution compares on par to state of the art reconstruction pipelines on standard evaluation datasets, while yielding significantly better results and generalization with realistic performance capture data.

This work has been published in the European Conference on Computer Vision 2018 [9] and Reconnaissance des Formes, Image, Apprentissage et Perception 2018 [8].

7.8. FeaStNet: Feature-Steered Graph Convolutions for 3D Shape Analysis

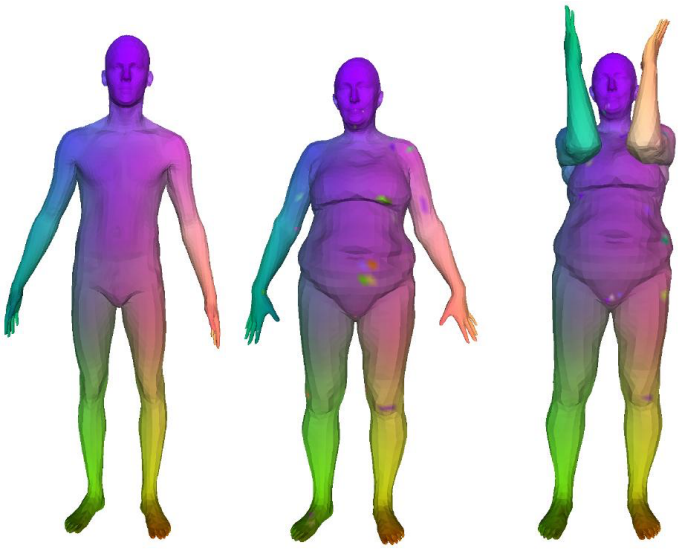


Figure 8. Two examples of texture transfer from a reference shape in neutral pose (left) using shape correspondences predicted by FeaStNet (multi-scale architecture, without refinement).

Convolutional neural networks (CNNs) have massively impacted visual recognition in 2D images, and are now ubiquitous in state-of-the-art approaches. CNNs do not easily extend, however, to data that are not represented by regular grids, such as 3D shape meshes or other graph-structured data, to which traditional local convolution operators do not directly apply. To address this problem, we propose a novel graph-convolution operator to establish correspondences between filter weights and graph neighborhoods with arbitrary connectivity. The key novelty of our approach is that these correspondences are dynamically computed from features learned by the network, rather than relying on predefined static coordinates over the graph as in previous work. We obtain excellent experimental results that significantly improve over previous state-of-the-art shape correspondence

results (see figure 8). This shows that our approach can learn effective shape representations from raw input coordinates, without relying on shape descriptors.

This work has been published in the IEEE Conference on Computer Vision and Pattern Recognition 2018 [11].

7.9. Analyzing Clothing Layer Deformation Statistics of 3D Human Motions

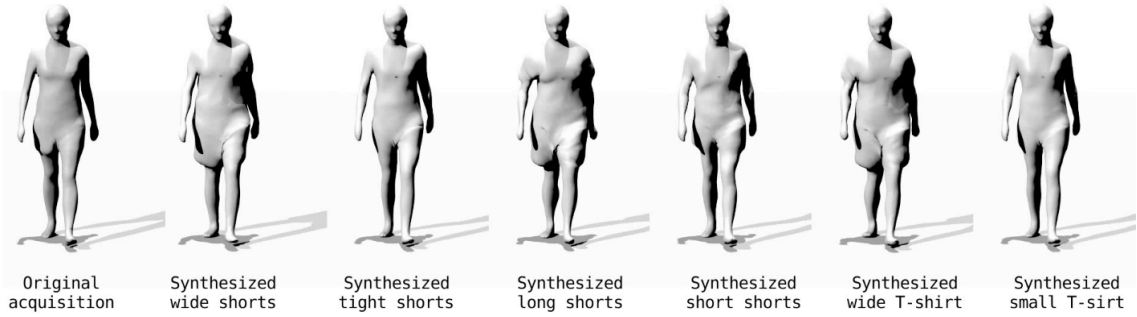


Figure 9. Examples of clothing re-synthesis based on our clothing layer regression model.

Recent capture technologies and methods allow not only to retrieve 3D model sequence of moving people in clothing, but also to separate and extract the underlying body geometry and motion component and separate the clothing as a geometric layer. So far this clothing layer has only been used as raw offsets for individual applications such as retargeting a different body capture sequence with the clothing layer of another sequence, with limited scope, e.g. using identical or similar motions. The structured, semantics and motion-correlated nature of the information contained in this layer has yet to be fully understood and exploited. To this purpose we propose a comprehensive analysis of the statistics of this layer with a simple two-component model, based on PCA subspace reduction of the layer information on one hand, and a generic parameter regression model using neural networks on the other hand, designed to regress from any semantic parameter whose variation is observed in a training set, to the layer parameterization space. We show that this model not only allows to reproduce previous motion retargeting works, but generalizes the data generation capabilities of the method to other semantic parameters such as clothing variation and size (see figure 9), or physical material parameters with synthetically generated training sequence, paving the way for many kinds of capture data-driven creation and augmentation applications.

This work has been published in the European Conference on Computer Vision 2018 [12].

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 [39]. 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 Dereverberation and Noise Reduction Using the Convolutive Transfer Function

We address the problems of blind multichannel identification and equalization for *joint speech dereverberation and noise reduction*. The standard time-domain cross-relation methods are hardly applicable for blind room impulse response identification due to the near-common zeros of the long impulse responses. We extend the cross-relation formulation to the short-time Fourier transform (STFT) domain, in which the time-domain impulse response is approximately represented by the convolutive transfer function (CTF) with much less coefficients. For the oversampled STFT, CTFs suffer from the common zeros caused by the non-flat-top STFT window. To overcome this, we propose to identify CTFs using the STFT framework with oversampled signals and critically sampled CTFs, which is a good trade-off between the frequency aliasing of the signals and the common zeros problem of CTFs. The phases of the identified CTFs are inaccurate due to the frequency aliasing of the CTFs, and thus only their magnitudes are used. This leads to a non-negative multichannel equalization method based on a non-negative convolution model between the STFT magnitude of the source signal and the CTF magnitude. To recover the STFT magnitude of the source signal and to reduce the additive noise, the ℓ_2-norm fitting error between the STFT magnitude of the microphone signals and the non-negative convolution is constrained to be less than a noise power related tolerance. Meanwhile, the ℓ_1-norm of the STFT magnitude of the source signal is minimized to impose the sparsity [38].
Website: https://team.inria.fr/perception/research/ctf-dereverberation/.

6.3. Speech Enhancement with a Variational Auto-Encoder

We addressed the problem of enhancing speech signals in noisy mixtures using a source separation approach. We explored the use of neural networks as an alternative to a popular speech variance model based on supervised non-negative matrix factorization (NMF). More precisely, we use a variational auto-encoder as a speaker-independent supervised generative speech model, highlighting the conceptual similarities that this approach shares with its NMF-based counterpart. In order to be free of generalization issues regarding the noisy recording environments, we follow the approach of having a supervised model only for the target

speech signal, the noise model being based on unsupervised NMF. We developed a Monte Carlo expectation-maximization algorithm for inferring the latent variables in the variational auto-encoder and estimating the unsupervised model parameters. Experiments show that the proposed method outperforms a semi-supervised NMF baseline and a state-of-the-art fully supervised deep learning approach.
Website: https://team.inria.fr/perception/research/ieee-mlsp-2018/.

6.4. Audio-Visual Speaker Tracking and Diarization

We are particularly interested in modeling the interaction between an intelligent device and a group of people. For that purpose we develop audio-visual person tracking methods [36]. As the observed persons are supposed to carry out a conversation, we also include speaker diarization into our tracking methodology. We cast the diarization problem into a tracking formulation whereby the active speaker is detected and tracked over time. A probabilistic tracker exploits the spatial coincidence of visual and auditory observations and infers a single latent variable which represents the identity of the active speaker. Visual and auditory observations are fused using our recently developed weighted-data mixture model [12], while several options for the speaking turns dynamics are fulfilled by a multi-case transition model. The modules that translate raw audio and visual data into image observations are also described in detail. The performance of the proposed method are tested on challenging datasets that are available from recent contributions which are used as baselines for comparison [36].
Websites:
https://team.inria.fr/perception/research/wdgmm/,
https://team.inria.fr/perception/research/speakerloc/,
https://team.inria.fr/perception/research/speechturndet/, and
https://team.inria.fr/perception/research/avdiarization/.

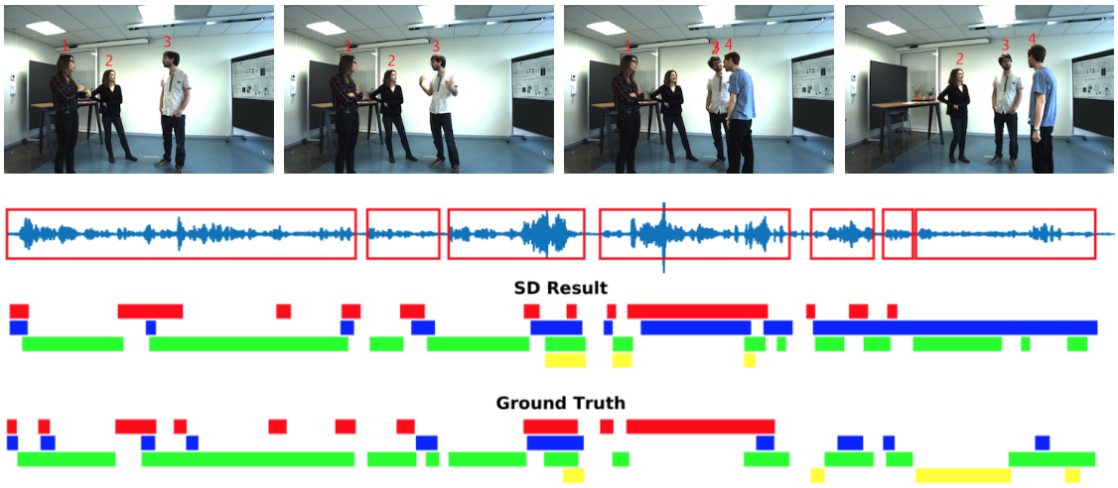


Figure 3. This figure illustrates the audiovisual tracking and diarization method that we have recently developed. First row: A number is associated with each tracked person. Second row: diarization result. Third row: the ground truth diarization. Fourth row: acoustic signal recorded by one of the two microphones.

6.5. Tracking Eye Gaze and of Visual Focus of Attention

The visual focus of attention (VFOA) has been recognized as a prominent conversational cue. We are interested in estimating and tracking the VFOAs associated with multi-party social interactions. We note that in this type of situations the participants either look at each other or at an object of interest; therefore their eyes are not always visible. Consequently both gaze and VFOA estimation cannot be based on eye detection and tracking. We propose a method that exploits the correlation between eye gaze and head movements. Both VFOA and gaze are modeled as latent variables in a Bayesian switching state-space model (also named switching Kalman filter). The proposed formulation leads to a tractable learning method and to an efficient online inference procedure that simultaneously tracks gaze and visual focus. The method is tested and benchmarked using two publicly available datasets, Vernissage and LAEO, that contain typical multi-party human-robot and human-human interactions [42].
Website: https://team.inria.fr/perception/research/eye-gaze/.

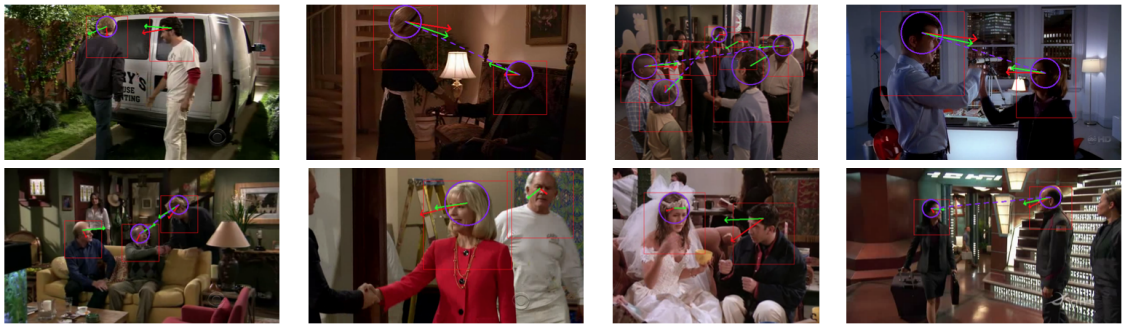


Figure 4. This figure shows some results obtained with the LAEO dataset. The top row shows results obtained with coarse head orientation and the bottom row shows results obtained with fine head orientation. Head orientations are shown with red arrows. The algorithm infers gaze directions (green arrows) and VFOAs (blue circles). People looking at each others are shown with a dashed blue line.

6.6. Variational Bayesian Inference of Multiple-Person Tracking

We addressed the problem of tracking multiple speakers using audio information or via the fusion of visual and auditory information. 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 propose a variational inference model which amounts to approximate the joint distribution with a factorized distribution. The solutions take the form of closed-form expectation maximization procedures using Gaussian distributions [44], [58], [56] or the von Mises distribution for circular variables [55]. We described in detail the inference algorithms, we evaluate their performance and we compared them with several baseline methods. These experiments show that the proposed audio and audio-visual trackers perform well in informal meetings involving a time-varying number of people.
Websites:
https://team.inria.fr/perception/research/var-av-track/,

https://team.inria.fr/perception/research/audiotrack-vonm/.

6.7. High-Dimensional and Deep Regression

One of the most important achievements for the last years has been the development of high-dimensional to low-dimensional regression methods. The motivation for investigating this problem raised from several problems that appeared both in audio signal processing and in computer vision. Indeed, often the task in data-driven methods is to recover low-dimensional properties and associated parameterizations from high-dimensional observations. Traditionally, this can be formulated as either an unsupervised method (dimensionality reduction of manifold learning) or a supervised method (regression). We developed a learning methodology at the crossroads of these two alternatives: the output variable can be either fully observed or partially observed. This was cast into the framework of linear-Gaussian mixture models in conjunction with the concept of inverse regression. It gave rise to several closed-form and approximate inference algorithms [8]. The method is referred to as *Gaussian locally linear mapping*, or GLLiM. As already mentioned, high-dimensional regression is useful in a number of data processing tasks because the sensory data often lies in high-dimensional spaces. Each one of these tasks required a special-purpose version of our general framework. Sound-source localization was the first to benefit from our formulation. Nevertheless, the sparse nature of speech spectrograms required the development of a GLLiM version that is able to with full-spectrum sounds and to test with sparse-spectrum ones [9]. This could be immediately applied to audio-visual alignment and to sound-source separation and localization [7].

In conjunction with our computer vision work, high-dimensional regression is a very useful methodology since visual features, obtained either by hand-crafted feature extraction methods or using convolutional neural networks, lie in high-dimensional spaces. Such properties as object pose lie in low-dimensional spaces and must be extracted from features. We took such an approach and proposed a head pose estimator [10]. Visual tracking can also benefit from GLLiM. Indeed, it is not practical to track objects based on high-dimensional features. We therefore combined GLLiM with switching linear dynamic systems. In 2018 we proposed a robust deep regression method [46]. In parallel we thoroughly benchmarked and analyzed deep regression tasks using several CNN architectures [57].

6.8. Human-Robot Interaction

Audio-visual fusion raises interesting problems whenever it is implemented onto a robot. Robotic platforms have their own hardware and software constraints. In addition, commercialized robots have economical constraints which leads to the use of cheap components. A robot must be reactive to changes in its environment and hence it must take fast decisions. This often implies that most of the computing resources must be onboard of the robot.

Over the last decade we have tried to do our best to take these constraints into account. Starting from our scientific developments, we put a lot of efforts into robotics implementations. For example, the audio-visual fusion method described in [2] used a specific robotic middleware that allowed fast communication between the robot and an external computing unit. Subsequently we developed a powerful software package that enables distributed computing. We also put a lot of emphasis on the implementation of low-level audio and visual processing algorithms. In particular, our single- and multiple audio source methods were implemented in real time onto the humanoid robot NAO [25], [50]. The multiple person tracker [4] was also implemented onto our robotic platforms [5], e.g. Figure 5 .

More recently, we investigated the use of reinforcement learning (RL) as an alternative to sensor-based robot control [45], [37]. 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. 6 . 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.

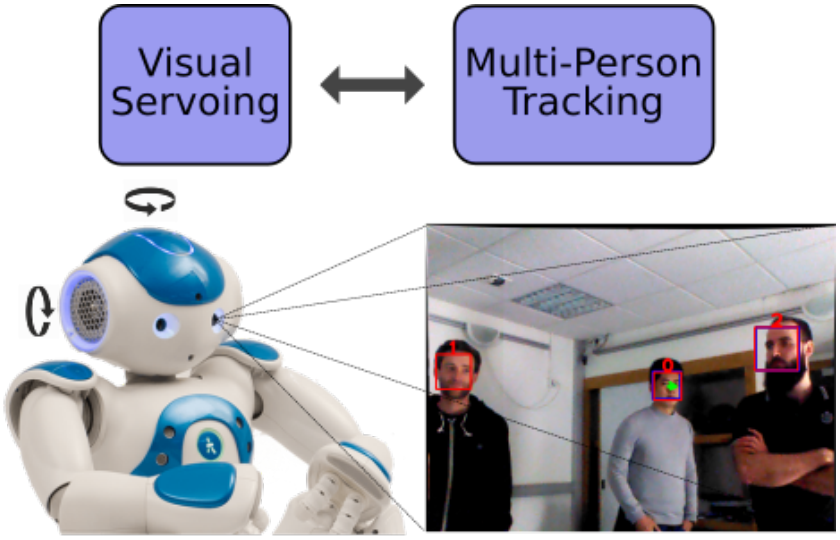


Figure 5. The multi-person tracking method is combined with a visual servoing module. The latter estimates the optimal robot commands and the expected impact of the tracked person locations. The multi-person tracking module refines the locations of the persons with the new observations and the information provided by the visual servoing.

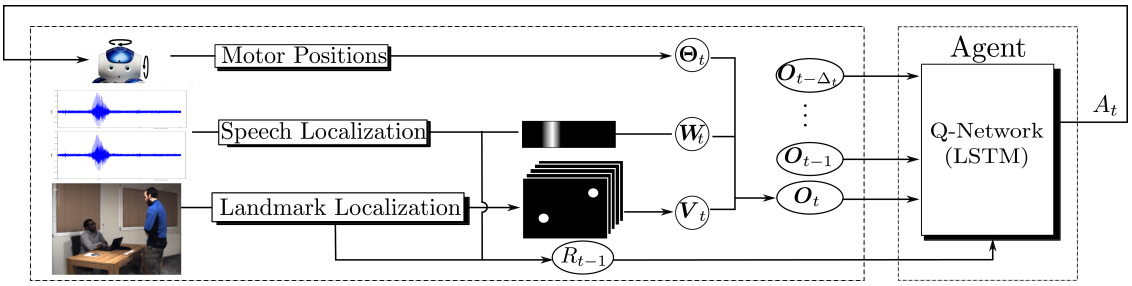


Figure 6. Overview of the proposed deep RL method for controlling the gaze of a robot. At each time index t, audio and visual data are represented as binary maps which, together with motor positions, form the set of observations $\mathbf{O}_t$. A motor action A_t (rotate the head left, right, up, down, 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 [45], [37] for further details.

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 [45], [37] we proposed the use of a simulated environment for pre-training the RL parameters, thus avoiding spending hours of tedious interaction.
Websites:
https://team.inria.fr/perception/research/deep-rl-for-gaze-control/,
https://team.inria.fr/perception/research/mot-servoing/.

6.9. Generation of Diverse Behavioral Data

We target the automatic generation of visual data depicting human behavior, and in particular how to design a method able to learn the generation of *data diversity*. In particular, we focus on smiles, because each smile is unique: one person surely smiles in different ways (e.g. closing/opening the eyes or mouth). We wonder if given one input image of a neutral face, we can generate multiple smile videos with distinctive characteristics. To tackle this one-to-many video generation problem, we propose a novel deep learning architecture named Conditional MultiMode Network (CMM-Net). To better encode the dynamics of facial expressions, CMM-Net explicitly exploits facial landmarks for generating smile sequences. Specifically, a variational auto-encoder is used to learn a facial landmark embedding. This single embedding is then exploited by a conditional recurrent network which generates a landmark embedding sequence conditioned on a specific expression (e.g. spontaneous smile), implemented as a Conditional LSTM. Next, the generated landmark embeddings are fed into a multi-mode recurrent landmark generator, producing a set of landmark sequences still associated to the given smile class but clearly distinct from each other, we call that a Multi-Mode LSTM. Finally, these landmark sequences are translated into face videos. Our experimental results, see Figure 7 , demonstrate the effectiveness of our CMM-Net in generating realistic videos of multiple smile expressions [52].

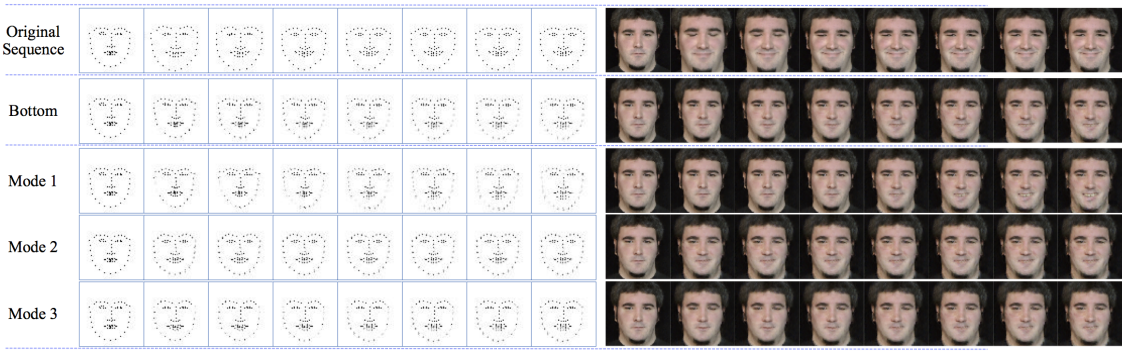


Figure 7. Multi-mode generation example with a sequence: landmarks (left) and associated face images (right) after the landmark-to-image decoding step based on Variational Auto-Encoders. The rows correspond to the original sequence (first), output of the Conditional LSTM (second), and output of the Multi-Mode LSTM (last three rows).

6.10. Registration of Multiple Point Sets

We have also addressed the rigid registration problem of multiple 3D point sets. While the vast majority of state-of-the-art techniques build on pairwise registration, we proposed a generative model that explains jointly registered multiple sets: back-transformed points are considered realizations of a single Gaussian mixture model (GMM) whose means play the role of the (unknown) scene points. Under this assumption, the joint registration problem is cast into a probabilistic clustering framework. We formally derive an expectation-maximization procedure that robustly estimates both the GMM parameters and the rigid transformations that map each individual cloud onto an under-construction reference set, that is, the GMM means. GMM variances carry rich information as well, thus leading to a noise- and outlier-free scene model as a by-product. A second version of the algorithm is also proposed whereby newly captured sets can be registered online. A thorough discussion and validation on challenging data-sets against several state-of-the-art methods confirm the potential of the proposed model for jointly registering real depth data [35].
Website: https://team.inria.fr/perception/research/jrmpc/

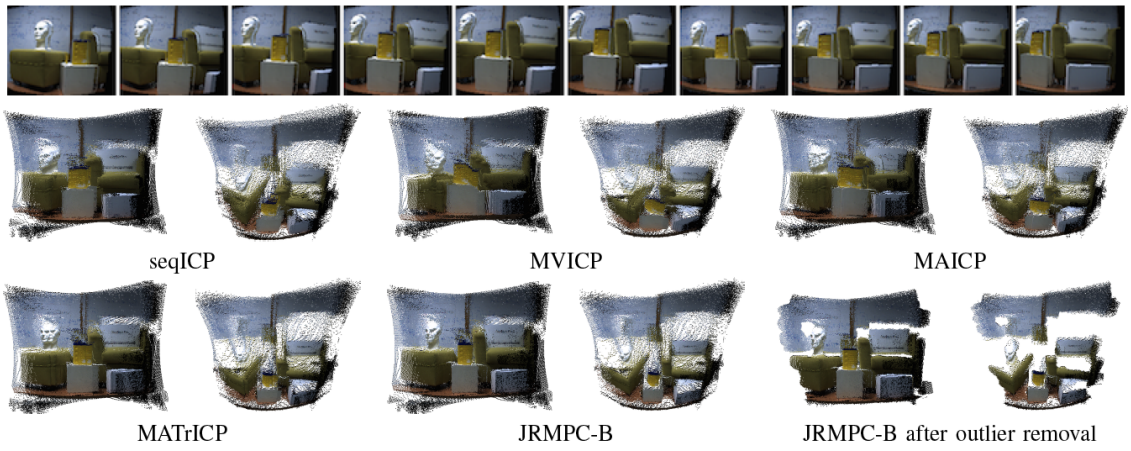


Figure 8. Integrated point clouds from the joint registration of 10 TOF images that record a static scene (EXBI data-set). Top: color images that roughly show the scene content of each range image (occlusions due to cameras baseline may cause texture artefacts). Bottom: front-view and top-view of integrated sets after joint registration. The results obtained with the proposed method (JRMPC-B) are compared with several other methods.

SIROCCO Project-Team

7. New Results

7.1. Visual Data Analysis

Scene depth, Scene flows, 3D modelling, Light-fields, 3D point clouds

7.1.1. Super-rays for efficient light fields processing

Participants: Matthieu Hog, Christine Guillemot.

Light field acquisition devices allow capturing scenes with unmatched post-processing possibilities. However, the huge amount of high dimensional data poses challenging problems to light field processing in interactive time. In order to enable light field processing with a tractable complexity, we have addressed, in collaboration with Neus Sabater (Technicolor) the problem of light field over-segmentation. We have introduced the concept of super-ray, which is a grouping of rays within and across views, as a key component of a light field processing pipeline. The proposed approach is simple, fast, accurate, easily parallelisable, and does not need a dense depth estimation. We have demonstrated experimentally the efficiency of the proposed approach on real and synthetic datasets, for sparsely and densely sampled light fields. As super-rays capture a coarse scene geometry information, we have also shown how they can be used for real-time light field segmentation and for correcting refocusing angular aliasing. The concept of super-rays has been extended to video light fields addressing problems of temporal tracking of super-rays using sparse scene flows[15].

7.1.2. Scene depth estimation from light fields

Participants: Christian Galea, Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

While there exist scene depth and scene flow 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 [17]. 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. Experiments show that, both for synthetic and real light fields, our algorithm achieves competitive performance compared to state-of-the-art algorithms which exploit the entire light field and usually generate the depth map for the center viewpoint only. Figure 2 shows the estimated depth map for a synthetic light field in comparison with the ground truth. The estimated depth maps allow us to construct accurate 3D point clouds of the captured scene [16]. This work is now pursued considering deep learning solutions.

7.1.3. Scene flow estimation from light fields

Participants: Pierre David, Christine Guillemot.

Temporal processing of dynamic 3D scenes requires estimating the displacement of the objects in the 3D space, i.e., so-called scene flows. 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. Estimating dense scene flows in light fields pose 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. We have addressed the complexity problem by designing a sparse estimation method followed by a densification step that avoids the difficulty of computing matches in occluded areas. 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.

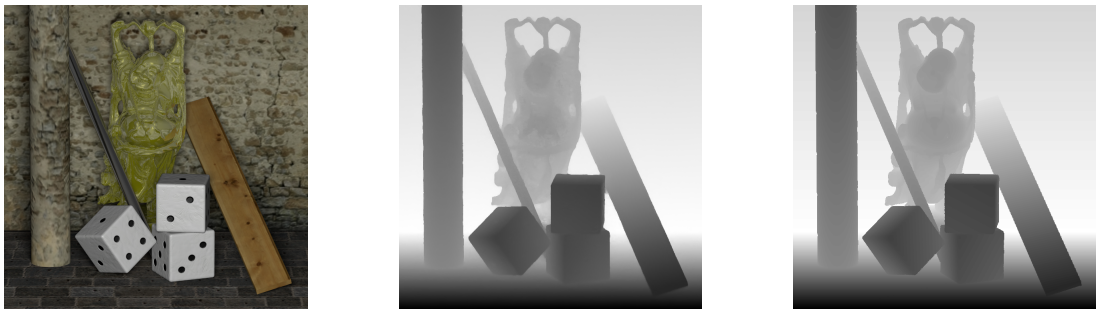


Figure 2. Estimated depth map (middle) for the light field 'Buddha' in comparison with the ground truth (right).

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. Multi-shot single sensor light field camera using a color coded mask

Participant: Christine Guillemot.

In collaboration with the University of Linkoping (Prof. J. unger, Dr. E. Miandji), we have proposed a compressive sensing framework for reconstructing a light field from a single-sensor consumer camera capture with color coded masks [19]. The proposed *camera architecture* captures incoherent measurements of the light field via a controllable color mask placed in front of the sensor. To enhance the incoherence, hence the reconstruction quality, we propose to utilize multiple shots where, for each shot, the mask configuration is changed to create a new random pattern. To reduce computations and increase the incoherence, we also perform a random sampling of the spatial domain. The compressive sensing framework relies on a dictionary trained over a light field data set. Numerical simulations show significant improvements compared with a similar coded aperture system for light field capture.

7.2.2. Compressive 4D light field reconstruction

Participants: Christine Guillemot, Fatma Hawary.

Exploiting the assumption that light field data is sparse in the Fourier domain, we have also developed a new method for reconstructing a 4D light field from a random set of measurements [14]. The reconstruction algorithm searches for these bases (i.e., their frequencies) which best represent the 4D Fourier spectrum of the sampled light field. The method has been further improved by introducing an orthogonality constraint on the residue, in the same vein as orthogonal matching pursuit but in the Fourier transform domain, as well as a refinement for non integer frequencies. The method achieves a very high reconstruction quality, in terms of PSNR (more than 1dB gain compared to state-of-the-art algorithms).

7.2.3. Light fields dimensionality reduction with low-rank models

Participants: Elian Dib, Christine Guillemot, Xiaoran Jiang.

We have further investigated low-rank approximation methods exploiting data geometry for dimensionality reduction of light fields. While our first solution was considering global low-rank models based on homographies, we have recently developed local low-rank models exploiting disparity. The local support of the approximation is given by super-rays (see section 7.1.1). The super-rays group super-pixels which are consistent across the views while being constrained to be of same shape and size. The corresponding super-pixels in all views are found thanks to disparity compensation. In order to do so, a novel method has been proposed

to estimate the disparity for each super-ray using a low rank prior, so that the super-rays are constructed to yield the lowest approximation error for a given rank. More precisely, the disparity for each super-ray is found in order to align linearly correlated sub-aperture images in such a way that they can be approximated by the considered low rank model. The rank constraint is expressed as a product of two matrices, where one matrix contains basis vectors (or eigen images) and where the other one contains weighting coefficients. The eigen images are actually splitted into two sets, one corresponding to light rays visible in all views and a second one, very sparse, corresponding to occluded rays (see Fig. 3). A light field compression algorithm has been designed encoding the different components of the resulting low rank approximation.

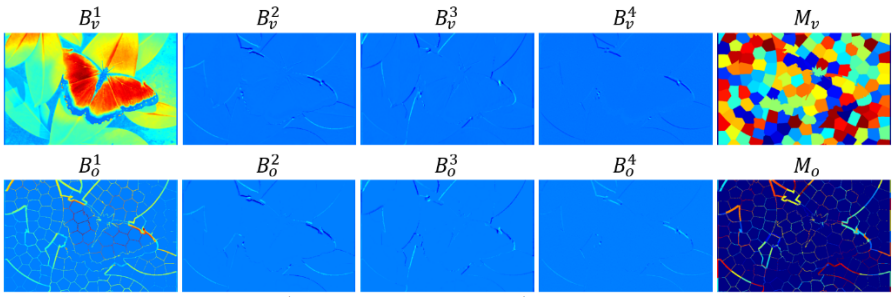


Figure 3. Eigen-images and segmentation maps for visible and occluded sets of pixels.

7.2.4. Graph-based transforms for light fields and omni-directional image compression

Participants: Christine Guillemot, Thomas Maugey, Mira Rizkallah, Xin Su.

Graph-based transforms are interesting tools for low-dimensional embedding of light field data. This embedding can be learned with a few eigenvectors 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, in [21], we have first developed *local super-rays based separable (spatial followed by angular)* weighted and unweighted transforms to jointly capture light fields correlation spatially and across views. While separable transforms on super-rays allow us to significantly decrease the eigenvector computation complexity, the basis functions of the spatial graph transforms to be applied on the super-ray pixels of each view are often not compatible, resulting in decreased correlation of the coefficients across views, 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 having the best alignment with those computed on a reference frame while still approximately diagonalizing their respective Laplacians. Fig.4 shows the second eigenvector of different super-pixels belonging to the same super-ray before and after optimization. A rate-distortion optimized graph partitioning algorithm has also been developed [20] for coding 360° videos signals, to achieve a good trade-off between distortion, smoothness of the signal on each subgraph, and the coding cost of the graph partition.

7.2.5. Neural networks for learning image transforms and predictors

Participants: Thierry Dumas, Christine Guillemot, Aline Roumy.

We have explored the problem of learning transforms for image compression via autoencoders. Learning a transform is equivalent to learning an autoencoder, which is of its essence unsupervised and therefore more difficult than classical supervised learning. In compression, the learning has in addition to be performed under a rate-distortion criterion, and not only a distortion criterion. Usually, the rate-distortion performances of image compression are tuned by varying the quantization step size. In the case of autoencoders, this in principle

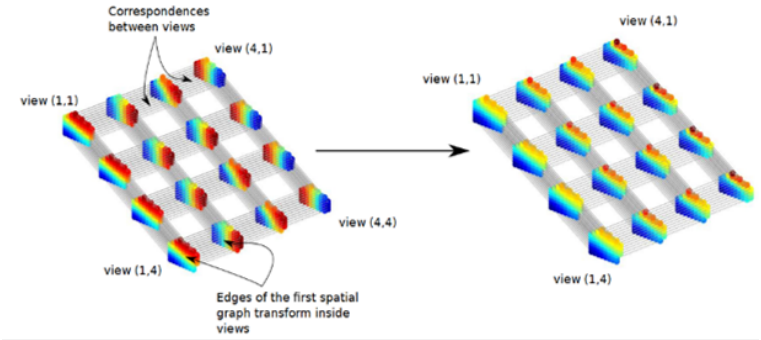


Figure 4. Second eigenvector of super-pixels forming a super-ray before and after optimization.

would require learning one transform per rate-distortion point at a given quantization step size. We have shown in [12] that comparable performances can be obtained with a unique learned transform. The different rate-distortion points are then reached by varying the quantization step size at test time. This approach saves a lot of training time.

Another important operator in compression algorithm is the predictor that aims at capturing spatial correlation. We have developed a set of neural network architectures, called Prediction Neural Networks Set (PNNS), based on both fully-connected and convolutional neural networks, for intra image prediction. It is shown that, while fully-connected neural networks give good performances for small block sizes, convolutional neural networks provide better predictions in large blocks with complex textures. Thanks to the use of masks of random sizes during training, the neural networks of PNNS well adapt to the available context that may vary, depending on the position of the image block to be predicted. Unlike the H.265 intra prediction modes, which are each specialized in predicting a specific texture, the proposed PNNS can model a large set of complex textures.

7.2.6. Cloud-based predictors and neural network temporal predictors video compression

Participants: Jean Begaint, Christine Guillemot.

Video codecs are primarily designed assuming that rigid, block-based, two-dimensional displacements are suitable models to describe the motion taking place in a scene. However, translational models are not sufficient to handle real world motion such as camera zoom, shake, pan, shearing or changes in aspect ratio. Building upon the region-based geometric and photometric model proposed in [5] to exploit correlation between images in the cloud, we have developed a region-based inter-prediction scheme for video compression. The proposed predictor is able to estimate multiple homography models in order to predict complex scene motion. We also introduce an affine photometric correction to each geometric model. Experiments on targeted sequences with complex motion demonstrate the efficiency of the proposed approach compared to the state-of- the-art HEVC video codec [11]. To further improve the accuracy of the temporal predictor, we have explored the use of deep neural networks for frame prediction and interpolation, and preliminary results have shown gains going up to 5% compared with the latest HEVC video codec.

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, Matthieu Hog, Xiaoran Jiang.

We have developed a lightweight convolutional neural network architecture able to perform view synthesis with occlusion handling in a stereo context, from one single, unlabelled and unannotated image, beyond state-of-the-art performance and with only a small amount of data required for training. In particular, it is able, at training and at test time, to estimate the disparity map corresponding to the problem at hand, and to evaluate a confidence in its prediction when using said disparity map for the synthesis. Knowing this confidence measure, it is then able to refine the value of the pixels wrongly estimated, with a refinement network component. The end result is a prediction built from a geometrical analysis of the scene, and completed in wrongly predicted areas by occlusion handling. Since 3D scene information is extracted in the course of the analysis, multiple new views can then be generated by interpolation.

Finally, in collaboration with Technicolor (N. Sabater and M. Hog), we have explored a novel way using recurrent neural networks to solve the problem of view synthesis in light fields. In particular, we proposed a novel solution using Long Short Term Memory Networks on a plane sweep volume. The approach has the advantage of having very few parameters and can be run on arbitrary sequence length. We have shown that the approach yields results that are competitive with the state of the art for dense light fields. Experimental results also show promising results when run on wider baselines.

7.3.2. Light field inpainting and restoration

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

With the increasing popularity of computational photography brought by light field, simple and intuitive editing of light field images is becoming a feature of high interest for users. Light field editing can be combined with the traditional refocusing feature, allowing a user to include or remove objects from the scene, change its color, its contrast or other features. A simple approach for editing a light field image can be obtained with an edit propagation, where first a particular subaperture view is edited (most likely the center one) and then a coherent propagation of this edit is performed through the other views. This problem is particularly challenging for the task of inpainting, as the disparity field is unknown under the occluding mask. We have developed a method that is computationally fast while giving coherent disparity in the masked region, allowing us to inpaint a light field of 81 views in a few seconds [10].

We have also developed a novel light field denoising algorithm using a vector-valued regularization operating in the 4D ray space. More precisely, the method performs a PDE-based anisotropic diffusion along directions defined by local structures in the 4D ray space. It does not require prior estimation of disparity maps. The local structures in the 4D light field are extracted using a 4D tensor structure. We use a diffusivity coefficient derived from the amount of local variations in the 4D space to control the smoothing along directions, surfaces, or volumes in the 4D ray space. The diffusivity coefficient is computed as a function of the 4 eigenvalues of the 4D structure tensor. Experimental results show that the proposed denoising algorithm performs well compared to state of the art methods, while keeping tractable complexity, even with high noise levels (see Fig.5).

7.3.3. High dynamic range light fields capture

Participant: Christine Guillemot.

In collaboration with Trinity College Dublin (Prof. A. Smolic, Dr. M. Le Pendu), we have proposed a method for capturing *High Dynamic Range (HDR) light fields* with dense viewpoint sampling. Analogously to the traditional HDR acquisition process, several light fields are captured at varying exposures with a plenoptic camera. The raw data are de-multiplexed to retrieve all light field viewpoints for each exposure. We then perform a soft detection of saturated pixels. Considering a matrix which concatenates all the vectorized views, we formulate the problem of recovering saturated areas as a Weighted Low Rank Approximation (WLRA) where the weights are defined from the soft saturation detection. The proposed WLRA method [18], extending the matrix completion algorithm of [7] to nonbinary weights, is shown to better handle the transition between the saturated and non-saturated areas. While the Truncated Nuclear Norm (TNN) minimization, traditionally used for single view HDR imaging, does not generalize to light fields, the proposed WLRA method successfully recovers the parallax in the over-exposed areas.

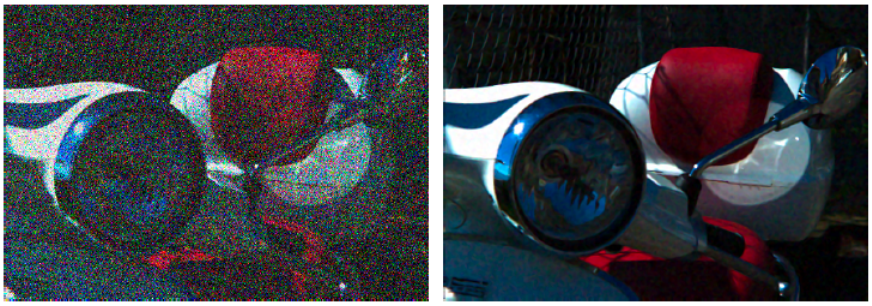


Figure 5. Illustration of denoising results, with additive white Gaussian noise of standard deviation $\sigma = 100$.

7.4. Distributed processing and robust communication

Information theory, stochastic modelling, 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. Information theoretic bounds for sequential massive random access to large database of correlated data

Participants: Thomas Maugey, Mai Quyen Pham, Aline Roumy.

Massive random access is a new source coding paradigm that we proposed. It allows us to extract arbitrary sources from an appropriately compressed database purely by bit extraction. We studied the sequential aspect of this problem where the clients successively access to one source after the other. Theoretical bounds have been derived, and it was shown that the extraction can be done at the same rate as if the database was decoded and the requested sources were re-encoded. As for the storage, a reasonable overhead is required. In [26], we derived the optimal storage and transmission rate regions to the case of more general sources, which occur in practical scenarios. For the lossless source coding problem, we considered non i.i.d. sources (i.e., with memory, but also non necessary ergodic). We also showed that, in the case source statistics are unknown, the rate is increased by a factor that vanishes as the length of the data goes to infinity. Lossy compression is another context of interest, in particular for the application to video. Therefore, we derived achievable storage and transmission rate regions under a distortion constraint for i.i.d. [26] and correlated [13] Gaussian sources. Similarly, the transmission rate-distortion region is the same as if re-encoding of the requested sources was allowed. We are currently extending this work, by studying the constraints of the successive user requests and their influence on the transmission-storage rates performance.

7.4.2. Correlation model selection for interactive video communication

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

One application of the sequential massive random access problem is interactive video communication for multi-view videos. In this scheme, the server has to store the views as compactly as possible while allowing interactive navigation. Interactive navigation refers to the possibility for the user to select one view or a subset of views. To achieve this goal, the compression must be done using a model-based coding in which the correlation between the predicted view generated on the user side and the original view has to be modeled by a statistical distribution. A question of interest is therefore how to select a model among a candidate set of models that incurs the lowest extra rate cost to the system. To answer this question, one should evaluate the effect on the transmission rate of using at the decoder a wrong model distribution. This question is related to an open problem in information theory called the mismatch capacity. So, we did not tackle the question for

any type of code as in the case of the mismatch capacity. In contrast, we focused on a type of code of practical interest: the linear codes. More precisely, we proposed a criterion to select the model when a linear block code is used for compression. We showed that, experimentally, the proposed bound is an accurate estimate of the effect of using a wrong model.

7.4.3. Compression of spatio-temporally correlated and massive georeferenced data

Participants: Thomas Maugey, Aline Roumy.

Another application of the sequential massive random access problem is interactive compression of spatio-temporally correlated sources. For example, highly instrumented smart cities are facing problems of management and storage of a large volume of data coming from an increasing number of sources. In [23] different compression schemes have been proposed that are able to exploit not only the temporal but also the spatial correlation between data sources. A special focus was made on a scheme where some sensors are used as references to predict the remaining sources. Finally, an adaptation of the scheme was proposed to offer interactivity and free selection of some sources by a client. This work was been done in collaboration with the Inria I4S project-team (A. Criniere), IFFSTAR (J. Dumoulin) and the L2S (M. Kieffer).

7.4.4. ICON 3D - Interactive COding for Navigation in 3D scenes

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey.

In the context of the ICON3D project, in collaboration with I3S-Nice (F. Payan), we have proposed a novel prediction tool for improving the compression performance of texture atlases of 3D meshes. This algorithm, called Geometry-Aware (GA) intra coding, takes advantage of the topology of the associated 3D meshes, in order to reduce the redundancies in the texture map. For texture processing, the general concept of the conventional intra prediction, used in video compression, has been adapted to utilize neighboring information on the 3D surface. We have also studied how this prediction tool can be integrated into a complete coding solution. In particular, a new block scanning strategy, as well as a graph-based transform for residual coding have been proposed. Experimental results show that the knowledge of the mesh topology can significantly improve the compression efficiency of texture atlases.

STARS Project-Team

7. New Results

7.1. Introduction

This year Stars has proposed new results related to its three main research axes : perception for activity recognition, semantic activity recognition and software engineering for activity recognition.

7.1.1. Perception for Activity Recognition

Participants: François Brémond, Juan Diego Gonzales Zuniga, Abhijit Das, Antitza Dancheva, Furqan Muhammad Khan, Michal Koperski, Thi Lan Anh Nguyen, Remi Trichet, Ujjwal Ujjval, Srijan Das, Vikas Thamizharasan, Monique Thonnat.

The new results for perception for activity recognition are:

- Late Fusion of multiple convolutional layers for pedestrian detection (see 7.2)
- Deep Learning applied on Embedded Systems for People Tracking (see 7.3)
- Cross Domain Residual Transfer Learning for Person Re-identification (see 7.4)
- Face-based Attribute Classification (see 7.5)
- Face Attribute manipulation
- From attribute-labels to faces: face generation using a conditional generative adversarial network (see 7.6)
- Face analysis in structured light images (see 7.7)

7.1.2. Semantic Activity Recognition

Participants: François Brémond, Antitza Dantcheva, Farhood Negin, Thanh Hung Nguyen, Michal Koperski, Srijan Das, Kaustubh Sakhalkar, Arpit Chaudhary, Abhishek Goel, Abdelrahman Abubakr, Abhijit Das, Yaohui Wang, S L Happy, Alexandra König, Guillaume Sacco, Philippe Robert, Soumik Mallick, Julien Badie, Monique Thonnat.

For this research axis, the contributions are :

- Deep-Temporal LSTM for Daily Living Action Recognition (see 7.8)
- A New Hybrid Architecture for Human Activity Recognition from RGB-D videos (see 7.10)
- Where to focus on for Human Action Recognition? (see 7.11)
- Online temporal detection of daily-living human activities in long untrimmed video streams (see 7.12)
- Activity Detection in Long-term Untrimmed Videos (see 7.13)
- Video based face analysis for health monitoring (see 7.14)
- Mobile biometrics (see 7.15)
- Comparing methods for assessment of facial dynamics in patients with major neurocognitive disorders (see 7.16)
- Combating the issue of low sample size in facial expression recognition (see 7.17)
- Serious exergames for Cognitive Stimulation (see 7.18)
- Fully Automatic Speech-Based Analysis of the Semantic Verbal Fluency Task (see 7.19)
- Language Modelling in the Clinical Semantic Verbal Fluency Task (see 7.19.2)
- Telephone-based Dementia Screening I: Automated Semantic Verbal Fluency Assessment (see 7.19.3)
- Automatic Detection of Apathy using Acoustic Markers extracted from Free Emotional Speech and using Automatic Speech Analysis (see 7.19.4)
- Monitoring the Behaviors of Retail Customers (see 7.20)

7.1.3. Software Engineering for Activity Recognition

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Ines Sarray, Daniel Gaffé, Julien Badie, François Brémond, Minh Khue Phan Tran.

The contributions for this research axis are:

- A Synchronous Approach to Activity Recognition (see 7.21)
- A Probabilistic Activity Description Language (see 7.22)

7.2. Late Fusion of Multiple Convolutional Layers for Pedestrian Detection

Participants: Ujjwal Ujjwal, François Brémond, Aziz Dziri [VEDECOM], Bertrand Leroy [VEDECOM].

One of the prominent problems in pedestrian detection is handling scale and occlusion. These problems are quite well aligned with the recent interests in autonomous vehicles. Successful detection of far-scale pedestrians can assist the vehicle in making safety maneuvers well ahead in time, thereby promoting a safer traffic environment. The same is true for surveillance systems in high security environment like airports and ports.

We propose a system design for pedestrian detection by leveraging the power of multiple convolutional layers explicitly (see Figure 5). We quantify the effect of different convolutional layers on the detection of pedestrians of varying scales and occlusion level. We show that earlier convolutional layers are better at handling small-scale and partially occluded pedestrians. We take cue from these conclusions and propose a pedestrian detection system design based on Faster-RCNN which leverages multiple convolutional layers by late fusion. In our design, we introduce height-awareness in the loss function to make the network emphasize on pedestrian heights which are misclassified during the training process. The proposed system design achieves a log-average miss-rate of 9.25% on the caltech-reasonable dataset. This is within 1.5% of the current state-of-art approach, while being a more compact system. The work was published in the 15^{th} IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS)-2018 [51].

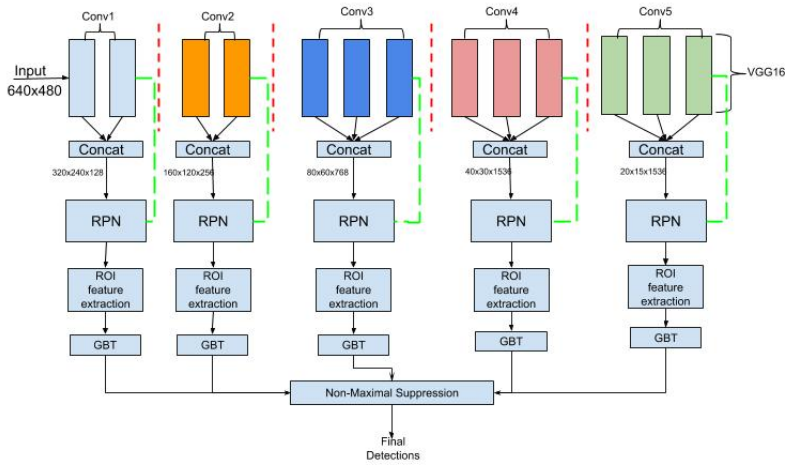


Figure 5. Block diagram of our proposed Multiple-RPN pedestrian detection system

7.3. Deep Learning applied on Embedded Systems for People Tracking

Participants: Juan Diego Gonzales Zuniga, Thi Lan Anh Nguyen, Francois Brémond, Serge Tissot [KONTRON].

Keywords: Deep Learning, Embedded Systems, Multiple Object Tracking

One of the main issues with people detection and tracking is the amount of resources it consumes for real time applications. 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. The applications for an embedded hardware with these capabilities are limitless: automotive, security and surveillance, augmented reality and health-care just to name a few. But the state-of-the-art architectures are mostly focused on accuracy rather than resource consumption.

In our work, we have to consider improving the systems' accuracy and reducing resources for real-time applications. We are creating a shared effort of hardware adaptation and agnostic software optimization for all deep learning based solutions.

We here focus our work on two separated but linked problems.

First, we improve the feature representation of tracklets for the Multiple Object Tracking challenge. This is based on the concept of Residual Transfer Learning [44]. Second, we are creating a viable platform to run our algorithms on different target hardware, mainly, Intel Xeon Processors, FPGAs and AMD GPUs.

7.3.1. Residual Transfer Learning :

We present a smart training alternative for transfer learning based on the concept of ResNet [65]. In ResNet, a layer learns the estimate residual between the input and output signals. We cast transfer learning as a residual learning problem, since the objective is to close the gap between the initial network and the desired one. Achieving this goal is done by adding residual units for a number of layers to an existing model that needs to be transferred from one task to another. The existing model can thus be able to perform a new task by adding and optimizing residual units as shown in Figure 6 . The main advantage of using residual units for transfer learning is the flexibility in terms of modelling the difference between two tasks.

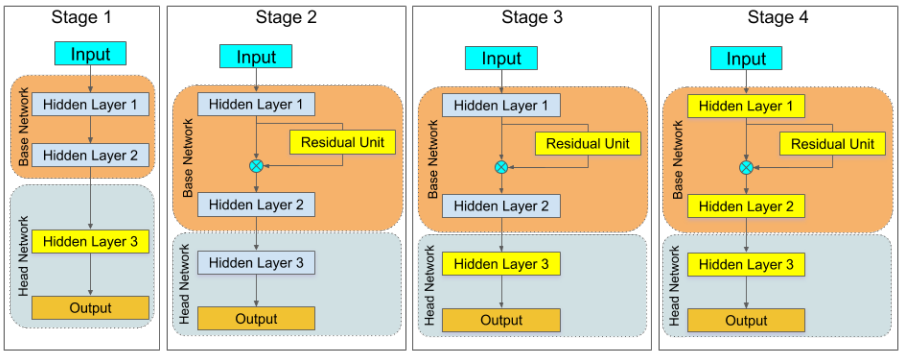


Figure 6. Training stages of Residual Transfer Learning method. Each stage only trains the layers shown in yellow, and fixes the layers in grey. The residual units are added at the second stage

7.3.2. Deep Learning Platform on Multiple Target Hardware :

Deep learning algorithms need an extensive allocation of resources to be executed, most of the research is accomplished under NVIDIA GPU's. This is limiting because it reduces the possibilities on how to optimize certain blocks that directly depend on the hardware configuration. The main cause is the lack of a flexible platform that would support different targets: AMD GPUs, Intel Xeon processors and specialized FPGAs.

We work with two hardware based platforms; ROCm and Openvino. The ROCm stack, shown in Figure 7 , allows us to perform a variety of layer computations on AMD GPUs. We have managed to import different deep learning networks such as VGG16, ResNet and Inception to AMD's Radeon graphics card. On the other hand, Openvino's main goal is to reduce the inference time of a network. For this solution, we count on the Openvino Optimizer, shown in Figure 8 , which main goal is to transform the network model from Caffe or Tensorflow into an Inference Model for Intel's processors and FPGAs.

We also built docker images on top of the above mention platforms, this is done to speed the deployment stage by being operating system independent.

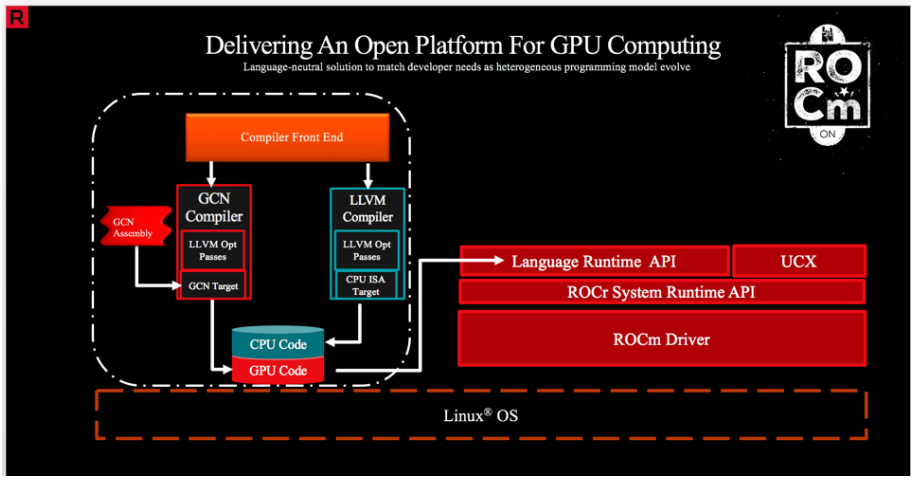


Figure 7. The ROCm System Runtime is language independent and makes heavy use of the Heterogeneous System Architecture.

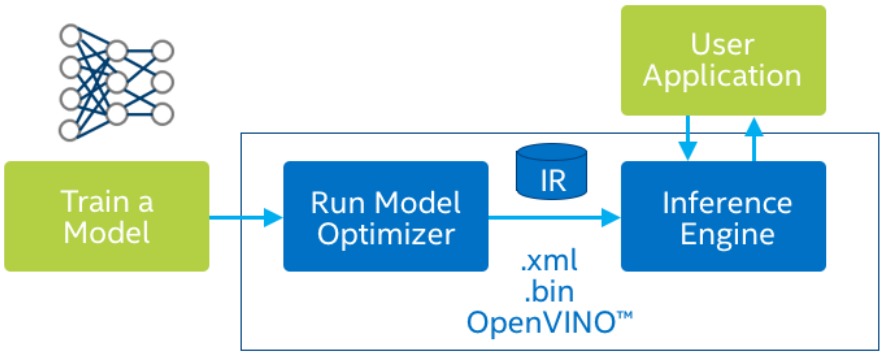


Figure 8. Openvino: When you run a pre-trained model through the Model Optimizer, your output is an Intermediate Representation of the network.

7.4. Cross Domain Residual Transfer Learning for Person Re-identification

Participants: Furqan Khan, Francois Brémond.

Keywords: multi-shot person re-identification, transfer learning, residual unit

Person re-identification (re-ID) refers to the retrieval task where the goal is to search for a given person (query) in disjoint camera views (gallery). Performance of appearance based person re-ID methods depends on the similarity metric and the feature descriptor used to build a person's appearance model from given image(s).

A novel way is proposed to transfer model weights from one domain to another using residual learning framework instead of direct fine-tuning. It also argues for hybrid models that use learned (deep) features and statistical metric learning for multi-shot person re-identification when training sets are small. This is in contrast to popular end-to-end neural network based models or models that use hand-crafted features with adaptive matching models (neural nets or statistical metrics). Our experiments demonstrate that a hybrid model with residual transfer learning can yield significantly better re-identification performance than an end-to-end model when training set is small. On iLIDS-VID [78] and PRID [67] datasets, we achieve rank1 recognition rates of 89.8% and 95%, respectively, which is a significant improvement over state-of-the-art.

7.4.1. Residual Transfer Learning

We use RTL to transfer a model trained on Imagenet [63] for object classification to perform person re-ID. We chose to use 16-layer VGG model due to its superior performance in comparison to AlexNet and overlooked ResNet for its extreme depth because our target datasets are small and do not warrant such a deep model for higher performance.

One advantage of using residual learning [66] for model transfer is that it allows more flexibility in terms of modeling the difference between two tasks through a number of residual units and their composition. We noted that when residual units are added to the network with a different network head, training loss is significantly higher in the beginning which pushes the network far away from pre-trained solution by trying to over compensate through residual units. To avoid this, we propose to train the network in 4 stages, with fourth stage being optional (Fig. 9). The proposed work has been published in [45].

- **Stage 1**: In the first stage, we replace original head of the network with a task specific head and initialize it randomly. At this stage, we do not add any residual units to the network and train only the parameters of the replaced head of the network. Thus only the head layers are considered to contribute to the loss. This allows the network to learn noisy high level representation for the desired task and decrease the network loss without affecting lower order layers.
- **Stage 2**: In the second stage, we add residual units to the network and initialize them randomly. Then we freeze all other layers, including the network head, and optimize the parameters of added residual units. As the head and other layers are fixed, residual units are considered as the source of loss. As we start with a reasonably low loss value, residual units are not forced to over compensate for the loss.
- **Stage 3**: In the third stage, we train the network by learning parameters of both added residual units and network head, thus allowing both the lower and higher order representations to adjust to the specific task.
- **Stage 4 (Optional)**: We noticed in our experiments on different datasets that the loss function generally gets low enough by the end of third stage. However, if needed, the whole network can be trained to further improve performance.

7.4.2. Conclusion

When using identity loss and large amount of training data, RTL gives comparable performance to direct fine-tuning of network parameters. However, the performance difference between two transfer learning approaches is considerably in favor of RTL when training sets are small. The reason is that when using RTL only a few parameters are modified to compensate for the residual error of the network. Still, the higher order layers of the network are prone to over-fitting. Therefore, we propose using hybrid models where higher order domain

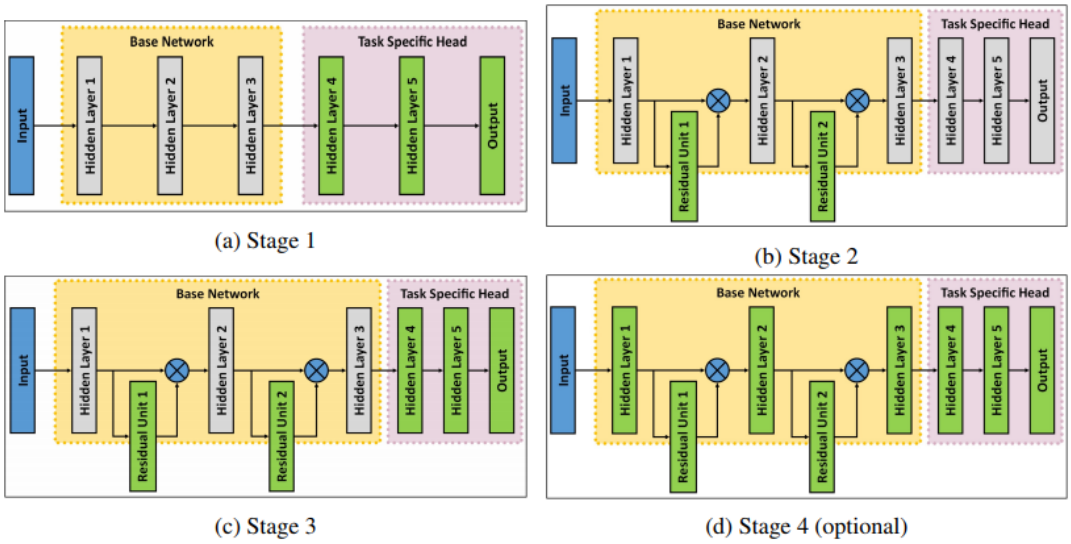


Figure 9. Residual Transfer Learning in 4 stages. During each stage only the selected layers (shown in green) are trained. Residual Units are added to the network after first stage of RTL.

specific layers are replaced with statistical metric learning. We demonstrate that the hybrid model performs significantly better on small datasets and gives comparable performance on large datasets. The ability of the model to generalize well from small amount of data is crucial for practical applications because frequent data collection in large amount for training is nit possible.

7.5. Face-based Attribute Classification and Manipulation

Participants: Abhijit Das, Antitza Dantcheva, Francois Brémond.

Keywords: Face, Attribute, GAN, Biometrics

Due to the biasness of face analytic datasets, with respect to factors such as age, gender, ethnicity, pose and resolution, systems based on a skewed training dataset are bound to produce skewed results. Further, it has been exhibited in the literature [59] that such biases may have serious impacts on performance in challenging situations where the outcome is critical. In order to progress toward balanced face recognition and attribute estimation, the 1st International Workshop on Bias Estimation in Face Analytics was organized in conjunction with ECCV 2018. The workshop also organized a challenge to introduce a well-balanced dataset across multiple factors: age, gender, ethnicity, pose and resolution and requested for algorithms to estimate biases.

We proposed a Multi-Task Convolutional Neural Network (MTCNN) algorithm that jointly learned [37] gender, age and ethnicity by a loss function involving joint dynamic loss weight adjustment and was successful, as well as relatively unbiased in estimating age, gender and ethnicity. Our algorithm was found to be the best algorithm focusing the aim of the competition and the above mentioned research problem.

7.5.1. Generative Adversal Network (GAN)

models are autoregressive models depending on the global information, which can be potentially affected by its employment on local feature/ attribute-based erasuring. In addition, these models are typically trained depending on the maximum likelihood to find the intense difference between the regression domains, as a result after a certain limit of learning it can produce very naive development in the interpolation of the

regression carried out for the purpose of local attribute removal. Hence, to mitigate an aforementioned couple of pitfalls we propose a method for localizing the Cycle GAN (C-GAN) for local feature-based regression. We trained the C-GAN with domain-specific local feature and end model was recurrently imposed on the testing images. We experimented the Local C-GAN (L-C-GAN) on facial attribute (eyeglass and moustache/ bearded) auto-regression. Our qualitative performance on partial CelebA dataset and a couple of datasets we collected is promising. Moreover, ensuring the facial attributes have also been found to achieve better performance accuracy with respect to the presence of these attributes.

7.6. From Attribute-labels to Faces: Face Generation using a Conditional Generative Adversarial Network ,

Participants: Yaohui Wang, Antitza Dantcheva, Francois Brémond.

Keywords: Generative Adversarial Networks, Face generation

Facial attributes are instrumental in semantically characterizing faces. Automated classification of such attributes (i.e., age, gender, ethnicity) has been a well studied topic. We here seek to explore the inverse problem, namely given attribute-labels the *generation of attribute-associated faces*. The interest in this topic is fueled by related applications in law enforcement and entertainment. In this work, we propose two models for attribute-label based facial image and video generation incorporating 2D (see Figure 10) and 3D (see Figure 11) deep conditional generative adversarial networks (DCGAN). The attribute-labels serve as a tool to determine the specific representations of generated images and videos. While these are early results (see Figure 12 and 13), our findings indicate the methods' ability to generate realistic faces from attribute labels.

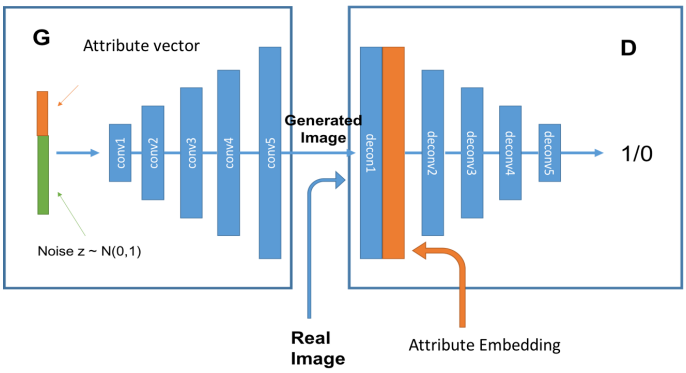


Figure 10. Architecture of proposed 2D method consisting of two modules, a discriminator D and a generator G. While D learns to distinguish between real and fake images, classifying based on attribute-labels, G accepts as input both, noise and attribute-labels in order to generate realistic face images.

7.7. Face Analysis in Structured Light Images

Participants: Vikas Thamizharasan, Antitza Dantcheva, Francois Brémond.

Keywords: Structured light, Face analysis

The main objective has been to perform face analysis tasks like authentication, gender, age and ethnicity classification by generating low-dimensional face embedding from the raw data acquired from structured light (see Figure 14) sensors using deep learning techniques. In this context we studied depth/disparity map extraction (see Figure 15), as well as other models.

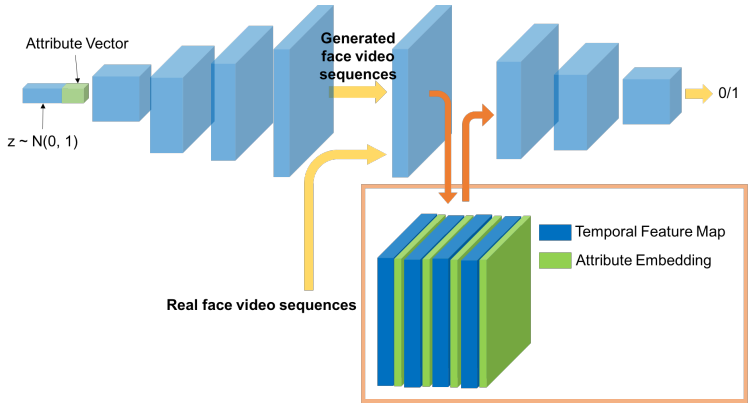


Figure 11. Architecture of proposed 3D model for face video generation

7.8. Deep-Temporal LSTM for Daily Living Action Recognition

Participants: Srijan Das, Michal Koperski, Francois Brémond, Gianpiero Francesca.

Keywords: Temporal sequences, Appearance, LSTM

We have proposed to improve the traditional use of RNNs by employing a many to many model for video classification. We analyzed the importance of modeling spatial layout and temporal encoding for daily living action recognition. Many RGB methods focus only on short term temporal information obtained from optical flow. Skeleton based methods on the other hand show that modeling long term skeleton evolution improves action recognition accuracy. In this work, we proposed a deep-temporal LSTM architecture (see fig. 16) which extends standard LSTM and allows better encoding of temporal information. In addition, we have proposed to fuse 3D skeleton geometry with deep static appearance. We validated our approach on publicly available datasets (CAD60, MSRDailyActivity3D and NTU-RGB+D), achieving competitive performance as compared to the state-of-the art. The proposed framework has been published in AVSS 2018 [39].

7.9. Spatio-Temporal Grids for Daily Living Action Recognition

Participants: Srijan Das, Kaustubh Sakhalkar, Michal Koperski, Francois Brémond.

Keywords: Spatio-temporal, Grids, Multi-modal

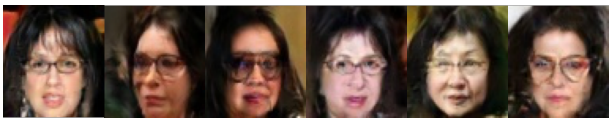
This work addresses the recognition of short-term daily living actions from RGB-D videos. Most of the existing approaches ignore spatio-temporal contextual relationships in the action videos. So, we have proposed to explore the spatial layout to better model the appearance. In order to encode temporal information, we divided the action sequence into temporal grids. We address the challenge of subject invariance by applying clustering on the appearance features and velocity features to partition the temporal grids. We validated our approach on four public datasets. The results show that our method is competitive with the state-of-the-art. The proposed architecture has been published in ICVGIP 2018 [40].

7.10. A New Hybrid Architecture for Human Activity Recognition from RGB-D videos

Participants: Srijan Das, Monique Thonnat, Kaustubh Sakhalkar, Michal Koperski, Francois Brémond, Gianpiero Francesca.

Keywords: Visual cues, Data fusion, RGB-D videos

(a) no glasses, female, black hair, smiling, young



(b) glasses, female, black hair, not smiling, old

(c) no glasses, male, no black hair, smiling, young

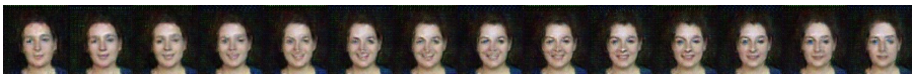
(d) glasses, male, no black hair, not smiling, old

Figure 12. Example images generated by the proposed 2D model.

(a) male, adolescent

(b) male, adult

(c) female, adolescent



(d) male, adult

Figure 13. Chosen output samples from 3DGAN

Figure 14. Structured light. A calibrated camera and projector (typically both near infrared) are placed at a fixed, known baseline. The structured light pattern helps establish correspondence between observed and projected pixels. Depth is derived for each corresponding pixel through triangulation. The process is akin to two stereo cameras, but with the projector system replacing the second camera, and aiding the correspondence problem.

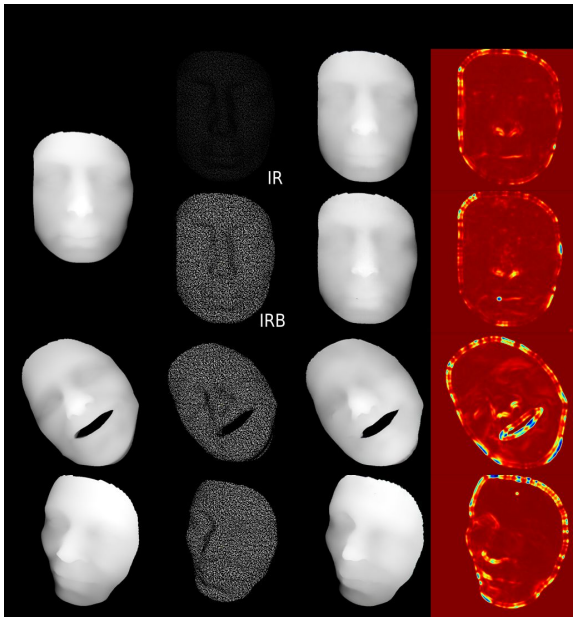


*Figure 15. *IR - Infrared image, IRB - Binarized Infrared image*

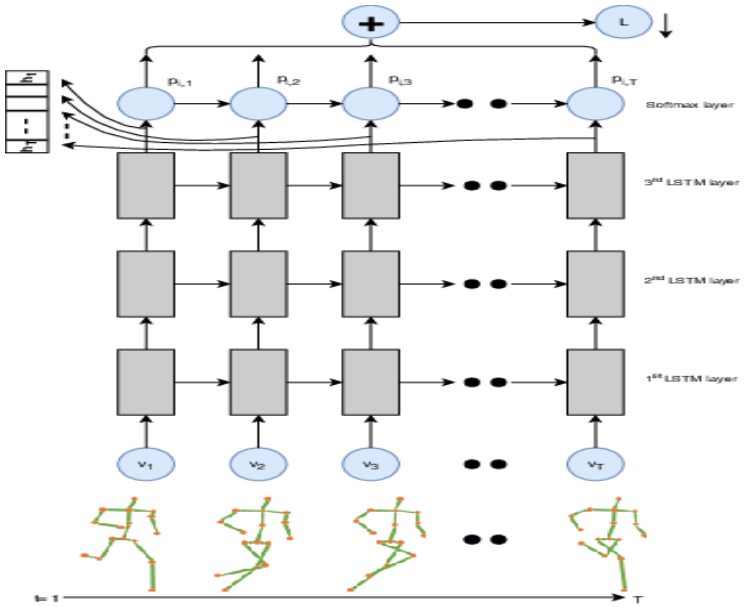


Figure 16. Framework of the deep-temporal LSTM proposed approach in [39]

Activity Recognition from RGB-D videos is still an open problem due to the presence of large varieties of actions. We have proposed a new architecture by mixing a high level handcrafted strategy and machine learning techniques. In order to address the problem of large variety of actions, we proposed a novel two level fusion strategy to combine motion, appearance and 3D pose information. For 3D pose information, we use the work published in AVSS 18 (described above). As similar actions are common in daily living activities, we also proposed a mechanism for similar action discrimination using dedicated SVMs. We validated our approach on four public datasets, CAD-60, CAD-120, MSRDailyActivity3D, and NTU-RGB+D improving the state-of-the-art results on them. The proposed architecture has been published in the industrial session of MMM 2019 [41].

7.11. Where to Focus on for Human Action Recognition?

Participants: Srijan Das, Arpit Chaudhary, Francois Brémond, Monique Thonnat.
Keywords: Spatial attention, Body parts, End-to-end

We proposed a spatial attention mechanism based on 3D articulated pose to focus on the most relevant body parts involved in the action. For action classification, we proposed a classification network compounded of spatio-temporal subnetworks modeling the appearance of human body parts and RNN attention subnetwork implementing our attention mechanism. Furthermore, we trained our proposed network end-to-end using a regularized cross-entropy loss, leading to a joint training of the RNN delivering attention globally to the whole set of spatio-temporal features, extracted from 3D ConvNets. Our method outperforms the State-of-the-art methods on the largest human activity recognition dataset available to-date (NTU RGB+D Dataset) which is also multi-views and on a human action recognition dataset with object interaction (Northwestern-UCLA Multiview Action 3D Dataset). The proposed framework will be published in WACV 2019. Sample visual results displaying the attention scores attained for each body parts can be seen in fig. 17 .

7.12. Online Temporal Detection of Daily-Living Human Activities in Long Untrimmed Video Streams

Participants: Abhishek Goel, Abdelrahman G. Abubakr, Michal Koperski, Francois Brémond.
keywords: Daily-living activity recognition, Human activity detection, Video surveillance, Smarthome

Many approaches were proposed to solve the problem of activity recognition in short clipped videos, which achieved impressive results with hand-crafted and deep features. However, it is not practical to have clipped videos in real life, where cameras provide continuous video streams in applications such as robotics, video surveillance, and smart-homes. Here comes the importance of activity detection to help recognizing and localizing each activity happening in long videos. Activity detection can be defined as the ability to localize starting and ending of each human activity happening in the video, in addition to recognizing each activity label. A more challenging category of human activities is the daily-living activities, such as eating, reading, cooking, etc, which have low inter-class variation and environment where actions are performed are similar. In this work we focus on solving the problem of detection of daily-living activities in untrimmed video streams. We introduce new online activity detection pipeline that utilizes single sliding window approach in a novel way; the classifier is trained with sub-parts of training activities, and an online frame-level early detection is done for sub-parts of long activities during detection. Finally, a greedy Markov model based post processing algorithm is applied to remove false detection and achieve better results. We test our approaches on two daily-living datasets, DAHLIA and GAADRD, outperforming state of the art results by more than 10%. The proposed work has been published in [43].

7.12.1. The Work Flow of processing untrimmed videos is composed of three tasks:

- **Feature extraction** consists in extracting the Person-Centered CNN (PC-CNN) features as shown in fig. 18 .
- **Classifier Training:** All training videos are first divided into relatively small windows of size W frames, which represent activity sub-videos (subparts). Then the features are generated for all these windows and the training is done with linear SVM classifier using all activities sub-videos.
- **Majority voting filtering**, as depicted in fig. 19 , looks up for neighbors within a certain range that have the same label apply majority-voting between the labels

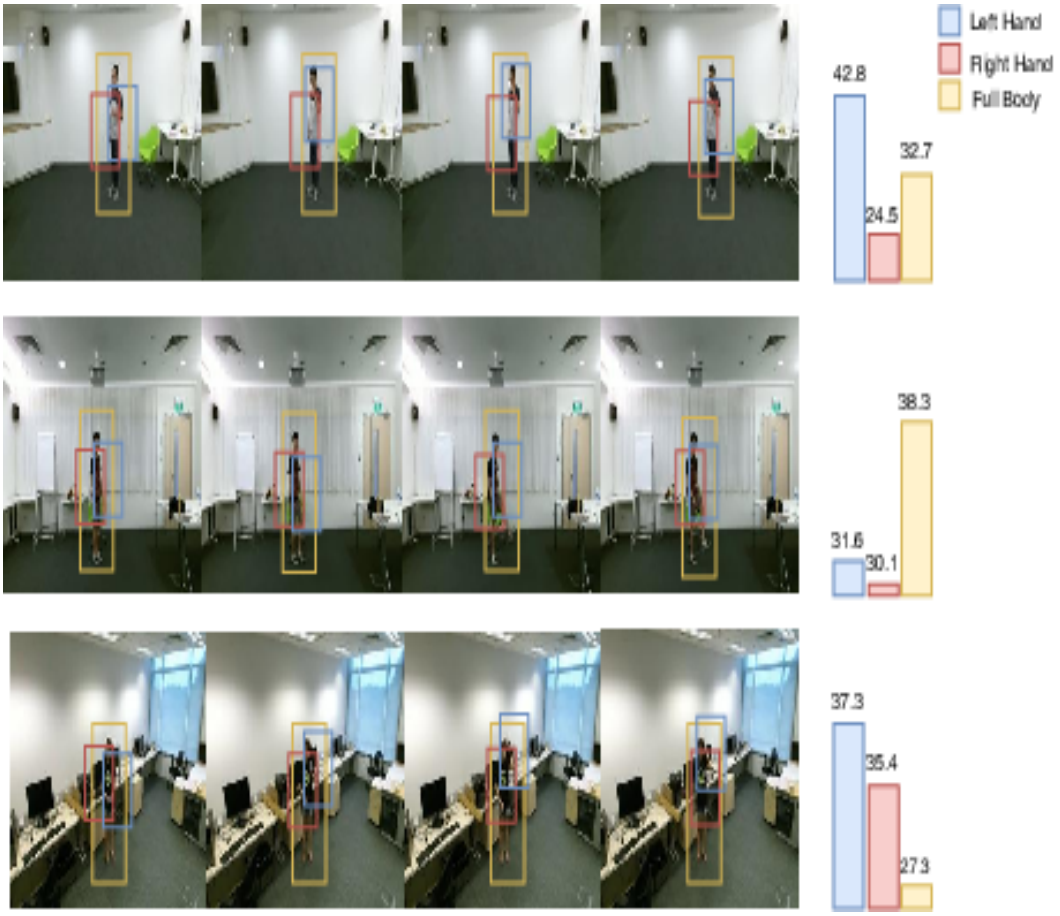


Figure 17. Example of video sequences with their respective attention scores. The action categories presented are drinking water with left hand (1st row), kicking (2nd row) and brushing hair with left hand (last row).

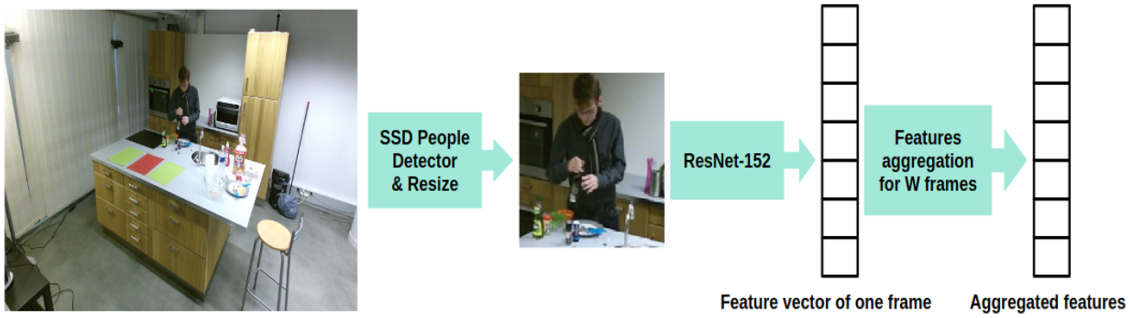


Figure 18. Extracting the Person-Centered CNN (PC-CNN) features

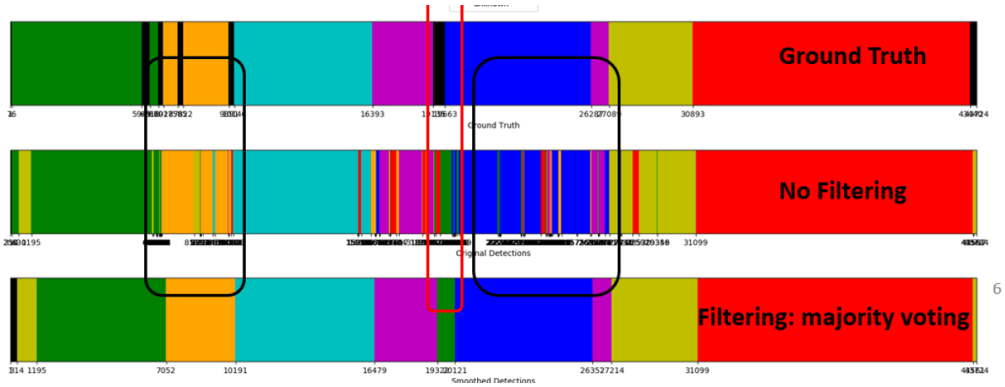


Figure 19. Post-filtering

7.13. Activity Detection in Long-term Untrimmed Videos by discovering sub-activities

Participants: Farhood Negin, Abhishek Goel, Abdelrahman G. Abubakr, Gianpiero Francesca, Francois Brémond.

Keywords: Activity detection, Semi-supervised learning, Sub-activity detection.

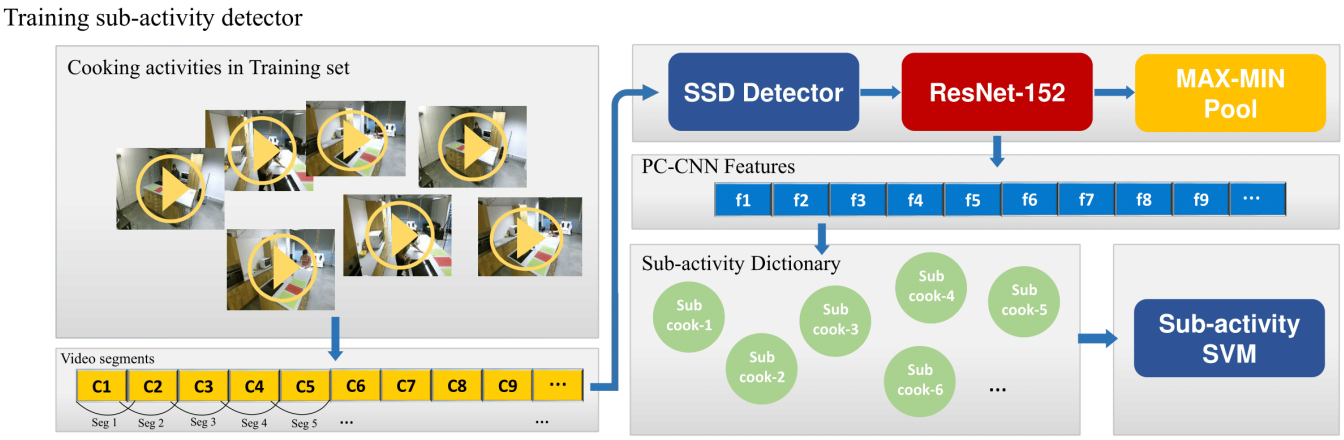


Figure 20. The process of extracting PC-CNN features and training of a weakly supervised sub-activity detector for the "Cooking" activity.

Detecting temporal delineation of activities is important to analyze large-scale videos. However, there are still challenges yet to be overcome in order to have an accurate temporal segmentation of activities. Detection of daily-living activities is even more challenging due to their high intra-class and low inter-class variations, complex temporal relationships of sub-activities performed in realistic settings. To tackle these problems, we

propose an online activity detection framework based on the discovery of sub-activities. We consider a long-term activity as a sequence of short-term sub-activities. Our contributions can be summarized as follows:

- We introduce a new online frame-level activity detection pipeline which uses single-sized window approach. A weakly supervised classifier is trained directly on sub-activities discovered by clustering and operates on test videos to capture sub-activities of long videos within a fixed temporal window.
- To alleviate the noisy detections especially in activity boundaries, we propose a novel greedy post-processing method based on Markov models.
- We have extensively evaluated our proposed method on untrimmed videos from DAHLIA [68] and GAADRD [77] datasets and achieved state-of-the-art performances.

7.13.1. Proposed Method:

Our framework produces frame-level activity labels in an online manner by two major steps followed by a novel greedy post-processing technique. In order to handle long activities, activities are decomposed into a sequence of fixed-length overlapping temporal clips. We then extract deep features from the clips. We suggested a person-centric feature (PC-CNN) based on SSD detector that satisfies required processing efficiency of online systems. We then proposed a weakly-supervised method for the discovery of sub-activities of long-term activities which benefits from clustering and model selection methods to find the optimal sub-activities of the given activities. In order to characterize each activity with constituent sub-activities, we use K-means to cluster that activity's clips and construct a specific sub-activity dictionary. Therefore, we have one sub-activity dictionary for each main activity. We represent an activity sequence with sub-activity assignments using the trained dictionary. Then, for each activity class, we train a binary SVM classifier (one versus all) based on its sub-activities (Figure 20). The trained classifiers are then simultaneously used to produce frame-level activity labels with the help of a sliding window architecture. It should be noticed that unlike multi-scale sliding window methods, we only use a single fixed-size temporal window thanks to recognition of fixed length sub-activities. Finally, assuming temporal progression of sub-activities, we developed a greedy algorithm based on Markov models to refine noisy sub-activity proposals in middle and boundary regions of long activities. We evaluated the proposed method on two daily-living activity datasets and achieved state-of-the-art performances.

Table 1. The activity detection results obtained on the DAHLIA. Values in bold represent the best performance.

	ELS			**Max Subgraph Search**			**DOHT (HOG)**			**Sub Activity**		
	FA_1	F_score	IoU	FA_1	F_score	IoU	FA_1	F_score	IoU	FA_1	F_score	IoU
View 1	0.18	0.18	0.11	-	0.25	0.15	0.80	0.77	0.64	**0.85**	**0.81**	**0.73**
View 2	0.27	0.26	0.16	-	0.18	0.10	0.81	0.79	0.66	**0.87**	**0.82**	**0.75**
View 3	0.52	0.55	0.39	-	0.44	0.31	0.80	**0.77**	0.65	**0.82**	0.76	**0.69**

Table 2. Detection results obtained on the GAADRD dataset.

Method	FA_1	F_score	IoU
simple sliding window(HOG)	0.68	0.52	0.40
simple sliding window(PC-CNN)	0.61	0.55	0.44

Tables 1 and 2 show the results of applying the developed frameworks on DAHLIA and GAADRD respectively. It can be noticed that in DAHLIA dataset (compared to [71], [61], [60]), we significantly outperformed state-of-the-art results in all of the categories except in camera view 3 when the F-Score metric is used. We reported the results of GAADRD dataset with the two types of features HOG and PC-CNN. As it can be seen, even with hand-crafted features our framework produces comparable results. In future work, we

are going to improve the sub-activity discovery algorithm by making it able to distinguish similar sub-activities in two different activities.

7.14. Video based Face Analysis for Health Monitoring

Participants: Abhijit Das, Antitza Dantcheva, Francois Brémond.

Keywords: Face, Attribute, GAN, Biometrics

Video based analysis in severely demented Alzheimer's Disease (AD) patients can be helpful for the analysis of their neuropsychiatric symptom such as apathy, depression. Even for the doctors it can be hard to know whether a person has depression or apathy. The main difference is that a person with depression will have feelings of sadness, be tearful, feel hopeless or have low self-esteem. Whereas, symptoms of person suffering from apathy can make the person's life less enjoyable. Therefore, a psychological protocol scenario can be used for video-based emotion analysis and facial movement can be used for discriminating apathetic person and non-apathetic person.

We proposed to use a) the facial expressions (neutral + 6 basic emotions: anger, disgust, happiness, surprise, sadness, fear) extracted using 50 layer Resnet, b) facial movements employing 68 facial landmark points, c) action unit intensity and frequency for AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, and 45 using OpenFace and d) lip movements employing the 3D mouth open vector using the mean of upper lip and mean of bottom lip extracted from the facial landmarks detected around the lip as feature for each frame of the video. We post-process the features and calculated the amplitude, SD (Standard Deviations) and mean of each clip (10 seconds per clip) and these features were passed inputs to GRU. The GRU is connected to the Fully Connected layers, these fully connected features are mean pooled to get the apathy/non-apathy classification.

7.15. Mobile Biometrics

Participants: Abhijit Das, Antitza Dantcheva, Francois Brémond.

Keywords: Mobile biometrics

The prevalent commercial deployment of mobile biometrics as a robust authentication method on mobile devices has fueled increasingly scientific attention. Motivated by this, in this work [38] we seek to provide insight on recent development in mobile biometrics. We present parallels and dissimilarities of mobile biometrics and classical biometrics, enumerate related strengths and challenges. Further, we provide an overview of recent techniques in mobile bio-metrics, as well as application systems adopted by industry. Finally, we discuss open research problems in this field.

7.16. Comparing Methods for Assessment of Facial Dynamics in Patients with Major Neurocognitive Disorders

Participants: Yaohui Wang, Antitza Dantcheva, Francois Brémond.

Keywords: Face Analysis

Assessing facial dynamics in patients with major neurocognitive disorders and specifically with Alzheimer's disease (AD) has shown to be highly challenging. Classically such assessment is performed by clinical staff, evaluating verbal and non-verbal language of AD-patients, since they have lost a substantial amount of their cognitive capacity, and hence communication ability. In addition, patients need to communicate important messages, such as discomfort or pain. Automated methods would support the current healthcare system by allowing for telemedicine, *i.e.,* lesser costly and logistically inconvenient examination. In this work [52], we compare methods for assessing facial dynamics such as talking, singing, neutral and smiling in AD-patients, captured during music mnemotherapy sessions. Specifically, we compare 3D ConvNets (see Figure 21), Very Deep Neural Network based Two-Stream ConvNets (see Figure 22), as well as Improved Dense Trajectories. We have adapted these methods from prominent action recognition methods and our promising results suggest that the methods generalize well to the context of facial dynamics. The Two-Stream ConvNets in combination with ResNet-152 obtains the best performance on our dataset (Table 3), capturing well even minor facial dynamics and has thus sparked high interest in the medical community.

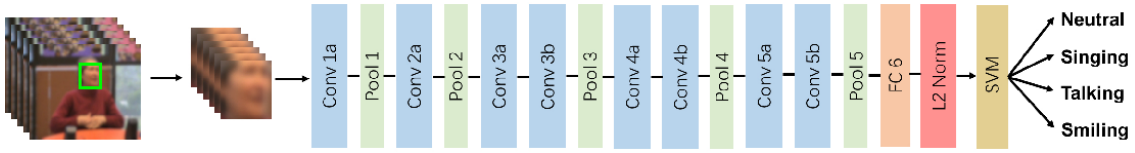
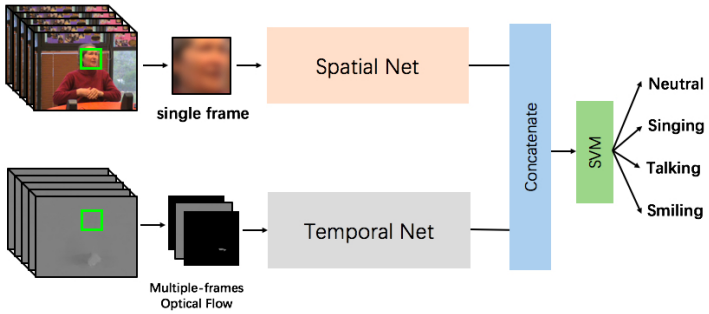
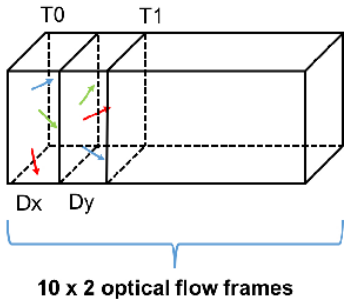


Figure 21. ***C3D based facial dynamics detection:*** *For each video sequence, faces are detected and the face sequences are passed into a pre-trained C3D network to extract a 4096-dim feature vector for each video. Finally a SVM classifier is trained to predict the final classification result. We have blurred the faces of the subject in this figure, in order to preserve the patient's privacy.*



(a) Two-Stream Architecture



(b) Stacked Optical Flow Field volume

Figure 22. (a) While the spatial ConvNet accepts a single RGB frame as input, the temporal ConvNet's input is the D_x and D_y of 10 consecutive frames, namely 20 input channels. Both described inputs are fed into the Two-stream ConvNets, respectively. We use in this work two variations of Very Deep Two Stream ConvNets, incorporating VGG-16 [76] ResNet-152 [65] for both streams respectively. (b) The optical flow of each frame has two components, namely D_x and D_y. We stack 10 times D_y after D_x for each frame to form a 20 frames length input volume.

Table 3. Classification accuracies of C3D, Very Deep Two-Stream ConvNets, iDT, as well as fusion thereof on the presented ADP-dataset. We report the Mean Accuracy (MA) associated to the compared methods. Abbreviations used: SN...Spatial Net, TN...Temporal Net.

Method	**MA (%)**
C3D	67.4
SN of Two-Stream ConvNets (VGG-16)	65.2
TN of Two-Stream ConvNets (VGG-16)	69.9
Two-Stream ConvNets (VGG-16)	76.1
SN of Two-Stream ConvNets (ResNet-152)	69.6
TN of Two-Stream ConvNets (ResNet-152)	75.8
Two-Stream ConvNets (ResNet-152)	76.4
iDT	61.2
C3D + iDT	71.1
Two-Stream ConvNets (VGG-16) + iDT	78.9
Two-Stream ConvNets (ResNet-152) + iDT	**79.5**

7.17. Combating the Issue of Low Sample Size in Facial Expression Recognition

Participants: S L Happy, Antitza Dantcheva, Francois Brémond.

Keywords: Face analysis, Expression recognition

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 contain images and videos with discrete disjoint labels of profound emotions, real-life data contain jointly occurring emotions and expressions of different intensities. Reliable data annotation is a major problem in this field, which results in publicly available datasets with low sample size. Transfer learning [73], [64] is usually used to combat the low sample size problem by capturing high level facial semantics learned on different tasks. However, 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 here propose a weakly supervised learning technique for expression classification, which leverages the information of unannotated data. In weak supervision scenarios, a portion of training data might not be annotated or wrongly annotated [79]. Crucial in our approach is that we first train a convolutional neural network (CNN) with label smoothing in a supervised manner and proceed to tune the CNN-weights with both labelled and unlabelled data simultaneously. The learning method learns the expression intensities in addition to classifying them into discrete categories. This bootstrapping of a fraction of unlabelled samples, replacing labelled data for model-update, while maintaining the confidence level of the model on supervised data improves the model performance.

Table 4. Cross database classification performance when using CK+ database for training.

Test databases	Percentage of training data		
	25%	50%	80%
CK+ (test-set)	88.79%	91.29%	95.16%
RaFD	64.25%	65.25%	78.46%
lifespan	35.13%	40.51%	60.83%

7.17.1. Experimental Results

Experiments were conducted on three publicly available expression datasets, namely CK+, RaFD, and lifespan. Substantial experiments on these datasets demonstrate large performance gain in cross-database performance,

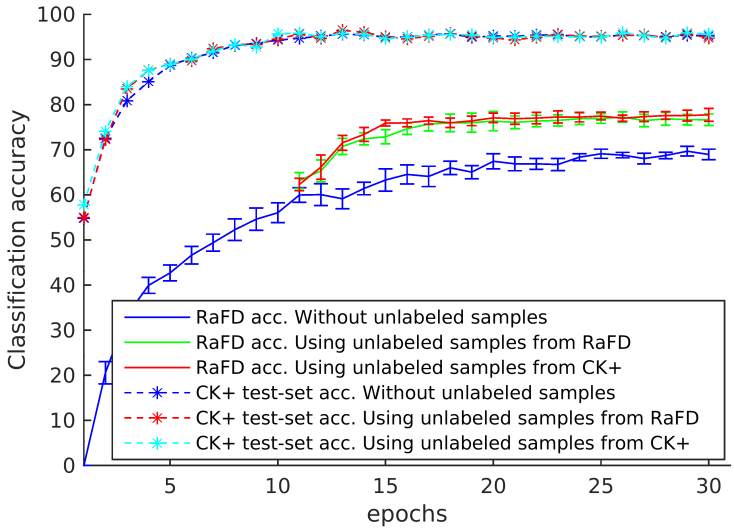


Figure 23. Cross-database experiments show significant performance improvement.

as well as show that the proposed method achieves to learn different expression intensities, even when trained with categorical samples. As can be seen in Fig. 23 , when the model is trained on CK+ with unlabelled data, the model-performance improved by 11% in RaFD cross database evaluation. We observe that the use of unlabelled data from either CK+ or RaFD results in similar performances. Utilizing unlabelled images from CK+, the network sees varying expression-intensities and adapts to it. Table 4 reports the self and cross-database classification results with respect to varying number of training samples. Significant classification accuracy has been obtained with merely 25% of the training data. Use of a larger labelled training set strikingly boosts the cross-database performance. In future, we are planning to further improve the performance with unsupervised learning of expression patterns.

7.18. Serious Exergames for Cognitive Stimulation

Participants: Guillaume Sacco, Monique Thonnat.

Keywords: Neurocognitive disorders, Serious games, Geriatrics, Executive functions, Physical exercise, Cognitive training

A PhD thesis has been defended on the 8th of June at Nice University on this topic by Guillaume Sacco. This thesis presents a clinical and therapeutic approach aiming to create new care for patients with neurocognitive disorder. Serious exergames are serious video games integrating physical activity. Serious exergames could be tools to product enriched environment associating physical exercise and cognitive training. The aim of this thesis is to investigate whether serious exergames can contribute to the non-pharmacological management of neurocognitive disorders. In this thesis we have made two types of contributions. The first type are general contributions. One presents our integrative clinical approach associating physical exercise and cognitive training using serious exergames. The other one presents recommendations concerning the use of serious exergames. The second type of contributions are experimental. The first one aims to confirm a theoretical base of our clinical approach. The two other experiments assess the implementation of our approached in a population of patients with neurocognitive disorder. This year the integrative clinical approach associating physical exercise and cognitive training using serious exergames has been published [32] and presented at the International Conference on Gerontechnology ISG in Saint Petersburg, Florida, USA in May 2018.

7.19. Speech-Based Analysis for older people with dementia

Participants: Alexandra König, Philippe Robert, Nicklas Linz, Johannes Tröger, Jan Alexandersson.

Keywords: Alzheimer's disease, Dementia, Mild cognitive impairment, Neuropsychology , Assessment, Semantic verbal fluency, Speech recognition, Speech processing, Machine learning

7.19.1. Fully Automatic Speech-Based Analysis of the Semantic Verbal Fluency Task:

Semantic verbal fluency (SVF) tests are routinely used in screening for mild cognitive impairment (MCI). In this task, participants name as many items as possible of a semantic category under a time constraint. Clinicians measure task performance manually by summing the number of correct words and errors. More fine-grained variables add valuable information to clinical assessment, but are time-consuming. Therefore, the aim of this study is to investigate whether automatic analysis of the SVF could provide measures as accurate as the manual ones and thus, support qualitative screening of neurocognitive impairment.

Methods: SVF data were collected from 95 older people with MCI (n = 47), Alzheimer's or related dementias (ADRD; n = 24), and healthy controls (HC; n = 24). All data were annotated manually and automatically with clusters and switches. The obtained metrics were validated using a classifier to distinguish HC, MCI, and ADRD.

Results: Automatically extracted clusters and switches were highly correlated (r = 0.9) with manually established values, and performed as well on the classification task, separating HC from persons with ADRD (area under curve [AUC] = 0.939) and MCI (AUC = 0.758).

Conclusion: The results show that it is possible to automate fine-grained analyses of SVF data for the assessment of cognitive decline [70].

7.19.2. Language Modelling in the Clinical Semantic Verbal Fluency Task:

We employed language modelling (LM) as a natural technique to model production in this task. Comparing different LMs, we show that perplexity of a person's SVF production predicts dementia well (F1 = 0.83). Demented patients show significantly lower perplexity, thus are more predictable. Persons in advanced stages of dementia differ in predictability of word choice and production strategy - people in early stages differ only in predictability of production strategy (Linz et al., 2018a).

7.19.3. Telephone-based Dementia Screening I: Automated Semantic Verbal Fluency Assessment:

Despite encouraging results, there are still two main issues in leveraging pervasive sensing technologies for automatic dementia screening: significant hardware costs or installation efforts and the challenge of an effective pattern recognition. Conversely, automatic speech recognition (ASR) and speech analysis have reached sufficient maturity and allow for low-tech remote telephone-based screening scenarios. Therefore, we examine the technological feasibility of automatically assessing a neuropsychological test—Semantic Verbal Fluency (SVF)–via a telephone-based solution. We investigate its suitability for inclusion into an automated dementia frontline screening and global risk assessment, based on concise telephone-sampled speech, ASR and machine learning classification. Results are encouraging showing an area under the curve (AUC) of 0.85. We observe a relatively low word error rate of 33% despite phone-quality speech samples and a mean age of 77 years of the participants. The automated classification pipeline performs equally well compared to the classifier trained on manual transcriptions of the same speech data. Our results indicate SVF as a prime candidate for inclusion into an automated telephone-screening system [50].

7.19.4. Using Acoustic Markers extracted from Free Emotional Speech:

Apathy is a frequent neuropsychiatric syndrome in people with dementia. It leads to diminished motivation for physical, cognitive and emotional activity. Apathy is highly underdiagnosed since its criteria have been only recently established and rely heavily on the subjective evaluation of human observers. We analyzed speech samples from demented people with and without apathy. Speech was provoked by asking patients two emotional questions. Acoustic features were extracted and used in a classification task. The resulting models

show performances of AUC = 0.71 and AUC = 0.63. This is a decent first step into the direction of automatic detection of apathy from speech. Usefulness of stimuli to elicit free speech is found to depend on patients' gender [46].

7.19.5. Using Automatic Speech Analysis:

Apathy is present in several psychiatric and neurological conditions and found to have a severe negative effect on patients' life. In older people, it can be a predictor of increased dementia risk. Current assessment methods seem insufficiently objective and sensitive, thus new diagnostic tools and broad-scale screening technologies are needed. This study is the first of its kind aiming to investigate whether automatic speech analysis could be used for characterization and detection of apathy.

Methods: A group of apathetic and non-apathetic patients (n = 60) was recorded while performing two short narrative speech tasks. Paralinguistic markers relating to prosodic, formant, source and temporal qualities of speech were automatically extracted, examined between the groups and compared to baseline assessments. Machine learning experiments were carried out to validate the diagnosis power of extracted markers.

Results: Correlations between apathy sub-scales and features revealed a relation between temporal aspects of speech and the subdomains of reduction in interest and initiative, as well as between prosody features and the affective domain. Group differences were found to vary for males and females, depending on the task. Differences in temporal aspects of speech were found to be the most consistent difference between apathetic and non-apathetic patients. Machine learning models trained on speech features achieved top performances of AUC = 0.88 for males and AUC = 0.77 for females (article under review).

An additional study in this context analyses transcripts of responses to emotional questions (positive and negative) for sentiment using a French emotion dictionary (FEEL) and for psycholinguistic properties (LIWC). Significant reductions in the number of words, the magnitude of sentiment, the overall sentiment and the range between sentiment in the positive and negative questions are found for the apathetic population. This effect is consistent between the positive and the negative stories. When training machine learning classifiers to detect apathy based on these features, the best model showed an AUC of 0.874 using only sentiment features. LIWC features mostly showed no predictive power. When ASR technology was introduced to automatically create transcripts, the performance of predictive models dropped slightly to AUC = 0.864. ASR errors were consistent over all categories of sentiment words. These results highlight the potential of computational linguistic analysis in screening for apathy (article under review).

7.20. Monitoring the Behaviors of Retail Customers

Participants: Soumik Mallick, Julien Badie, Francois Brémond.

Keywords: Ontology, Event detection, Multi-sensor data fusion, Real-time person tracking

The future shops will be connected and distributors as well as shopkeepers need to fulfill their promise to provide a personalized shopping experience to the customers, for example: advising and guiding customers in real time. It could not only enrich the productivity of the staffs but also increase the product sale. Implementing digital service and information in the store (like using beacons) is of primary importance. Sellers can keep their promise by providing the customer's contextual support tool in order to sell more product. To improve the performance of the store, this digital service can help to analyze customer displacement and the reaction to the product which can help to reduce the operational costs of the store by optimizing store process. It can also help to adjust store prices, merchandising and commercial operation. Thus connected digital store is a major level for new consumer services and an efficient way to manage the store.

We use multiple video cameras to detect customer in real-time inside the store. Furthermore, data are collected from different sensors like mobile phone, video camera, GPS location or Beacon. It helps to provide us with the trajectory information of the customer. A trajectory is composed of a set of points. The trajectory points are collected with the help of sensor API. Then, the calculation of distance of points in subsequent frames is performed. Every point has a minimum distance to a certain threshold of time. If there is a difference between a distance on a certain period of time that will be considered as a moving subject. For example, if we have

2 tracklets from different sensors (and generally with a different frequency of points), we cut both tracklets just to keep the intersection (in terms of time) and then apply Dynamic Time Warping (DTW) on this section. When we have the results for all tracklet pairs, we order them by distance and we decide to authorize to merge the data from the different sensors or not, with help of fusion algorithms to pass the information from the sensors to the ontology. After that, only one trajectory is sent to the ontology. Then we create a SPARQL request to extract trajectory-based events and execute it.

In this storeConnect project, we are investigating to improve the event recognition model. It will help to identify customer activity in the different zone inside the store as well as moving and stopping positions of the customer. Furthermore, inside the ontology, we want to add different attributes such as emotion, gender etc.

7.21. 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 artefacts), that follow the predefined model of an activity. We propose to describe activities as a series of actions, triggered and driven by environmental events.

Due to the large range of application domains (surveillance, safety, health care ...), we propose a generic approach to design activity recognition systems that interact continuously with their environment and react to its stimuli at run-time. Such recognition systems must satisfy stringent requirements: dependability, real time, cost effectiveness, security and safety, correctness, completeness ... To enforce most of these properties, our approach is to base the configuration of the system as well as its execution on formal techniques. We chose the *Synchronous Approach* which provides formal bases to perform static analysis, verification and validation, but also direct implementation.

Based on the synchronous approach, we designed a new user-oriented activity description language (named ADeL) to express activities and to automatically generate recognition automata. This language relies on two formal semantics, a behavioral and an equational one [48]. We also developed a component, called Synchronizer, to transform asynchronous sensor events into synchronous “instants”, necessary for the synchronous approach. This year, we mainly worked on the ADeL compiler to generate synchronous automata, on the graphical tool of this language and on the Synchronizer component.

7.21.1. ADeL Compilation:

To compile an ADeL program, we first transform it into an equation system which represents its synchronous automaton. Then we directly implement this equation system, transforming it into a Boolean equation system. This equation system provides an effective implementation of the initial ADeL program for our runtime recognition engine. The internal representation as Boolean equation systems also makes it possible to verify and validate ADeL programs, by generating a format suitable for a dedicated model checker such as the off-the-shelf NuSMV model-checker.

7.21.2. Synchronizer:

The role of the Synchronizer is to filter physical asynchronous events, to decide which ones may be considered as “simultaneous” and to aggregate the latter into logical instants. The sequence of these instants constitutes the logical time of our recognition systems. The runtime recognition engine interacts with the synchronizer and uses these instants to run the automata corresponding to the activities currently recognized. In general, no exact decision algorithm exists but several empirical strategies and heuristics may be used e.g., for determining instant boundaries. This year we completed the specification and implementation of a first version of the Synchronizer. It is parametrized by heuristics to manage events and data coming from various sensors, to define instant boundaries, and to cope with possible high level interruptions (preemptions).

Moreover, to facilitate the job of the synchronizer (to build the instants) and of the runtime engine (to wake up only the relevant automata), each automaton provides information about the awaited events at each state, i.e the events which may trigger transitions to a next state. The ADeL compiler has in charge to generate this information. In a first attempt, we computed statically all the awaited events in all states of an automaton. However, this approach implied to build the entire explicit automaton from an equation system, which was not realistic. Thus, this year we added specific equations to the equation systems of the operational semantics to compute the awaited events of each operator of the language. The information about next awaited events is now computed at runtine, when a state of the automaton is reached.

7.22. Probabilistic Activity Description Language

Participants: Elisabetta de Maria, Sabine Moisan, Jean-Paul Rigault.

Since the arrival of E. De Maria in the STARS team in September 2018, we work on the conception of a probabilistic framework for human behavior representation. The goal is to propose (i) a textual language for the description of activities which takes uncertainty into account; (ii) a formal probabilistic model to represent behaviors. Such a model will be tested and validated using experimental data coming from Alzheimer's patients. We will use temporal data resulting from different sensors and corresponding to patients playing with serious games. This will be the topic of T. L'Yvonnet's PhD starting in December. E. De Maria's main researches concern the investigation of the dynamic behavior of biological neuronal networks, using Leaky Integrate and Fire (LIF) neuronal networks, whose temporal dimension is crucial (the state of each neuron is computed taking into account not only present inputs but also past ones). This year, we used the PRISM language to model LIF neuronal networks as probabilistic reactive systems and we proposed an algorithm which aims at reducing the number of neurons and synaptical connections of these networks [42].

THOTH Project-Team

7. New Results

7.1. Visual Recognition in Images and Videos

7.1.1. Actor and Observer: Joint Modeling of First and Third-Person Videos

Participants: Gunnar Sigurdsson [CMU], Abhinav Gupta [CMU], Cordelia Schmid, Ali Farhadi [AI2, Univ. Washington], Karteek Alahari.

Several theories in cognitive neuroscience suggest that when people interact with the world, or simulate interactions, they do so from a first-person egocentric perspective, and seamlessly transfer knowledge between third-person (observer) and first-person (actor). Despite this, learning such models for human action recognition has not been achievable due to the lack of data. Our work in [33] takes a step in this direction, with the introduction of Charades-Ego, a large-scale dataset of paired first-person and third-person videos, involving 112 people, with 4000 paired videos. This enables learning the link between the two, actor and observer perspectives. Thereby, we address one of the biggest bottlenecks facing egocentric vision research, providing a link from first-person to the abundant third-person data on the web. We use this data to learn a joint representation of first and third-person videos, with only weak supervision, and show its effectiveness for transferring knowledge from the third-person to the first-person domain.

7.1.2. Learning to Segment Moving Objects

Participants: Pavel Tokmakov, Cordelia Schmid, Karteek Alahari.

We study the problem of segmenting moving objects in unconstrained videos [14]. Given a video, the task is to segment all the objects that exhibit independent motion in at least one frame. We formulate this as a learning problem and design our framework with three cues: (i) independent object motion between a pair of frames, which complements object recognition, (ii) object appearance, which helps to correct errors in motion estimation, and (iii) temporal consistency, which imposes additional constraints on the segmentation. The framework is a two-stream neural network with an explicit memory module. The two streams encode appearance and motion cues in a video sequence respectively , while the memory module captures the evolution of objects over time, exploiting the temporal consistency. The motion stream is a convolutional neural network trained on synthetic videos to segment independently moving objects in the optical flow field. The module to build a visual memory in video, i.e., a joint representation of all the video frames, is realized with a convolutional recurrent unit learned from a small number of training video sequences. For every pixel in a frame of a test video, our approach assigns an object or background label based on the learned spatio-temporal features as well as the 'visual memory' specific to the video. We evaluate our method extensively on three benchmarks, DAVIS, Freiburg-Berkeley motion seg-mentation dataset and SegTrack. In addition, we provide an extensive ablation study to investigate both the choice of the training data and the influence of each component in the proposed framework. An overview of our model is shown in Figure 1 .

7.1.3. Unsupervised Learning of Artistic Styles with Archetypal Style Analysis

Participants: Daan Wynen, Cordelia Schmid, Julien Mairal.

In [36], we introduce an unsupervised learning approach to automatically discover, summarize, and manipulate artistic styles from large collections of paintings. Our method (summarized in Figure 2) is based on archetypal analysis, which is an unsupervised learning technique akin to sparse coding with a geometric interpretation. When applied to neural style representations from a collection of artworks, it learns a dictionary of archetypal styles, which can be easily visualized. After training the model, the style of a new image, which is characterized by local statistics of deep visual features, is approximated by a sparse convex combination of archetypes. This enables us to interpret which archetypal styles are present in the input image, and in which proportion. Finally, our approach allows us to manipulate the coefficients of the latent archetypal decomposition, and achieve various special effects such as style enhancement, transfer, and interpolation between multiple archetypes.

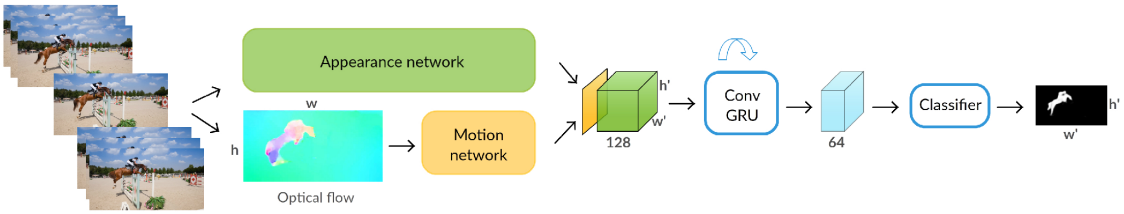


Figure 1. Overview of our segmentation approach [14]. Each video frame is processed by the appearance (green) and the motion (yellow) networks to produce an intermediate two-stream representation. The ConvGRU module combines this with the learned visual memory to compute the final segmentation result. The width (w') and height (h') of the feature map and the output are $w/8$ and $h/8$ respectively.

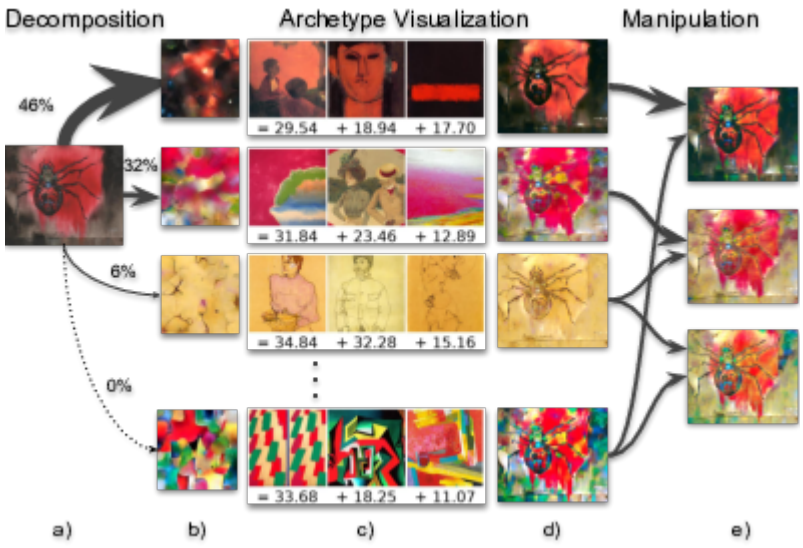


Figure 2. Using deep archetypal style analysis, we can represent the style of an artwork (a) as a convex combination of archetypes. The archetypes can be visualized as synthesized textures (b), as a convex combination of artworks (c) or, when analyzing a specific artwork, as stylized versions of that artwork itself (d). Free recombination of the archetypal styles then allows for novel stylizations of the input.

7.1.4. Learning from Web Videos for Event Classification

Participants: Nicolas Chesneau, Karteek Alahari, Cordelia Schmid.

Traditional approaches for classifying event videos rely on a manually curated training dataset. While this paradigm has achieved excellent results on benchmarks such as TrecVid multimedia event detection (MED) challenge datasets, it is restricted by the effort involved in careful annotation. Recent approaches have attempted to address the need for annotation by automatically extracting images from the web, or generating queries to retrieve videos. In the former case, they fail to exploit additional cues provided by video data, while in the latter, they still require some manual annotation to generate relevant queries. We take an alternate approach in [4], leveraging the synergy between visual video data and the associated textual metadata, to learn event classifiers without manually annotating any videos. Specifically, we first collect a video dataset with queries constructed automatically from textual description of events, prune irrelevant videos with text and video data, and then learn the corresponding event classifiers. We evaluate this approach in the challenging setting where no manually annotated training set is available, i.e., EK0 in the TrecVid challenge, and show state-of-the-art results on MED 2011 and 2013 datasets.

7.1.5. How good is my GAN?

Participants: Konstantin Shmelkov, Cordelia Schmid, Karteek Alahari.

Generative adversarial networks (GANs) are one of the most popular methods for generating images today. While impressive results have been validated by visual inspection, a number of quantitative criteria have emerged only recently. We argue here that the existing ones are insufficient and need to be in adequation with the task at hand. In [32] introduce two measures based on image classification—GAN-train and GAN-test (illustrated in Figure 3), which approximate the recall (diversity) and precision (quality of the image) of GANs respectively. We evaluate a number of recent GAN approaches based on these two measures and demonstrate a clear difference in performance. Furthermore, we observe that the increasing difficulty of the dataset, from CIFAR10 over CIFAR100 to ImageNet, shows an inverse correlation with the quality of the GANs, as clearly evident from our measures.

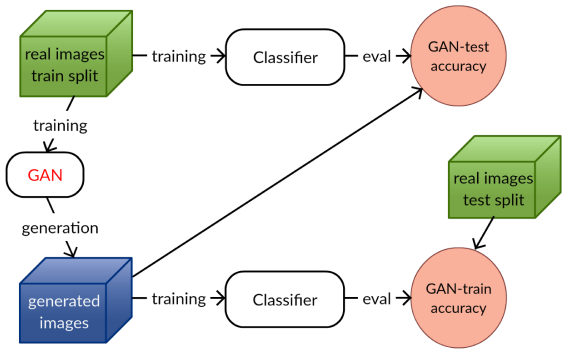


Figure 3. Illustration of GAN-train and GAN-test. GAN-train learns a classifier on GAN generated images and measures the performance on real test images. This evaluates the diversity and realism of GAN images. GAN-test learns a classifier on real images and evaluates it on GAN images. This measures how realistic GAN images are.

7.1.6. Modeling Visual Context is Key to Augmenting Object Detection Datasets

Participants: Nikita Dvornik, Julien Mairal, Cordelia Schmid.

Performing data augmentation for learning deep neural networks is well known to be important for training visual recognition systems. By artificially increasing the number of training examples, it helps reducing overfitting and improves generalization. For object detection, classical approaches for data augmentation consist of generating images obtained by basic geometrical transformations and color changes of original training images. In [23], we go one step further and leverage segmentation annotations to increase the number of object instances present on training data. For this approach to be successful, we show that modeling appropriately the visual context surrounding objects is crucial to place them in the right environment. Otherwise, we show that the previous strategy actually hurts. Clear difference between the two approaches can is presented in Figure 4 . With our context model, we achieve significant mean average precision improvements when few labeled examples are available on the VOC'12 benchmark.

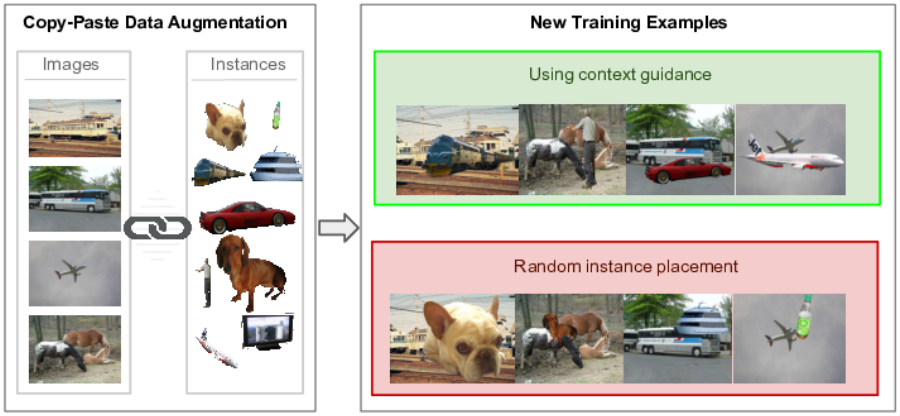


Figure 4. ***Examples of data-augmented training examples produced by our approach.*** *Images and objects are taken from the VOC'12 dataset that contains segmentation annotations. We compare the output obtained by pasting the objects with our context model vs. those obtained with random placements. Even though the results are not perfectly photorealistic and display blending artefacts, the visual context surrounding objects is more often correct with the explicit context model.*

7.1.7. On the Importance of Visual Context for Data Augmentation in Scene Understanding

Participants: Nikita Dvornik, Julien Mairal, Cordelia Schmid.

Performing data augmentation for learning deep neural networks is known to be important for training visual recognition systems. By artificially increasing the number of training examples, it helps reducing overfitting and improves generalization. While simple image transformations such as changing color intensity or adding random noise can already improve predictive performance in most vision tasks, larger gains can be obtained by leveraging task-specific prior knowledge. In [42], we consider object detection and semantic segmentation and augment the training images by blending objects in existing scenes, using instance segmentation annotations. We observe that randomly pasting objects on images hurts the performance, unless the object is placed in the right context. To resolve this issue, we propose an explicit context model by using a convolutional neural network, which predicts whether an image region is suitable for placing a given object or not. In our experiments, we show that by using copy-paste data augmentation with context guidance we are able to improve detection and segmentation on the PASCAL VOC12 and COCO datasets, with significant gains when few labeled examples are available. The way to augment for different tasks and annotations is presented

in Figure 5 . We also show that the method is not limited to datasets that come with expensive pixel-wise instance annotations and can be used when only bounding box annotations are available, by employing weakly-supervised learning for instance masks approximation.

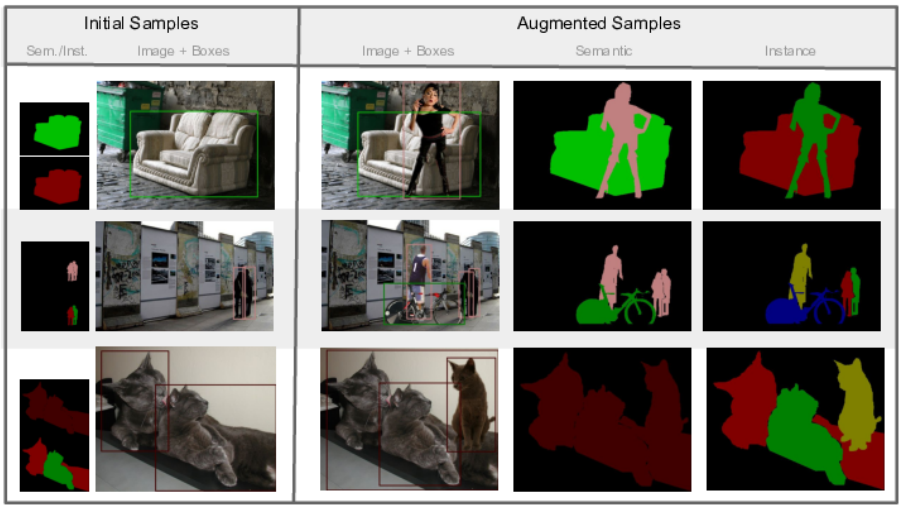


Figure 5. ***Data augmentation for different types of annotations.*** *The first column contains samples from the training dataset with corresponding semantic/instance segmentation and bounding box annotations. Columns 2-4 present the result of applying context-driven augmentation to the initial sample with corresponding annotations.*

7.1.8. Predicting future instance segmentation by forecasting convolutional features

Participants: Pauline Luc, Camille Couprie [Facebook AI Research], Yann Lecun [Facebook AI Research], Jakob Verbeek.

Anticipating future events is an important prerequisite towards intelligent behavior. Video forecasting has been studied as a proxy task towards this goal. Recent work has shown that to predict semantic segmentation of future frames, forecasting at the semantic level is more effective than forecasting RGB frames and then segmenting these. In [28], we consider the more challenging problem of future instance segmentation, which additionally segments out individual objects. To deal with a varying number of output labels per image, we develop a predictive model in the space of fixed-sized convolutional features of the Mask R-CNN instance segmentation model. We apply the "detection head" of Mask R-CNN on the predicted features to produce the instance segmentation of future frames. Experiments show that this approach significantly improves over strong baselines based on optical flow and repurposed instance segmentation architectures. We show an overview of the proposed method in Figure 6 .

7.1.9. Joint Future Semantic and Instance Segmentation Prediction

Participants: Camille Couprie [Facebook AI Research], Pauline Luc, Jakob Verbeek.

The ability to predict what will happen next from observing the past is a key component of intelligence. Methods that forecast future frames were recently introduced towards better machine intelligence. However, predicting directly in the image color space seems an overly complex task, and predicting higher level representations using semantic or instance segmentation approaches were shown to be more accurate. In [20], we introduce a novel prediction approach that encodes instance and semantic segmentation information in a single representation based on distance maps. Our graph-based modeling of the instance segmentation

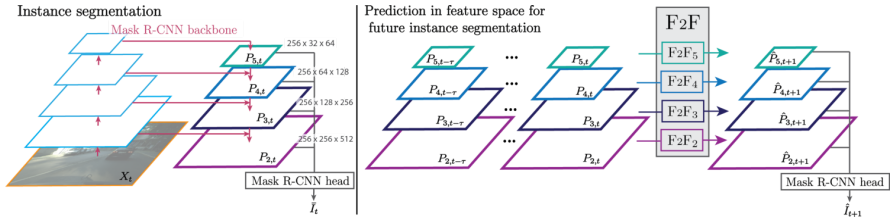


Figure 6. For future instance segmentation, we extract a pyramid of features from frames $t-\tau$ to t, and use them to predict the pyramid features for frame $t+1$. We learn separate feature-to-feature prediction models for each level of the pyramid. The predicted features are then given as input to a downstream network to produce future instance segmentation.

prediction problem allows us to obtain temporal tracks of the objects as an optimal solution to a watershed algorithm. Our experimental results on the Cityscapes dataset present state-of-the-art semantic segmentation predictions, and instance segmentation results outperforming a strong baseline based on optical flow. We show an overview of the proposed method in Figure 7 .

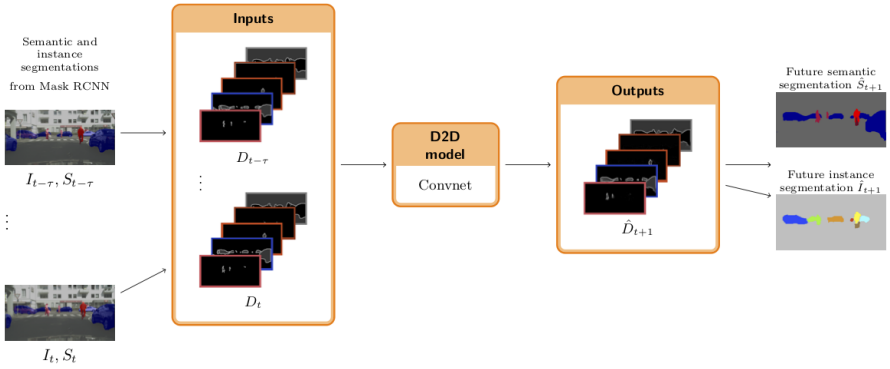


Figure 7. Our representation enables both future semantic and instance segmentation prediction. It relies on distance maps from the different objects contours: For each channel of an input segmentation, corresponding to a specific class, the segmentation is decomposed into zeros for background, ones for objects and high values for contours. Then a convnet is trained to predict the future representation. Taking its argmax lets us recover the future semantic segmentation, and computing a watershed from it leads to the future instance segmentation.

7.1.10. Depth-based Hand Pose Estimation: Methods, Data, and Challenges

Participants: James S. Supancic [UC Irvine], Grégory Rogez, Yi Yang [Baidu Research], Jamie Shotton [Microsoft Research], Deva Ramanan [Carnegie Mellon University].

Hand pose estimation has matured rapidly in recent years. The introduction of commodity depth sensors and a multitude of practical applications have spurred new advances. In [13], we provide an extensive analysis of the state-of-the-art, focusing on hand pose estimation from a single depth frame. We summarize important conclusions here: (1) Pose estimation appears roughly solved for scenes with isolated hands. However,

methods still struggle to analyze cluttered scenes where hands may be interacting with nearby objects and surfaces. To spur further progress we introduce a challenging new dataset with diverse, cluttered scenes. (2) Many methods evaluate themselves with disparate criteria , making comparisons difficult. We define a consistent evaluation criteria, rigorously motivated by human experiments. (3) We introduce a simple nearest-neighbor baseline that outperforms most existing systems (see results in Fig. 8). This implies that most systems do not generalize beyond their training sets. This also reinforces the under-appreciated point that training data is as important as the model itself. We conclude with directions for future progress.

Figure 8. We evaluate a broad collection of hand pose estimation algorithms on different training and testsets under consistent criteria. Test sets which contained limited variety, in pose and range, or which lacked complex backgrounds were notably easier. To aid our analysis, we introduce a simple 3D exemplar (nearest-neighbor) baseline that both detects and estimates pose suprisingly well, outperforming most existing systems. We show the best-matching detection window in (middle) and the best-matching exemplar in (bottom). We use our baseline to rank dataset difficulty, compare algorithms, and show the importance of training set design.

7.1.11. Image-based Synthesis for Deep 3D Human Pose Estimation

Participants: Grégory Rogez, Cordelia Schmid.

In [11], we address the problem of 3D human pose estimation in the wild. A significant challenge is the lack of training data, i.e., 2D images of humans annotated with 3D poses. Such data is necessary to train state-of-the-art CNN architectures. Here, we propose a solution to generate a large set of photorealistic synthetic images of humans with 3D pose annotations. We introduce an image-based synthesis engine that artificially augments a dataset of real images with 2D human pose annotations using 3D Motion Capture (MoCap) data. Given a candidate 3D pose our algorithm selects for each joint an image whose 2D pose locally matches the projected 3D pose. The selected images are then combined to generate a new synthetic image by stitching local image patches in a kinematically constrained manner. See examples in Figure 9 . The resulting images are used to train an end-to-end CNN for full-body 3D pose estimation. We cluster the training data into a large number of pose classes and tackle pose estimation as a K-way classification problem. Such an approach is viable only with large training sets such as ours. Our method outperforms the state of the art in terms of 3D pose estimation in controlled environments (Human3.6M) and shows promising results for in-the-wild images (LSP). This demonstrates that CNNs trained on artificial images generalize well to real images. Compared to data generated from more classical rendering engines, our synthetic images do not require any domain adaptation or fine-tuning stage.

7.1.12. LCR-Net++: Multi-person 2D and 3D Pose Detection in Natural Images

Participants: Grégory Rogez, Philippe Weinzaepfel [Naver Labs Europe], Cordelia Schmid.

Figure 9. Given a candidate 3D pose, our algorithm selects for each joint an image whose annotated 2D pose locally matches the projected 3D pose. The selected images are then combined to generate a new synthetic image by stitching local image patches in a kinematically constrained manner. We show 6 examples corresponding to the same 3D pose observed from 6 different camera viewpoints.

In [12], we propose an end-to-end architecture for joint 2D and 3D human pose estimation in natural images. Key to our approach is the generation and scoring of a number of pose proposals per image, which allows us to predict 2D and 3D pose of multiple people simultaneously. See example in Figure 10 . Hence, our approach does not require an approximate localization of the humans for initialization. Our architecture, named LCR-Net, contains 3 main components: 1) the pose proposal generator that suggests potential poses at different locations in the image; 2) a classifier that scores the different pose proposals ; and 3) a regressor that refines pose proposals both in 2D and 3D. All three stages share the convolutional feature layers and are trained jointly. The final pose estimation is obtained by integrating over neighboring pose hypotheses , which is shown to improve over a standard non maximum suppression algorithm. Our approach significantly outperforms the state of the art in 3D pose estimation on Human3.6M, a controlled environment. Moreover, it shows promising results on real images for both single and multi-person subsets of the MPII 2D pose benchmark and demonstrates satisfying 3D pose results even for multi-person images.

7.1.13. Link and code: Fast indexing with graphs and compact regression codes

Participants: Matthijs Douze [Facebook AI Research], Alexandre Sablayrolles, Hervé Jégou [Facebook AI Research].

Similarity search approaches based on graph walks have recently attained outstanding speed-accuracy trade-offs, taking aside the memory requirements. In [21], we revisit these approaches by considering, additionally, the memory constraint required to index billions of images on a single server. This leads us to propose a method based both on graph traversal and compact representations. We encode the indexed vectors using quantization and exploit the graph structure to refine the similarity estimation, see Figure 11 . In essence, our method takes the best of these two worlds: the search strategy is based on nested graphs, thereby providing high precision with a relatively small set of comparisons. At the same time it offers a significant memory compression. As a result, our approach outperforms the state of the art on operating points considering 64–128 bytes per vector, as demonstrated by our results on two billion-scale public benchmarks.

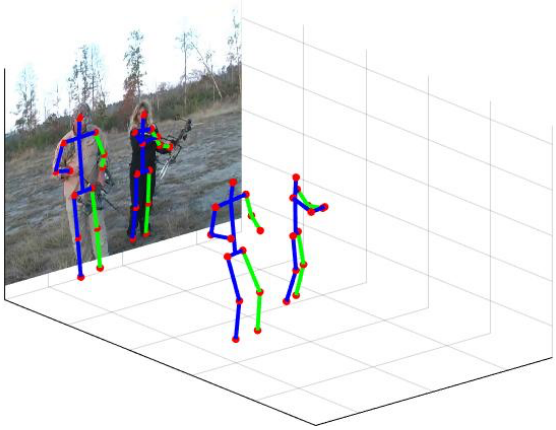


Figure 10. Examples of joint 2D-3D pose detections in a natural image. Even in case of occlusion or truncation, we estimate the joint locations by reasoning in term of full-body 2D-3D poses.

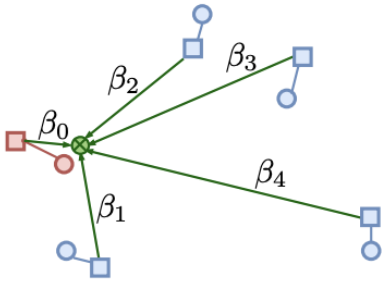


Figure 11. Illustration of our approach: we adopt a graph traversal strategy that maintains a connectivity between all database points. We further improve the estimate by regressing each database vector from its encoded neighbors.

7.1.14. Sparse weakly supervised models for object localization in road environment

Participants: Valentina Zadrija [Univ. Zagreb], Josip Krapac [Univ. Zagreb], Sinisa Segvic [Univ. Zagreb], Jakob Verbeek.

In [16] we propose a novel weakly supervised object localization method based on Fisher-embedding of low-level features (CNN, SIFT), and model sparsity at the component level. Fisher-embedding provides an interesting alternative to raw low-level features, since it allows fast and accurate scoring of image subwindows with a model trained on entire images. Model sparsity reduces overfitting and enables fast evaluation. We also propose two new techniques for improving performance when our method is combined with nonlinear normalizations of the aggregated Fisher representation of the image. These techniques are i) intra-component metric normalization and ii) first-order approximation to the score of a normalized image representation. We evaluate our weakly supervised localization method on real traffic scenes acquired from driver's perspective. The method dramatically improves the localization AP over the dense non-normalized Fisher vector baseline (16 percentage points for zebra crossings, 21 percentage points for traffic signs) and leads to a huge gain in execution speed (91× for zebra crossings, 74× for traffic signs). See Figure 12 for several example outputs.

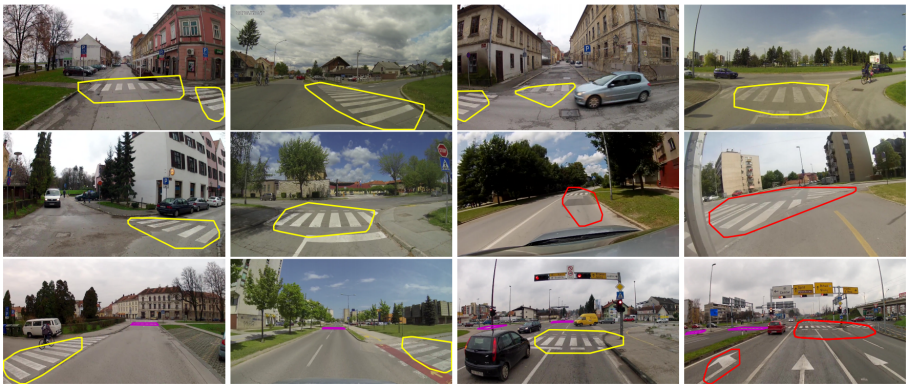


Figure 12. Localization results on test images: correct localization polygons (yellow), false postive responses (red), and ground-truth polygons for false negatives (magenta).

7.1.15. Scene Coordinate Regression with Angle-Based Reprojection Loss for Camera Relocalization

Participants: Xiaotian Li [Aalto Univ.], Juha Ylioinas [Aalto Univ.], Jakob Verbeek, Juho Kannala [Univ. Oulu].

Image-based camera relocalization is an important problem in computer vision and robotics. Recent works utilize convolutional neural networks (CNNs) to regress for pixels in a query image their corresponding 3D world coordinates in the scene. The final pose is then solved via a RANSAC-based optimization scheme using the predicted coordinates, see Figure 13 . Usually, the CNN is trained with ground truth scene coordinates, but it has also been shown that the network can discover 3D scene geometry automatically by minimizing single-view reprojection loss. However, due to the deficiencies of reprojection loss, the network needs to be carefully initialized. In [27], we present a new angle-based reprojection loss which resolves the issues of the original reprojection loss. With this new loss function, the network can be trained without careful initialization, and the system achieves more accurate results. The new loss also enables us to utilize available multi-view constraints, which further improve performance.

7.1.16. FeaStNet: Feature-Steered Graph Convolutions for 3D Shape Analysis

Participants: Nitika Verma, Edmond Boyer [Inria, MORPHEO], Jakob Verbeek.

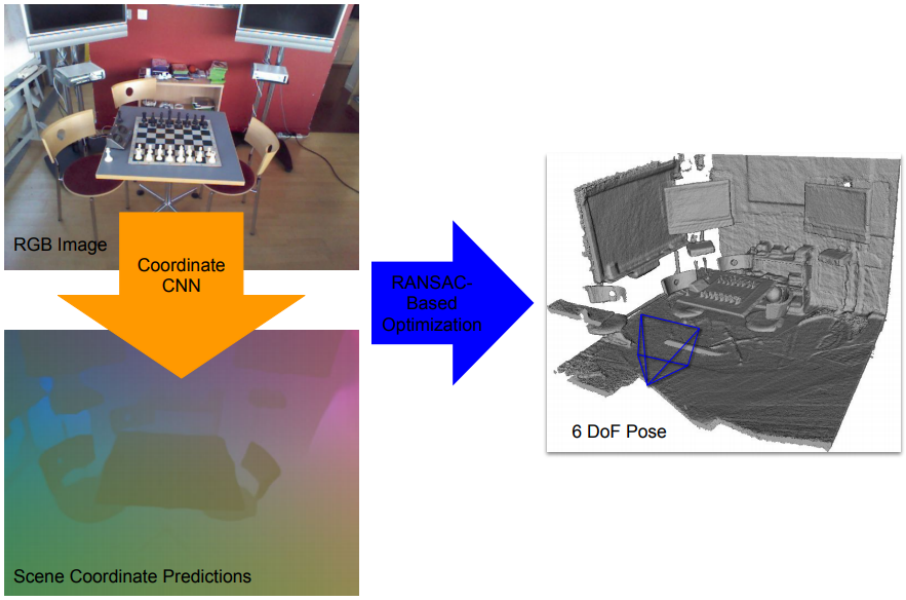


Figure 13. Localization pipeline. In this two-stage pipeline, a coordinate CNN first produces scene coordinate predictions from an RGB image, and then the predicted correspondences are fed into a RANSAC-based solver to determine the camera pose.

Convolutional neural networks (CNNs) have massively impacted visual recognition in 2D images, and are now ubiquitous in state-of-the-art approaches. While CNNs naturally extend to other domains, such as audio and video, where data is also organized in rectangular grids, they do not easily generalize to other types of data such as 3D shape meshes, social network graphs or molecular graphs. In our recent paper [35], we propose a novel graph-convolutional network architecture to handle such data. The architecture builds on a generic formulation that relaxes the 1-to-1 correspondence between filter weights and data elements around the center of the convolution, see Figure 14 for an illustration. The main novelty of our architecture is that the shape of the filter is a function of the features in the previous network layer, which is learned as an integral part of the neural network. Experimental evaluations on digit recognition and 3D shape correspondence yield state-of-the-art results, significantly improving over previous work for shape correspondence.

7.2. Statistical Machine Learning

7.2.1. Modulated Policy Hierarchies

Participants: Alexander Pashevich, Danijar Hafner [Google Brain], James Davidson [Vernalis (R&D) Ltd.], Rahul Sukthankar [Google], Cordelia Schmid.

Solving tasks with sparse rewards is a main challenge in reinforcement learning. While hierarchical controllers are an intuitive approach to this problem, current methods often require manual reward shaping, alternating training phases, or manually defined sub tasks. In [45], we introduce modulated policy hierarchies (MPH), that can learn end-to-end to solve tasks from sparse rewards. To achieve this, we study different modulation signals and exploration for hierarchical controllers. Specifically, we find that communicating via bit-vectors is more efficient than selecting one out of multiple skills, as it enables mixing between them (see Figure 15). To facilitate exploration, MPH uses its different time scales for temporally extended intrinsic motivation at each level of the hierarchy. We evaluate MPH on the robotics tasks of pushing and sparse block stacking, where it outperforms recent baselines.

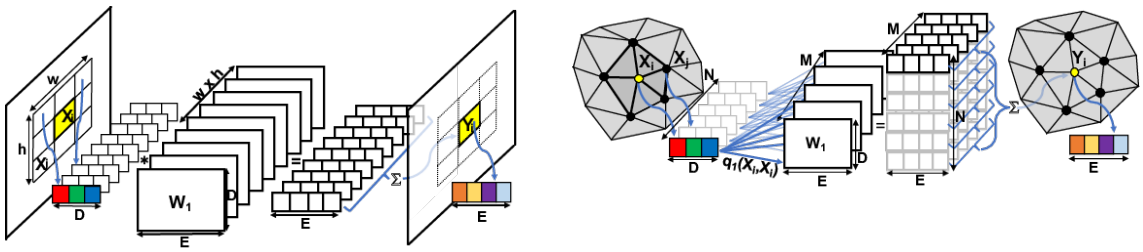


Figure 14. Left: Illustration of a standard CNN, representing the parameters as a set of $M = w \times h$ weight matrices, each of size $D \times E$. Each weight matrix is associated with a single relative position in the input patch. Right: Our graph convolutional network, where each node in the input patch is associated in a soft manner to each of the M weight matrices based on its features using the weight $q_m(\mathbf{x}_i, \mathbf{x}_j)$.

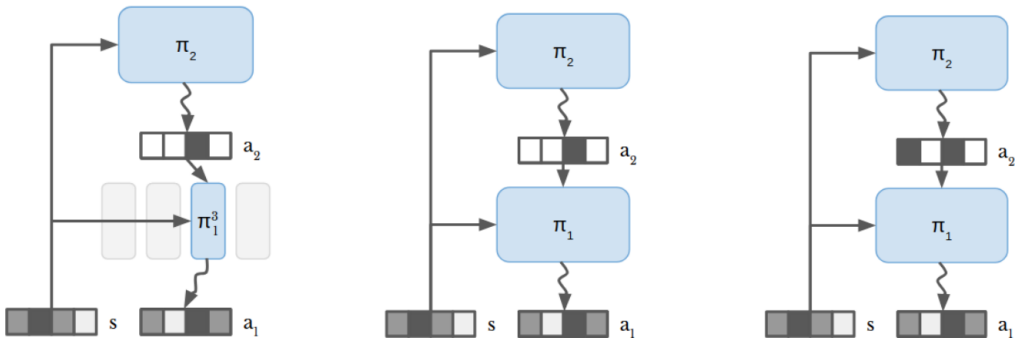


Figure 15. Overview of hierarchical policies. (a) The options agent selects between separate skill networks using a categorical master policy. (b) The one-hot agent combines the skills into a single network and is modulated by a 1-hot signal. (c) Our modulated policy hierarchy sends a binary vector, allowing for richer communication and mixing of skills.

7.2.2. 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 [38], 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.

7.2.3. A Contextual Bandit Bake-off

Participants: Alberto Bietti, Alekh Agarwal [Microsoft Research], John Langford [Microsoft Research].

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 [37], 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.

7.2.4. Learning Disentangled Representations with Reference-Based Variational Autoencoders

Participants: Adria Ruiz, Oriol Martinez [Universitat Pompeu Fabra, Barcelona], Xavier Binefa [Universitat Pompeu Fabra, Barcelona], 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 [47] 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. 16 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.

7.2.5. On Regularization and Robustness of Deep Neural Networks

Participants: Alberto Bietti, Grégoire Mialon, Julien Mairal.

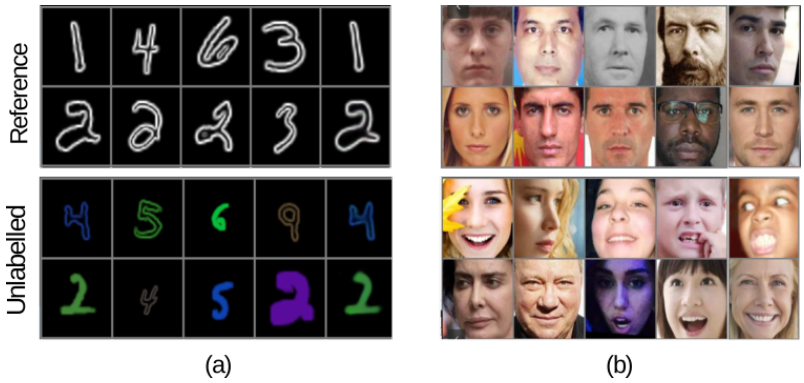


Figure 16. Illustration of different reference-based disentangling problems. (a) Disentangling style from digits. The reference distribution is composed by numbers with a fixed style (b) Disentangling factors of variations related with facial expressions. Reference images correspond to neutral faces. Note that pairing information between unlabelled and reference images is not available during training.

For many supervised learning tasks, deep neural networks are known to work well when large amounts of annotated data are available. Yet, Despite their success, deep neural networks suffer from several drawbacks: they lack robustness to small changes of input data known as "adversarial examples" and training them with small amounts of annotated data is challenging. In [39], we study the connection between regularization and robustness of deep neural networks by viewing them as elements of a reproducing kernel Hilbert space (RKHS) of functions and by regularizing them using the RKHS norm. Even though this norm cannot be computed, we consider various approximations based on upper and lower bounds. These approximations lead to new strategies for regularization, but also to existing ones such as spectral norm penalties or constraints, gradient penalties, or adversarial training. Besides, the kernel framework allows us to obtain margin-based bounds on adversarial generalization. We show that our new algorithms lead to empirical benefits for learning on small datasets and learning adversarially robust models. We also discuss implications of our regularization framework for learning implicit generative models.

7.2.6. Mixed batches and symmetric discriminators for GAN training

Participants: Thomas Lucas, Corentin Tallec [Inria, TAU], Jakob Verbeek, Yann Ollivier [Facebook AI Research].

Generative adversarial networks (GANs) are powerful generative models based on providing feedback to a generative network via a discriminator network. However, the discriminator usually assesses individual samples. This prevents the discriminator from accessing global distributional statistics of generated samples, and often leads to *mode dropping*: the generator models only part of the target distribution. In [29] we propose to feed the discriminator with *mixed batches* of true and fake samples, and train it to predict the ratio of true samples in the batch. The latter score does not depend on the order of samples in a batch. Rather than learning this invariance, we introduce a generic permutation-invariant discriminator architecture, which is illustrated in Figure 17 . This architecture is provably a universal approximator of all symmetric functions. Experimentally, our approach reduces mode collapse in GANs on two synthetic datasets, and obtains good results on the CIFAR10 and CelebA datasets, both qualitatively and quantitatively.

7.2.7. Auxiliary Guided Autoregressive Variational Autoencoders

Participants: Thomas Lucas, Jakob Verbeek.

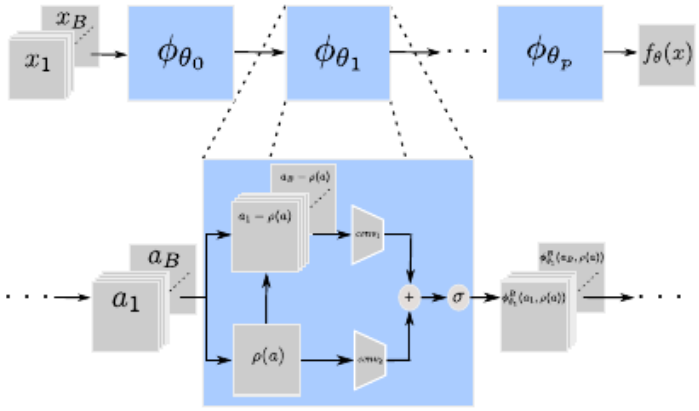


Figure 17. Graphical representation of our discriminator architecture. Each convolutional layer of an otherwise classical CNN architecture is modified to include permutation invariant batch statistics, denoted $\rho(x)$. This is repeated at every layer so that the network gradually builds up more complex statistics.

Generative modeling of high-dimensional data is a key problem in machine learning. Successful approaches include latent variable models and autoregressive models. The complementary strengths of these approaches, to model global and local image statistics respectively, suggest hybrid models combining the strengths of both. Our contribution in [30] is to train such hybrid models using an auxiliary loss function that controls which information is captured by the latent variables and what is left to the autoregressive decoder, as illustrated in Figure 18 . In contrast, prior work on such hybrid models needed to limit the capacity of the autoregressive decoder to prevent degenerate models that ignore the latent variables and only rely on autoregressive modeling. Our approach results in models with meaningful latent variable representations, and which rely on powerful autoregressive decoders to model image details. Our model generates qualitatively convincing samples, and yields state-of-the-art quantitative results.

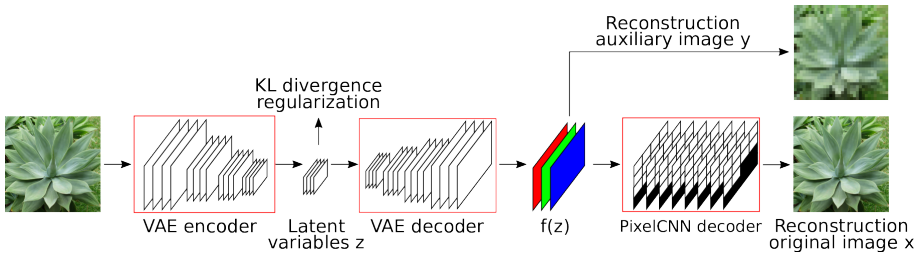


Figure 18. Schematic illustration of our auxiliary guided autoregressive variational autoencoder (AGAVE). An input image is encoded into a latent representation and decoded back into an image. This first reconstruction is guided by an auxiliary maximum likelihood loss and regularized with a Kullback-Liebler divergence. An autoregressive model is then conditionned on the auxiliary reconstruction and also trained with maximum likelihood.

7.2.8. End-to-End Incremental Learning

Participants: Francisco Castro [Univ. Malaga], Manuel Marin-Jimenez [Univ. Cordoba], Nicolas Guil [Univ. Malaga], Cordelia Schmid, Karteek Alahari.

Although deep learning approaches have stood out in recent years due to their state-of-the-art results, they continue to suffer from catastrophic forgetting, a dramatic decrease in overall performance when training with new classes added incrementally. This is due to current neural network architectures requiring the entire dataset, consisting of all the samples from the old as well as the new classes, to update the model—a requirement that becomes easily unsustainable as the number of classes grows. We address this issue with our approach [17] to learn deep neural networks incrementally, using new data and only a small exemplar set corresponding to samples from the old classes. This is based on a loss composed of a distillation measure to retain the knowledge acquired from the old classes, and a cross-entropy loss to learn the new classes. Our incremental training is achieved while keeping the entire framework end-to-end, i.e., learning the data representation and the classifier jointly, unlike recent methods with no such guarantees. We evaluate our method extensively on the CIFAR-100 and ImageNet (ILSVRC 2012) image classification datasets, and show state-of-the-art performance.

7.3. Large-scale Optimization for Machine Learning

7.3.1. Stochastic Subsampling for Factorizing Huge Matrices

Participants: Julien Mairal, Arthur Mensch [Inria, Parietal], Gael Varoquaux [Inria, Parietal], Bertrand Thirion [Inria, Parietal].

In [10], we present a matrix-factorization algorithm that scales to input matrices with both huge number of rows and columns. Learned factors may be sparse or dense and/or non-negative, which makes our algorithm suitable for dictionary learning, sparse component analysis, and non-negative matrix factorization. Our algorithm streams matrix columns while subsampling them to iteratively learn the matrix factors. At each iteration, the row dimension of a new sample is reduced by subsampling, resulting in lower time complexity compared to a simple streaming algorithm. Our method comes with convergence guarantees to reach a stationary point of the matrix-factorization problem. We demonstrate its efficiency on massive functional Magnetic Resonance Imaging data (2 TB), and on patches extracted from hyperspectral images (103 GB). For both problems, which involve different penalties on rows and columns, we obtain significant speed-ups compared to state-of-the-art algorithms. The main principle of the method is illustrated in Figure 19 .

7.3.2. An Inexact Variable Metric Proximal Point Algorithm for Generic Quasi-Newton Acceleration

Participants: Hongzhou Lin, Julien Mairal, Zaid Harchaoui [Univ. Washington].

In [43], we propose a generic approach to accelerate gradient-based optimization algorithms with quasi-Newton principles. The proposed scheme, called QuickeNing, can be applied to incremental first-order methods such as stochastic variance-reduced gradient (SVRG) or incremental surrogate optimization (MISO). It is also compatible with composite objectives, meaning that it has the ability to provide exactly sparse solutions when the objective involves a sparsity-inducing regularization. QuickeNing relies on limited-memory BFGS rules, making it appropriate for solving high-dimensional optimization problems. Besides, it enjoys a worst-case linear convergence rate for strongly convex problems. We present experimental results where QuickeNing gives significant improvements over competing methods for solving large-scale high-dimensional machine learning problems, see Figure 20 for example.

7.3.3. Catalyst Acceleration for First-order Convex Optimization: from Theory to Practice

Participants: Hongzhou Lin, Julien Mairal, Zaid Harchaoui [Univ. Washington].

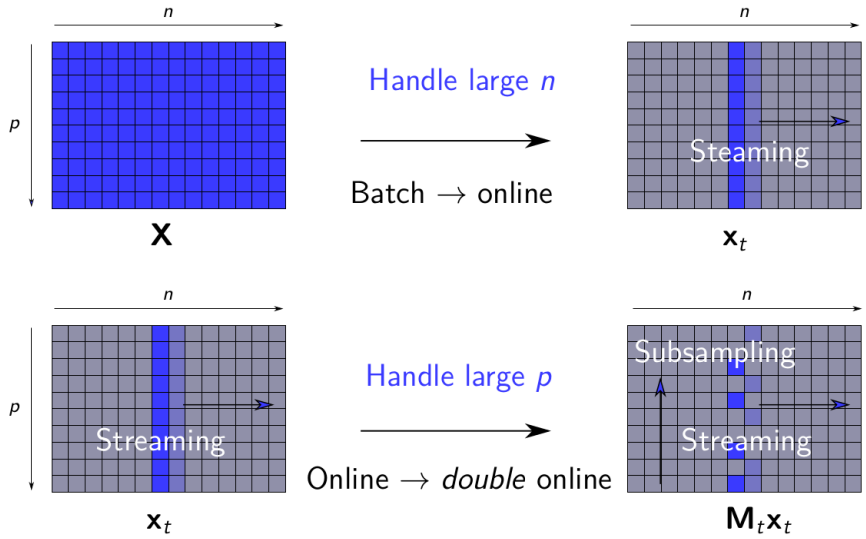


Figure 19. Illustration of the matrix factorization algorithm, which streams columns in one dimension while subsampling them.

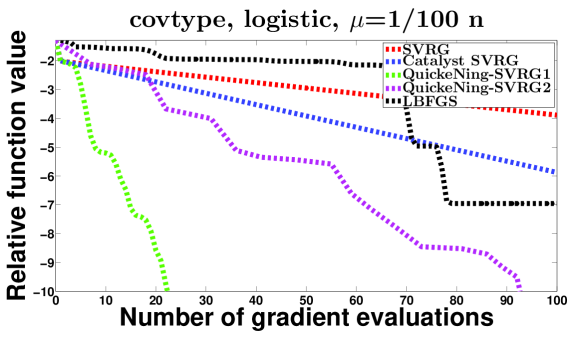


Figure 20. An illustration of the minimization of logistic regression. Significant improvement is observed by applying QuickeNing.

In [9], we introduce a generic scheme for accelerating gradient-based optimization methods in the sense of Nesterov. The approach, called Catalyst, builds upon the inexact accelerated proximal point algorithm for minimizing a convex objective function, and consists of approximately solving a sequence of well-chosen auxiliary problems, leading to faster convergence. One of the key to achieve acceleration in theory and in practice is to solve these sub-problems with appropriate accuracy by using the right stopping criterion and the right warm-start strategy. In this work, we give practical guidelines to use Catalyst and present a comprehensive theoretical analysis of its global complexity. We show that Catalyst applies to a large class of algorithms, including gradient descent, block coordinate descent, incremental algorithms such as SAG, SAGA, SDCA, SVRG, Finito/MISO, and their proximal variants. For all of these methods, we provide acceleration and explicit support for non-strongly convex objectives. We conclude with extensive experiments showing that acceleration is useful in practice, especially for ill-conditioned problems.

7.3.4. Catalyst Acceleration for Gradient-Based Non-Convex Optimization

Participants: Courtney Paquette [Univ. Washington], Hongzhou Lin, Dmitriy Drusvyatskiy [Univ. Washington], Julien Mairal, Zaid Harchaoui [Univ. Washington].

In [31], we introduce a generic scheme to solve nonconvex optimization problems using gradient-based algorithms originally designed for minimizing convex functions. When the objective is convex, the proposed approach enjoys the same properties as the Catalyst approach of Lin et al, 2015. When the objective is nonconvex, it achieves the best known convergence rate to stationary points for first-order methods. Specifically, the proposed algorithm does not require knowledge about the convexity of the objective; yet, it obtains an overall worst-case efficiency of $O(\epsilon^{-2})$ and, if the function is convex, the complexity reduces to the near-optimal rate $O(\epsilon^{-2/3})$. We conclude the paper by showing promising experimental results obtained by applying the proposed approach to SVRG and SAGA for sparse matrix factorization and for learning neural networks (see Figure 21).

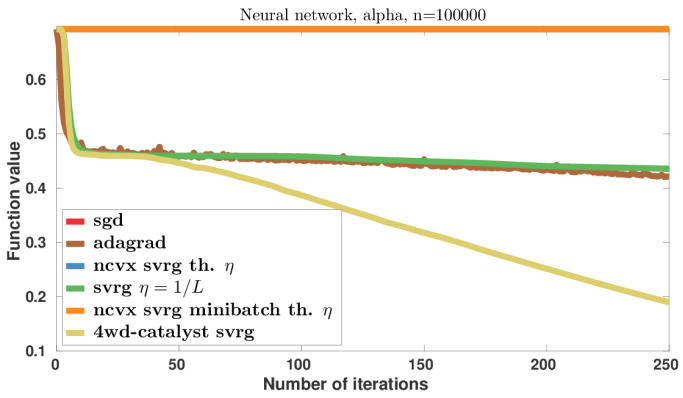


Figure 21. Comparison of different algorithms for the minimization of a two-layer neural network. Applying our method provides a clear acceleration in terms of function value.

7.4. Pluri-disciplinary Research

7.4.1. Biological Sequence Modeling with Convolutional Kernel Networks

Participants: Dexiong Chen, Laurent Jacob [CNRS, LBBE Laboratory], 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 [40], we introduce a hybrid approach between convolutional neural networks and kernel methods to model biological sequences. Our method 22 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.

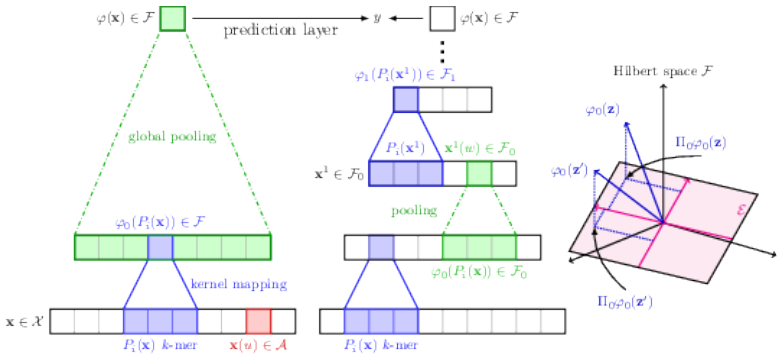


Figure 22. Construction of single-layer (left) and multilayer (middle) CKN-seq and the approximation of one layer (right). For a single-layer model, each k-mer $P_i(\mathbf{x})$ is mapped to $\varphi_0(P_i(\mathbf{x}))$ in $\mathcal{F}$ and projected to $\Pi\varphi_0(P_i(\mathbf{x}))$ parametrized by $\psi_0(P_i(\mathbf{x}))$. 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. Token-level and sequence-level loss smoothing for RNN language models

Participants: Maha Elbayad, Laurent Besacier [LIG], Jakob Verbeek.

In [25] we investigate the limitations of the maximum likelihood estimation (MLE) used when training recurrent neural network language models. First, the MLE treats all sentences that do not match the ground truth as equally poor, ignoring the structure of the output space. Second, it suffers from "exposure bias": during training tokens are predicted given ground-truth sequences, while at test time prediction is conditioned on generated output sequences. To overcome these limitations we build upon the recent reward augmented maximum likelihood approach i.e., sequence-level smoothing that encourages the model to predict sentences close to the ground truth according to a given performance metric. We extend this approach to token-level loss smoothing, and propose improvements to the sequence-level smoothing approach. Our experiments on two different tasks, image captioning (see Fig. 23) and machine translation, show that token-level and sequence-level loss smoothing are complementary, and significantly improve results.

7.4.3. Pervasive Attention: 2D Convolutional Neural Networks for Sequence-to-Sequence Prediction

Participants: Maha Elbayad, Laurent Besacier [LIG], Jakob Verbeek.

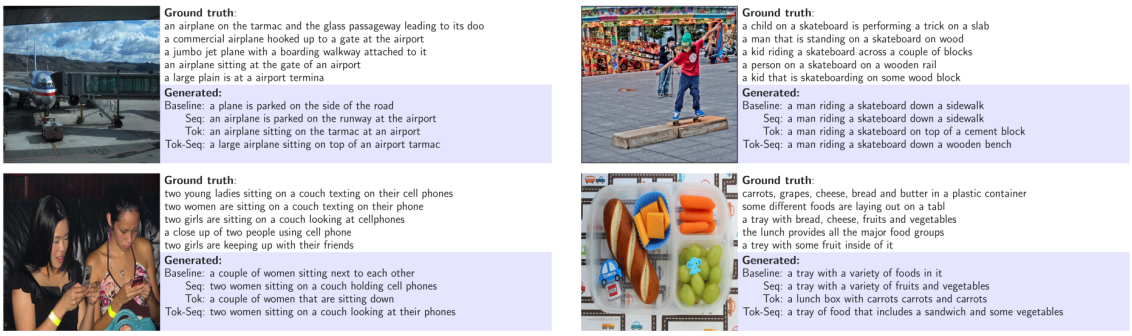


Figure 23. Examples of generated captions with the baseline MLE and our models with attention.

Current state-of-the-art machine translation systems are based on encoder-decoder architectures, that first encode the input sequence, and then generate an output sequence based on the input encoding. Both are interfaced with an attention mechanism that recombines a fixed encoding of the source tokens based on the decoder state. In [24], we propose an alternative approach which instead relies on a single 2D convolutional neural network across both sequences as illustrated in Figure 24 . Each layer of our network re-codes source tokens on the basis of the output sequence produced so far. Attention-like properties are therefore pervasive throughout the network. Our model yields excellent results, outperforming state-of-the-art encoder-decoder systems, while being conceptually simpler and having fewer parameters.

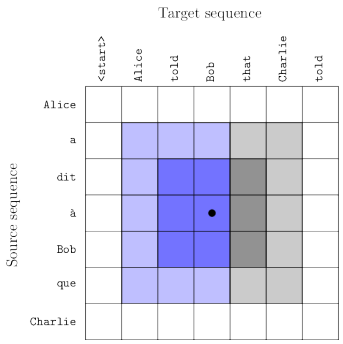


Figure 24. Convolutional layers in our model use masked 3×3 filters so that features are only computed from previous output symbols. Illustration of the receptive fields after one (dark blue) and two layers (light blue), together with the masked part of the field of view of a normal 3×3 filter (gray)

7.4.4. Probabilistic Count Matrix Factorization for Single Cell Expression Data Analysis

Participant: Ghislain Durif.

The development of high-throughput biology technologies now allows the investigation of the genome-wide diversity of transcription in single cells. This diversity has shown two faces: the expression dynamics (gene to gene variability) can be quantified more accurately, thanks to the measurement of lowly-expressed genes. Second, the cell-to-cell variability is high, with a low proportion of cells expressing the same gene at the same

time/level. Those emerging patterns appear to be very challenging from the statistical point of view, especially to represent and to provide a summarized view of single-cell expression data. PCA is one of the most powerful framework to provide a suitable representation of high dimensional datasets, by searching for latent directions catching the most variability in the data. Unfortunately, classical PCA is based on Euclidean distances and projections that work poorly in presence of over-dispersed counts that show drop-out events (zero-inflation) like single-cell expression data. In [22], we propose a probabilistic Count Matrix Factorization (pCMF) approach for single-cell expression data analysis, that relies on a sparse Gamma-Poisson factor model. This hierarchical model is inferred using a variational EM algorithm. We show how this probabilistic framework induces a geometry that is suitable for single-cell data visualization, and produces a compression of the data that is very powerful for clustering purposes. Our method is competed to other standard representation methods like t-SNE, and we illustrate its performance for the representation of zero-inflated over-dispersed count data. We also illustrate our work with results on a publicly available data set, being single-cell expression profile of neural stem cells. Our work is implemented in the pCMF R-package.

7.4.5. Extracting Universal Representations of Cognition across Brain-Imaging Studies

Participants: Arthur Mensch [Inria, Parietal], Julien Mairal, Bertrand Thirion [Inria, Parietal], Gael Varoquaux [Inria, Parietal].

We show in [44] how to extract shared brain representations that predict mental processes across many cognitive neuroimaging studies. Focused cognitive-neuroimaging experiments study precise mental processes with carefully-designed cognitive paradigms; however the cost of imaging limits their statistical power. On the other hand, large-scale databasing efforts increase considerably the sample sizes, but cannot ask precise cognitive questions. To address this tension, we develop new methods that turn the heterogeneous cognitive information held in different task-fMRI studies into common-universal-cognitive models. Our approach does not assume any prior knowledge of the commonalities shared by the studies in the corpus; those are inferred during model training. The method uses deep-learning techniques to extract representations - task-optimized networks - that form a set of basis cognitive dimensions relevant to the psychological manipulations, as illustrated in Figure 25 . In this sense, it forms a novel kind of functional atlas, optimized to capture mental state across many functional-imaging experiments. As it bridges information on the neural support of mental processes, this representation improves decoding performance for 80% of the 35 widely-different functional imaging studies that we consider. Our approach opens new ways of extracting information from brain maps, increasing statistical power even for focused cognitive neuroimaging studies, in particular for those with few subjects.

7.4.6. Loter: Inferring local ancestry for a wide range of species

Participants: Thomas Dias-Alves, Julien Mairal, Michael Blum [CNRS, TIMC Laboratory].

Admixture between populations provides opportunity to study biological adaptation and phenotypic variation. Admixture studies can rely on local ancestry inference for admixed individuals, which consists of computing at each locus the number of copies that originate from ancestral source populations, as illustrated in Figure 26 . Existing software packages for local ancestry inference are tuned to provide accurate results on human data and recent admixture events. In [5], we introduce Loter, an open-source software package that does not require any biological parameter besides haplotype data in order to make local ancestry inference available for a wide range of species. Using simulations, we compare the performance of Loter to HAPMIX, LAMP-LD, and RFMix. HAPMIX is the only software severely impacted by imperfect haplotype reconstruction. Loter is the less impacted software by increasing admixture time when considering simulated and admixed human genotypes. LAMP-LD and RFMIX are the most accurate method when admixture took place 20 generations ago or less; Loter accuracy is comparable or better than RFMix accuracy when admixture took place of 50 or more generations; and its accuracy is the largest when admixture is more ancient than 150 generations. For simulations of admixed Populus genotypes, Loter and LAMP-LD are robust to increasing admixture times by contrast to RFMix. When comparing length of reconstructed and true ancestry tracts, Loter and LAMP-LD provide results whose accuracy is again more robust than RFMix to increasing admixture times. We apply Loter to admixed Populus individuals and lengths of ancestry tracts indicate that admixture took place around

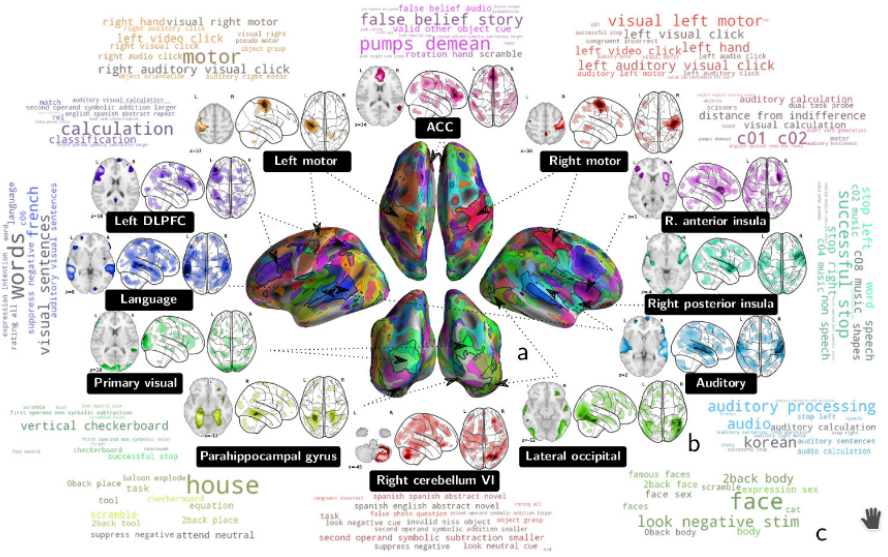


Figure 25. Visualization of some of task-optimized networks. Our approach allows to learn networks that are important for inter-subject decoding across studies. These networks, individually focal and collectively well spread across the cortex, are readily associated with the cognitive tasks that they contribute to predict. We display a selection of these networks, named with the salient anatomical brain region they recruit, along with a word-cloud representation of the stimuli whose likelihood increases with the network activation.

100 generations ago. The Loter software package and its source code are available at https://github.com/bcm-uga/Loter.

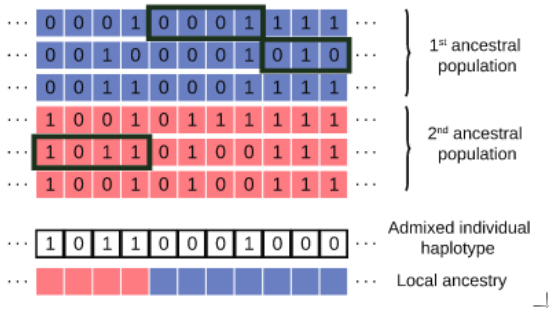


Figure 26. Graphical description of Local Ancestry Inference as implemented in the software Loter. Given a collection of parental haplotypes from the source populations depicted in blue and red, Loter assumes that an haplotype of an admixed individuals is modeled as a mosaic of existing parental haplotypes.

WILLOW Project-Team

7. New Results

7.1. 3D object and scene modeling, analysis, and retrieval

7.1.1. Indoor Visual Localization with Dense Matching and View Synthesis

Participants: Hajime Taira, Masatoshi Okutomi, Torsten Sattler, Mircea Cimpoi, Marc Pollefeys, Josef Sivic, Tomas Pajdla, Akihiko Torii.

In [20], we seek to predict the 6 degree-of-freedom (6DoF) pose of a query photograph with respect to a large indoor 3D map. The contributions of this work are three-fold. First, we develop a new large-scale visual localization method targeted for indoor environments. The method proceeds along three steps: (i) efficient retrieval of candidate poses that ensures scalability to large-scale environments, (ii) pose estimation using dense matching rather than local features to deal with textureless indoor scenes, and (iii) pose verification by virtual view synthesis to cope with significant changes in viewpoint, scene layout, and occluders. Second, we collect a new dataset with reference 6DoF poses for large-scale indoor localization. Query photographs are captured by mobile phones at a different time than the reference 3D map, thus presenting a realistic indoor localization scenario. Third, we demonstrate that our method significantly outperforms current state-of-the-art indoor localization approaches on this new challenging data. Figure 1 presents some example results.

7.1.2. Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions

Participants: Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, Frederik Kahl, Tomas Pajdla.

Visual localization enables autonomous vehicles to navigate in their surroundings and augmented reality applications to link virtual to real worlds. Practical visual localization approaches need to be robust to a wide variety of viewing condition, including day-night changes, as well as weather and seasonal variations, while providing highly accurate 6 degree-of-freedom (6DOF) camera pose estimates. In [19], we introduce the first benchmark datasets specifically designed for analyzing the impact of such factors on visual localization. Using carefully created ground truth poses for query images taken under a wide variety of conditions, we evaluate the impact of various factors on 6DOF camera pose estimation accuracy through extensive experiments with state-of-the-art localization approaches. Based on our results, we draw conclusions about the difficulty of different conditions, showing that long-term localization is far from solved, and propose promising avenues for future work, including sequence-based localization approaches and the need for better local features. Our benchmark is available at visuallocalization.net. Figure 2 presents some example results.

7.1.3. Changing Views on Curves and Surfaces

Participants: Kathlen Kohn, Bernd Sturmfels, Matthew Trager, Boris Bukh, Xavier Goaoc, Alfredo Hubard, Matthew Trager.

Visual events in computer vision are studied from the perspective of algebraic geometry. Given a sufficiently general curve or surface in 3-space, we consider the image or contour curve that arises by projecting from a viewpoint. Qualitative changes in that curve occur when the viewpoint crosses the visual event surface as illustrated in 3 . We examine the components of this ruled surface, and observe that these coincide with the iterated singular loci of the coisotropic hypersurfaces associated with the original curve or surface. We derive formulas, due to Salmon and Petitjean, for the degrees of these surfaces, and show how to compute exact representations for all visual event surfaces using algebraic methods. This work has been published in [8].

subsectionConsistent Sets of Lines with no Colorful Incidence

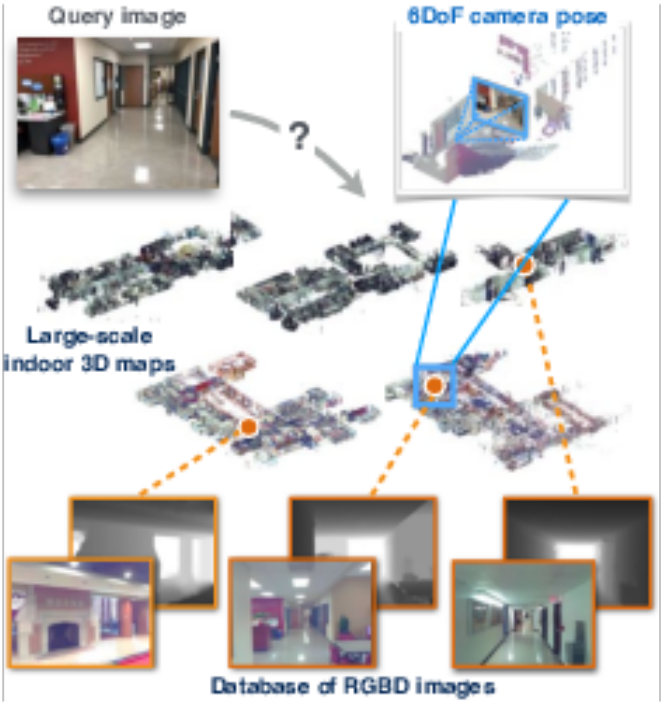


Figure 1. Large-scale indoor visual localization. Given a database of geometrically-registered RGBD images, we predict the 6DoF camera pose of a query RGB image by retrieving candidate images, estimating candicate camera poses, and selecting the best matching camera pose. To address inherent difficulties in indoor visual localization, we introduce the ?InLoc? approach that performs a sequence of progressively stricter verification steps.

Figure 2. Visual localization in changing urban conditions. We present three new datasets, Aachen Day-Night, RobotCar Seasons (shown) and CMU Seasons for evaluating 6DOF localization against a prior 3D map (top) using registered query images taken from a wide variety of conditions (bottom), including day-night variation, weather, and seasonal changes over long periods of time.

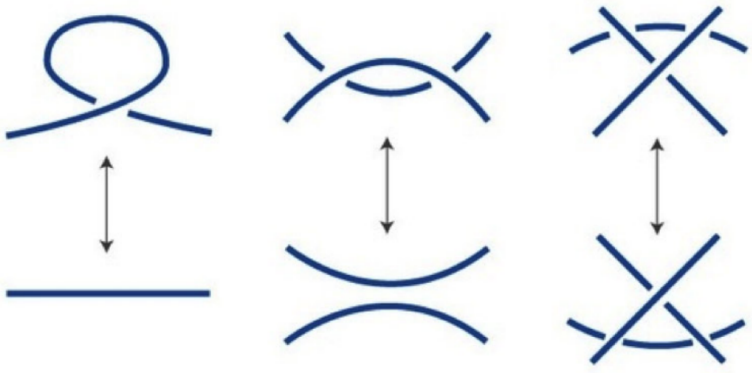


Figure 3. Changing views of a curve correspond to Reidemeister moves. The viewpoint z crosses the tangential surface (left), edge surface (middle), or trisecant surface (right).

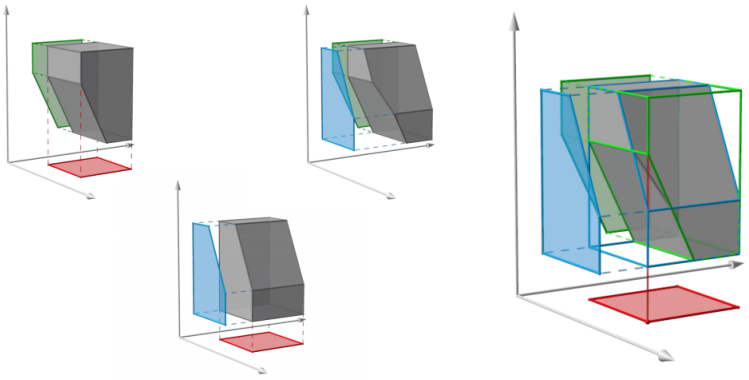


Figure 4. Three silhouettes that are 2-consistent but not globally consistent for three orthogonal projections. Each of the first three figures shows a three-dimensional set that projects onto two of the three silhouettes. The fourth figure illustrates that no set can project simultaneously onto all three silhouettes: the highlighted red image point cannot be lifted in 3D, since no point that projects onto it belongs to the pre-images of both the blue and green silhouettes.

We consider incidences among colored sets of lines in $\mathbb{R}^d$ and examine whether the existence of certain concurrences between lines of k colors force the existence of at least one concurrence between lines of $k+1$ colors. This question is relevant for problems in 3D reconstruction in computer vision such as the one illustrated in Figure 4 . This work has been published in [12].

7.1.4. On the Solvability of Viewing Graphs

Participants: Matthew Trager, Brian Osserman, Jean Ponce.

A set of fundamental matrices relating pairs of cameras in some configuration can be represented as edges of a ” viewing graph ". Whether or not these fundamental matrices are generically sufficient to recover the global camera configuration depends on the structure of this graph. We study characterizations of ” solvable " viewing graphs, and present several new results that can be applied to determine which pairs of views may be used to recover all camera parameters. We also discuss strategies for verifying the solvability of a graph computationally. This work has been published in [21].

7.1.5. In Defense of Relative Multi-View Geometry

Participants: Matthew Trager, Jean Ponce.

The idea of studying multi-view geometry and structure-from-motion problems *relative* to the scene and camera configurations, without appeal to external coordinate systems, dates back to the early days of modern geometric computer vision. Yet, it has a bad rap, the scene reconstructions obtained often being deemed as inaccurate despite careful implementations. The aim of this article is to correct this perception with a series of new results. In particular, we show that using a small subset of scene and image points to parameterize their relative configurations offers a natural coordinate-free formulation of Carlsson-Weinshall duality for arbitrary numbers of images. An example is shown in Figure 5 . For three views, this approach also yields novel purely- and quasi-linear formulations of structure from motion using *reduced trilinearities*, without the complex polynomial constraints associated with trifocal tensors, revealing in passing the strong link between “3D” ($\mathbb{P}^3 \to \mathbb{P}^2$) and “2D” ($\mathbb{P}^2 \to \mathbb{P}^1$) models of trinocular vision. Finally, we demonstrate through preliminary experiments that the proposed relative reconstruction methods gives good results on real data. This works is available as a preprint [32].

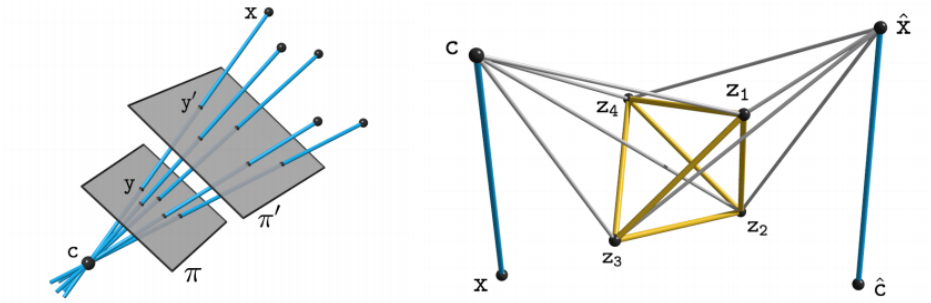


Figure 5. Configurations. ***Left****: Image point and viewing ray configurations are isomorphic and independent of the retinal plane.* ***Right****: Geometric Carlsson-Weinshall duality between scene point and pinhole configurations.*

7.1.6. Multigraded Cayley-Chow Forms

Participants: Brian Osserman, Matthew Trager.

We introduce a theory of multigraded Cayley-Chow forms associated to subvarieties of products of projective spaces. Figure 6 illustrares some examples of projective speces. Two new phenomena arise: first, the construction turns out to require certain inequalities on the dimensions of projections; and second, in positive characteristic the multigraded Cayley-Chow forms can have higher multiplicities. The theory also provides a natural framework for understanding multifocal tensors in computer vision. This works is available as a preprint [30].

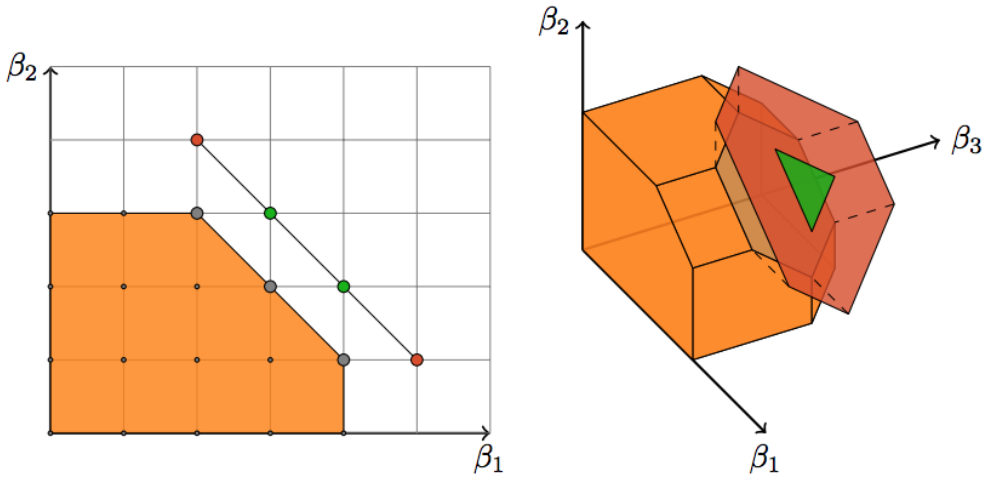


Figure 6. Two polymatroids. The sets of bases (corresponding to our multidegree supports) are in gray; while the sets of circuits and of non-circuit 1-deficient vectors are in green and red, respectively.

7.2. Category-level object and scene recognition

7.2.1. Detecting rare visual relations using analogies

Participants: Julia Peyre, Cordelia Schmid, Ivan Laptev, Josef Sivic.

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 rare or unseen at training such as shown in Figure 7 . 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 embeddings 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 two challenging datasets involving rare and unseen relations : on HICO-DET, our model achieves significant improvement over a strong baseline, and we confirm this improvement on retrieval of unseen triplets on the UnRel rare relation dataset. This work, currently under review, can be found at [31].

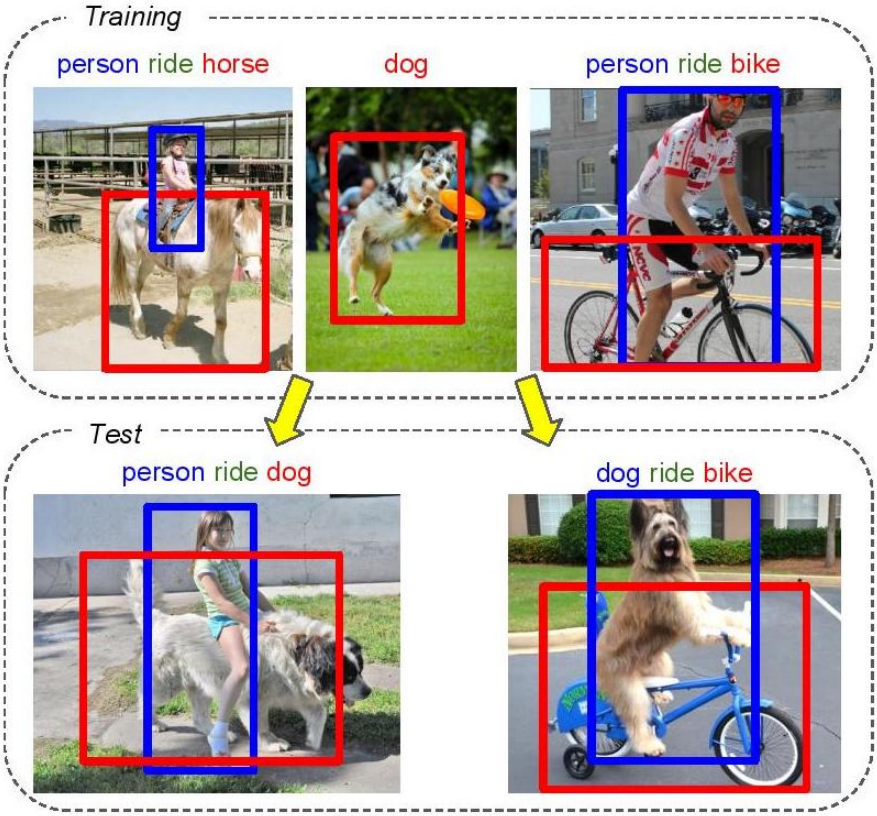


Figure 7. Illustration of transfer by analogy from seen training triplets (e.g. "person ride horse") to unseen or rare ones (e.g. "person ride dog"))

7.2.2. Convolutional neural network architecture for geometric matching

Participants: Ignacio Rocco, Relja Arandjelović, Josef Sivic.

In [9], we address the problem of determining correspondences between two images in agreement with a geometric model such as an affine, homography or thin-plate spline transformation, and estimating its parameters. The contributions of this work are threefold. First, we propose a convolutional neural network architecture for geometric matching. The architecture is based on three main components that mimic the standard steps of feature extraction, matching and simultaneous inlier detection and model parameter estimation, while being trainable end-to-end. Second, we demonstrate that the network parameters can be trained from synthetically generated imagery without the need for manual annotation and that our matching layer significantly increases generalization capabilities to never seen before images. Finally, we show that the same model can perform both instance-level and category-level matching giving state-of-the-art results on the challenging PF, TSS and Caltech-101 datasets.

7.2.3. End-to-end weakly-supervised semantic alignment

Participants: Ignacio Rocco, Relja Arandjelović, Josef Sivic.

In [17], we tackle the task of semantic alignment where the goal is to compute dense semantic correspondence aligning two images depicting objects of the same category. This is a challenging task due to large intra-class variation, changes in viewpoint and background clutter. We present the following three principal contributions. First, we develop a convolutional neural network architecture for semantic alignment that is trainable in an end-to-end manner from weak image-level supervision in the form of matching image pairs. The outcome is that parameters are learnt from rich appearance variation present in different but semantically related images without the need for tedious manual annotation of correspondences at training time. Second, the main component of this architecture is a differentiable soft inlier scoring module, inspired by the RANSAC inlier scoring procedure, that computes the quality of the alignment based on only geometrically consistent correspondences thereby reducing the effect of background clutter. Third, we demonstrate that the proposed approach achieves state-of-the-art performance on multiple standard benchmarks for semantic alignment. Figure 8 presents some example results.

7.2.4. Neighbourhood Consensus Networks

Participants: Ignacio Rocco, Mircea Cimpoi, Relja Arandjelović, Akihiko Torii, Tomas Pajdla, Josef Sivic.

In [18], we address the problem of finding reliable dense correspondences between a pair of images. This is a challenging task due to strong appearance differences between the corresponding scene elements and ambiguities generated by repetitive patterns. The contributions of this work are threefold. First, inspired by the classic idea of disambiguating feature matches using semi-local constraints, we develop an end-to-end trainable convolutional neural network architecture that identifies sets of spatially consistent matches by analyzing neighbourhood consensus patterns in the 4D space of all possible correspondences between a pair of images without the need for a global geometric model. Second, we demonstrate that the model can be trained effectively from weak supervision in the form of matching and non-matching image pairs without the need for costly manual annotation of point to point correspondences. Third, we show the proposed neighbourhood consensus network can be applied to a range of matching tasks including both category- and instance-level matching, obtaining the state-of-the-art results on the PF Pascal dataset and the InLoc indoor visual localization benchmark. Figure 9 shows the network architecture of the proposed Neighbourhood Consensus Network, that features 3 layers of 4D convolutions.

7.2.5. Compressing the Input for CNNs with the First-Order Scattering Transform

Participants: Edouard Oyallon, Eugene Belilovsky, Sergey Zagoruyko, Michal Valko.

In [16], we study the first-order scattering transform as a candidate for reducing the signal processed by a convolutional neural network (CNN). We study this transformation and show theoretical and empirical evidence that in the case of natural images and sufficiently small translation invariance, this transform preserves most of the signal information needed for classification while substantially reducing the spatial resolution and total signal size. We show that cascading a CNN with this representation performs on par with ImageNet classification models commonly used in downstream tasks such as the ResNet-50. We subsequently apply our trained hybrid ImageNet model as a base model on a detection system, which has typically larger

Figure 8. Each row corresponds to one example and shows the (right) automatic semantic alignment of the (left) source and (middle) target images.

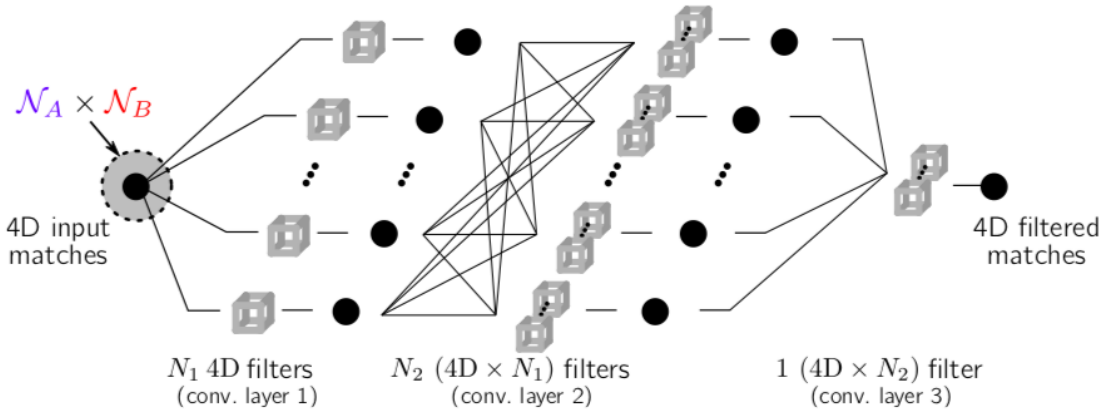


Figure 9. A neighbourhood consensus CNN operates on the 4D space of feature matches. The first 4D convolutional layer filters span $\mathcal{N}_A \times \mathcal{N}_B$, the Cartesian product of local neighbourhoods $\mathcal{N}_A$ and $\mathcal{N}_B$ in images A and B respectively. The proposed 4D neighbourhood consensus CNN can learn to identify the matching patterns of reliable and unreliable matches, and filter the matches accordingly

image inputs. On Pascal VOC and COCO detection tasks we deliver substantial improvements in the inference speed and training memory consumption compared to models trained directly on the input image.

7.2.6. Exploring Weight Symmetry in Deep Neural Networks

Participants: Xu Shell Hu, Sergey Zagoruyko, Nikos Komodakis.

In [27], 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, 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.

7.3. Image restoration, manipulation and enhancement

7.3.1. Neural Embedding of an Iterative Deconvolution Algorithm for Motion Blur Estimation and Removal

Participants: Thomas Eboli, Jian Sun, Jean Ponce.

We introduce a new two-steps learning-based approach to motion blur estimation and removal decomposed into two trainable modules. A local linear motion model is estimated at each pixel using a first convolutional neural network (CNN) in a regression setting. It is then used to drive an algorithm that casts non-blind, non-uniform image deblurring as a least-squares problem regularized by natural image priors in the form of sparsity constraints. This problem is solved by combining the alternative direction method of multipliers with an iterative residual compensation algorithm, with a finite number of iterations embedded into a second CNN whose trainable parameters are deconvolution filters. The second network outputs the sharp image, and the

two CNNs can be trained together in an end-to-end manner. Our experiments demonstrate that the proposed method is significantly faster than existing ones, and provides competitive results with the state of the art on synthetic and real data. This work is available as a pre-print[25] and an example is illustrated in Figure 10 .

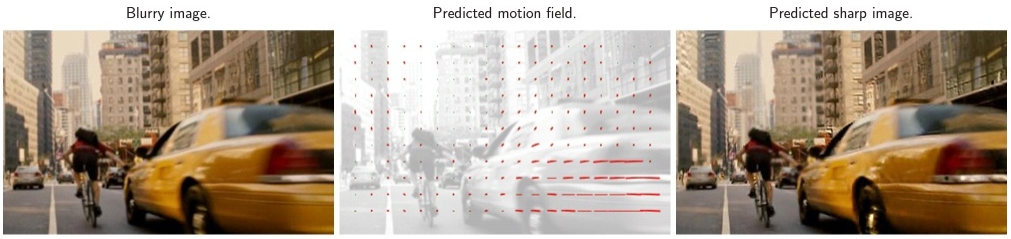


Figure 10. From a blurry image, we first use CNN-based regressor to predict a motion field with local linear motions before using it in a trainable iterative residual compensation algorithm to restore the image.

7.3.2. Deformable Kernel Networks for Joint Image Filtering

Participants: Beomjun Kim, Jean Ponce, Bumsub Ham.

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. 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 by 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 by 3 kernels outperforms the state of the art by a significant margin. This works has been submitted to the IEEE Trans. on Pattern Analysis and Machine Intelligence and is available as a pre-print [28].

7.4. Human activity capture and classification

7.4.1. Learning a Text-Video Embedding from Incomplete and Heterogeneous Data

Participants: Antoine Miech, Ivan Laptev, Josef Sivic.

Joint understanding of video and language is an active research area with many applications. Prior work in this domain typically relies on learning text-video embeddings. One difficulty with this approach, however, is the lack of large-scale annotated video-caption datasets for training. To address this issue, in [29] we aim at learning text-video embeddings from heterogeneous data sources. To this end, we propose a Mixture-of-Embedding-Experts (MEE) model with ability to handle missing input modalities during training. As a result, our framework can learn improved text-video embeddings simultaneously from image and video datasets. We also show the generalization of MEE to other input modalities such as face descriptors. We evaluate our method on the task of video retrieval and report results for the MPII Movie Description and MSR-VTT datasets. The proposed MEE model demonstrates significant improvements and outperforms previously reported methods on both text-to-video and video-to-text retrieval tasks. Figure 11 illustrates application of our method in text-to-video retrieval.

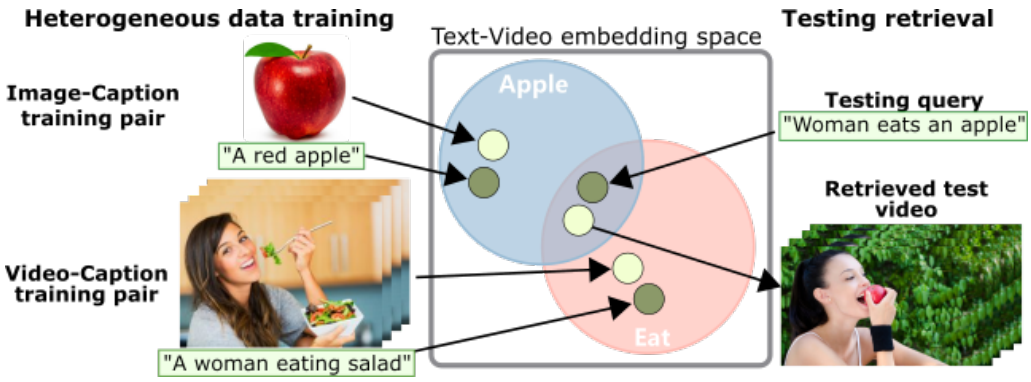


Figure 11. We learn a text-video embedding from heterogenous (here Image-Text and Video-Text) data sources. At test time, we can query concepts learnt from both Image-Caption and Video-Caption training pair (e.g. the eating notion being learnt from video and the apple notion from image).

7.4.2. A flexible model for training action localization with varying levels of supervision

Participants: Guilhem Chéron, Jean-Baptiste Alayrac, Ivan Laptev, Cordelia Schmid.

Spatio-temporal action detection in videos is typically addressed in a fully-supervised setup with manual annotation of training videos required at every frame. Since such annotation is extremely tedious and prohibits scalability, there is a clear need to minimize the amount of manual supervision. In this work we propose a unifying framework that can handle and combine varying types of less-demanding weak supervision. Our model is based on discriminative clustering and integrates different types of supervision as constraints on the optimization as illustrated in Figure 12 . We investigate applications of such a model to training setups with alternative supervisory signals ranging from video-level class labels to the full per-frame annotation of action bounding boxes. Experiments on the challenging UCF101-24 and DALY datasets demonstrate competitive performance of our method at a fraction of supervision used by previous methods. The flexibility of our model enables joint learning from data with different levels of annotation. Experimental results demonstrate a significant gain by adding a few fully supervised examples to otherwise weakly labeled videos. This work has been published in [14].

7.4.3. BodyNet: Volumetric Inference of 3D Human Body Shapes

Participants: Gül Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan Laptev, Cordelia Schmid.

Human shape estimation is an important task for video editing, animation and fashion industry. Predicting 3D human body shape from natural images, however, is highly challenging due to factors such as variation in human bodies, clothing and viewpoint. Prior methods addressing this problem typically attempt to fit parametric body models with certain priors on pose and shape. In this work we argue for an alternative representation and propose BodyNet, a neural network for direct inference of volumetric body shape from a single image. BodyNet is an end-to-end trainable network that benefits from (i) a volumetric 3D loss, (ii) a multi-view re-projection loss, and (iii) intermediate supervision of 2D pose, 2D body part segmentation, and 3D pose. Each of them results in performance improvement as demonstrated by our experiments. To evaluate the method, we fit the SMPL model to our network output and show state-of-the-art results on the SURREAL and Unite the People datasets, outperforming recent approaches. Besides achieving state-of-the-art performance, our method also enables volumetric body-part segmentation. Figure 13 illustrates the volumetric outputs given two sample input images. This work has been published at ECCV 2018 [22].

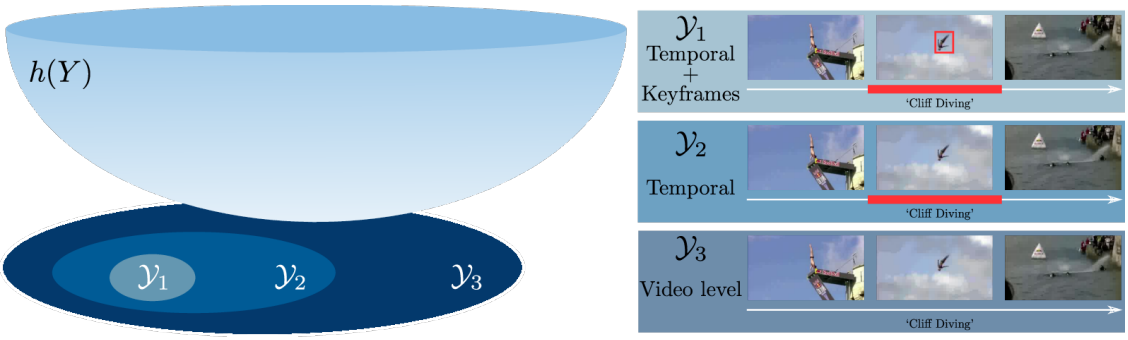


Figure 12. Our method estimates a matrix Y assigning human tracklets to action labels in training videos by optimizing an objective function $h(Y)$ under constraints $\mathcal{Y}_s$. Different types of supervision define particular constraints $\mathcal{Y}_s$ and do not affect the form of the objective function. The increasing level of supervision imposes stricter constraints, e.g. $\mathcal{Y}_1 \supset \mathcal{Y}_2 \supset \mathcal{Y}_3 \supset \mathcal{Y}_4$ as illustrated for the Cliff Diving example above.

Figure 13. Our BodyNet predicts a volumetric 3D human body shape and 3D body parts from a single image. We show the input image, the predicted human voxels, and the predicted part voxels.

7.4.4. Localizing Moments in Video with Temporal Language

Participants: Lisa Anne Hendricks, Oliver Wang, Eli Schechtman, Josef Sivic, Trevor Darrell, Bryan Russell.

Localizing moments in a longer video via natural language queries is a new, challenging task at the intersection of language and video understanding. Though moment localization with natural language is similar to other language and vision tasks like natural language object retrieval in images, moment localization offers an interesting opportunity to model temporal dependencies and reasoning in text. In [15], we propose a new model that explicitly reasons about different temporal segments in a video, and shows that temporal context is important for localizing phrases which include temporal language. To benchmark whether our model, and other recent video localization models, can effectively reason about temporal language, we collect the novel TEMPO-ral reasoning in video and language (TEMPO) dataset. Our dataset consists of two parts: a dataset with real videos and template sentences (TEMPO - Template Language) which allows for controlled studies on temporal language, and a human language dataset which consists of temporal sentences annotated by humans (TEMPO - Human Language).

7.4.5. 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.

In this work, 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.

7.4.6. Modeling Spatio-Temporal Human Track Structure for Action Localization

Participants: Guilhem Chéron, Anton Osokin, Ivan Laptev, Cordelia Schmid.

This paper [24] addresses spatio-temporal localization of human actions in video. In order to localize actions in time, we propose a recurrent localization network (RecLNet) designed to model the temporal structure of actions on the level of person tracks. Our model is trained to simultaneously recognize and localize action classes in time and is based on two layer gated recurrent units (GRU) applied separately to two streams, i.e. appearance and optical flow streams. When used together with state-of-the-art person detection and tracking, our model is shown to improve substantially spatio-temporal action localization in videos. The gain is shown to be mainly due to improved temporal localization as illustrated in Figure 14 . We evaluate our method on two recent datasets for spatio-temporal action localization, UCF101-24 and DALY, demonstrating a significant improvement of the state of the art.

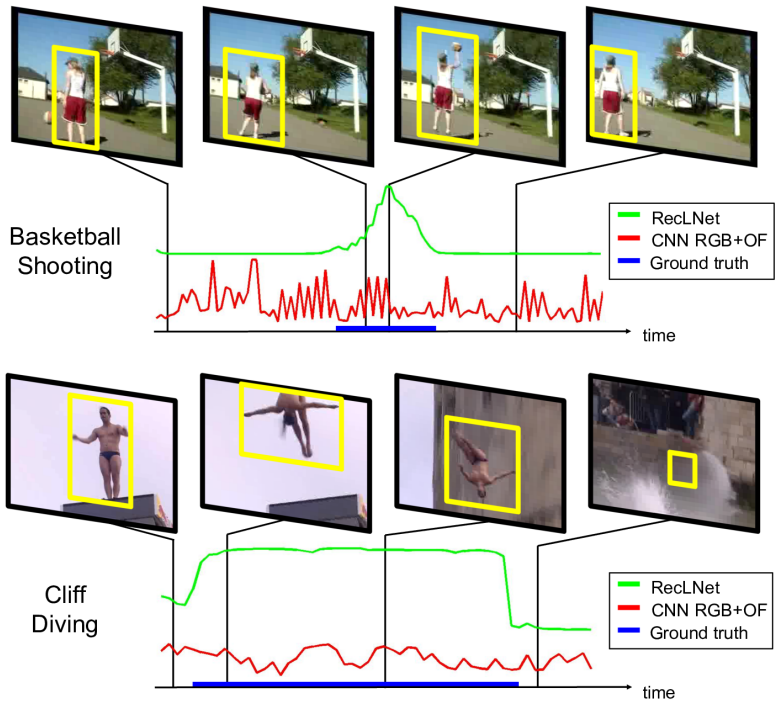


Figure 14. Spatio-temporal action localization using a CNN baseline (red) and our RecLNet (green) both applied on the level of person tracks. Our approach provides accurate temporal boundaries when the action happens.