RESEARCH CENTER

FIELD
Perception, Cognition and Interaction

Activity Report 2016

Section New Results

Edition: 2017-08-25

CEDAR Team

7. New Results

7.1. Scalable Heterogeneous Stores

To improve data querying performance within polystores (Section 3.1), we developed Estocada, a novel system capable of exploiting side-by-side a practically unbound variety of data management system, all the while guaranteeing the soundness and completeness of the store, and striving to extract the best performance out of the various DMSs. Estocada leverages recent advances in the area of query rewriting under constraints, which we use to capture the various data models and describe the fragments stored within each data management system. Estocada was demonstrated at the IEEE ICDE conference [12]; recent experimental results demonstrated performance improvements by many orders of magnitude brought by the fragments Estocada supports, with respect to the setting where data is stored only in the system it originates from. This work continues, in collaboration with Alin Deutsch and Rana Alotaibi from UCSD.

7.2. Semantic Query Answering

This is a core topic for the team, in which the year has been particularly fruitful.

First, we investigated efficient query answering techniques in knowledge bases. A large and useful set of ontologies enjoys FOL (first-order logic) reducibility of query answering, that is: answering a query q can be reduced to evaluating a certain first-order logic (FOL) formula (obtained from the query and ontology) against only the explicit facts. We devised a novel query optimization framework for ontology-based data access settings enjoying FOL reducibility. Our framework is based on searching within a set of alternative equivalent FOL queries, that is, FOL reformulations, one with minimal evaluation cost when evaluated through a relational database system. We applied this framework to the DL-Lite$_{\mathcal{R}}$ Description Logic underpinning the W3C's OWL2 QL ontology language, and demonstrated through experiments its performance benefits when two leading SQL systems, one open-source and one commercial, are used for evaluating the FOL query reformulations. This work has lead to a major publication in the PVLDB journal [13], and a demonstration at the Semantic Web conference [4], while the complete details appear in [16] and the PhD thesis of the student author. [2].

Second, we initiated a study of extensions of conjunctive queries to conjunctive regular path queries. The first step has been to study regular path queries under linear existential rules, generalizing previous work on DL-Lite$_{\mathcal{R}}$, which is at the core of the Semantic Web OWL 2 QL profile. Regular path queries are queries that check for a path between two individuals, which is labeled by a word belonging to a given regular language. Such navigational languages are very popular for graph-based data representation, such as RDF. We have studied the complexity for this query language, and shown that it is NL-complete in data complexity, and EXPTIME-complete in combined complexity (and PTIME complete with bounded arity). This work has received the best paper award at RR'16 , and is currently being extented to conjunctive regular path queries.

Last, we studied the expressivity of several variants of Datalog, the classical language for deductive databases. In particular, we have studied its expressivity when given access (or not) to input negation (the ability to check if an extensional atoms hold or not) and to a linear order. We provided a complete Venn diagram regarding the expressivity of all the variants when considering homomorphism-closed query. The trickiest (and most surprising) points is the existence of polynomial-time computable homomorphism closed queries that are not expressible within Datalog with linear order but without input negation. These results have been published at IJCAI'16 [7].

7.3. Multi-model Querying

We have proposed a lightweight data integration architecture implemented within Tatooine (see Section 6.1.5); the system was demonstrated on a data journalism use case at the prestigious VLDB conference [9].

A separate effort in the area of multi-model querying considered querying databases of interconnected documents, users and concepts, by means of keywords. In this context, it is important that query results reflect not only the keywords present in documents but also the links between users and documents (so as to return to one user first the results authored in his social neighborhood), links between documents (for instance when a tweet answers another or an article has a link to another), and last but not least semantic information which allows interconnecting and interpreting terms mentioned in text. This research was finalized as part of the PhD of Raphaël Bonaque [1] and appeared at the EDBT conference 2016 [11].

7.4. Interactive Data Exploration at Scale

In the work with Enhui Huang (PhD student at Ecole Polytechnique), we seek to minimize the number of samples presented to the user for reviewing in order to build an accurate model of the user interest. In particular, as the dimensionality of the data space increases, the number of samples needed to build an accurate user interest model increases fast. We examine a range of popular feature selection techniques for data exploration, and for the best-performing feature selection technique, Gradient boosting regression trees (GBRT), we propose optimizations to overcome the issue of unbalanced training data and to dynamically determine the number of relevant features to select. Experimental results show that our optimized GBRT improves F-measure from nearly 0 without feature selection, to high F-measure (>0.8), by adaptively choosing the number of relevant features.

This work is currently under submission to a database conference.

7.5. Exploratory Querying of Semantic Graphs

We have started work with an intern (Zheng Zhang) toward automatically exploring the structure of an RDF graph and visualizing it with the help of a D3.js (https://d3js.org/) visualization library. These initial steps should serve to guide the beginning of an interactive exploration of the RDF graph in order to identify interesting analytical queries to be asked and evaluated. This work continues.

Separately, with a different intern (Javier Letelier), we have investigated efficient algorithms for keyword search in an RDF graph, exploiting structural and semantic knowledge about the graph; such knowledge is organized as an RDF summary which is an RDF graph itself. The algorithm was implemented and integrated as a text search tool within the Tatooine prototype; the work is ongoing.

DAHU Project-Team

6. New Results

6.1. Specification and verification of data-driven systems

Verification of Hierarchical Artifact Systems

Data-driven workflows, of which "business artifacts" are a prime exponent, have been successfully deployed in practice, adopted in industrial standards, and have spawned a rich body of research in academia, focused primarily on static analysis. Over the past few years, we have embarked upon a study of the verification problem for artifact systems. This is a challenging problem because of the presence of unbounded data. In order to deal with the resulting infinite-state system, we developed in earlier work a symbolic approach allowing a reduction to finite-state model checking and yielding a pspace verification algorithm for the simplest variant of the model (no database dependencies and uninterpreted data domain). Subsequently, we extended our approach to allow for database dependencies and numeric data testable by arithmetic constraints. In [19], we make significant progress on several fronts, by considering a much richer and more realistic model than in previous work, incorporating core elements of IBM's successful Guard-Stage-Milestone model. In particular, the model features task hierarchy, concurrency, and richer artifact data. It also allows database key and foreign key dependencies, as well as arithmetic constraints. The results require qualitatively novel techniques, because the reduction to finite-state model checking used in previous work is no longer possible. Instead, the richer model requires the use of a hierarchy of Vector Addition Systems with States. The arithmetic constraints are handled using quantifier elimination techniques, adapted to our setting.

Process-centric views of data-driven workflows.

We also studied the models of *data Petri nets* and *ν-Petri nets*. While these models were introduced in the verification community to analyse protocols and process algebra, they can also be seen as (very limited) data-driven workflows with only unary predicates. Our results this year show that various boundedness problems (e.g. can the database grow unbounded?) are decidable in data Petri nets [22], and pinpoint the exact complexity of safety analysis in ν-Petri nets [23].

Complexity in counter systems and in proof systems.

The static analysis of queries on XML trees and data streams relies in a majority of cases on decision procedures expressed in terms of formal systems like counter systems or proof systems. For instance, two-variables first-order data queries on words can be related to reachability in vector addition systems (VAS), and the same queries on trees to reachability in a branching extension of VAS [12]. We are at the forefront on the complexity analysis for such systems [15], [13], [16], [14].

We investigate in the ANR PRODAQ project a different angle on the static analysis of queries, relying on proof systems. Our first results on the subject [18] provide a sequent calculus for a modal data logic with an optimal proof-search algorithm.

6.2. Personal information management.

Thymeflow We developed Thymeflow, a personal knowledge base with spatio-temporal data [24].

The typical Internet user has data spread over several devices and across several online systems. We demonstrate an open-source system for integrating user's data from different sources into a single Knowledge Base. Our system integrates data of different kinds into a coherent whole, starting with email messages, calendar, contacts, and location history. It is able to detect event periods in the user's location data and align them with calendar events. We will demonstrate how to query the system within and across different dimensions, and perform analytics over emails, events, and locations.

EXMO Project-Team

6. New Results

6.1. Ontology matching and alignments

6.1.1. Evaluation

Participant: Jérôme Euzenat [Correspondent].

Since 2004, we run the Ontology Alignment Evaluation Initiative (OAEI) which organises evaluation campaigns for assessing the degree of achievement of actual ontology matching algorithms [3].

This year, we used again our generator for generating a new version of benchmarks. The Alignment API was used for manipulating alignments and evaluating results [8].

The participating systems and evaluation results were presented in the 11th Ontology Matching workshop [14], [15], held Kobe (JP). More information on OAEI can be found at http://oaei.ontologymatching.org/.

6.1.2. Algebras of alignment relations

Participants: Manuel Atencia Arcas, Jérôme Euzenat [Correspondent], Armen Inants.

Qualitative calculi are central in qualitative binary constraint satisfaction problems. All these qualitative calculi share an implicit assumption that the universe is homogeneous, i.e., consists of objects of the same kind. However, objects of different kinds may also entertain relations. Many applications discriminate between different kinds of objects. For example, some spatial models discriminate between regions, lines and points, and different relations are used for each kind of objects. In ontology matching, qualitative calculi were shown useful for expressing alignments between only one kind of entities, such as concepts or individuals. However, relations between individuals and concepts, which impose additional constraints, are not exploited.

We introduced modularity in qualitative calculi and provided a methodology for modeling qualitative calculi with heterogeneous universes [5]. It is based on a special class of partition schemes which we call modular. For a qualitative calculus generated by a modular partition scheme, we can define a structure that associates each relation symbol with an abstract domain and codomain from a Boolean lattice of sorts. A module of such a qualitative calculus is a sub-calculus restricted to a given sort, which is obtained through an operation called relativisation to a sort. Of a greater practical interest is the opposite operation, which allows for combining several qualitative calculi into a single calculus. We defined an operation called combination modulo glue, which combines two or more qualitative calculi over different universes, provided some glue relations between these universes. The framework is general enough to support most known qualitative spatio-temporal calculi.

In 2012, we introduced a semantics supporting confidences in correspondences as weights. However, it introduced a discontinuity between weighted and non-weighted interpretations. Moreover, it does not provide a calculus for reasoning with weighted ontology alignments. We introduced a calculus for such alignments [11] provided by an infinite relation-type algebra, the elements of which are weighted taxonomic relations. In addition, it approximates the non-weighted case in a continuous manner.

This work has been part of the PhD of Armen Inants [5] partially funded by the LINDICLE project (§7.1.1).

6.2. Data interlinking

The web of data uses semantic web technologies to publish data on the web in such a way that they can be interpreted and connected together. It is thus important to be able to establish links between these data [7], both for the web of data and for the semantic web that it contributes to feed. We consider this problem from different perspectives.

6.2.1. Interlinking cross-lingual RDF data sets

Participants: Tatiana Lesnikova, Jérôme David [Correspondent], Jérôme Euzenat.

RDF data sets are being published with labels that may be expressed in different languages. Even systems based on graph structure, ultimately rely on anchors based on language fragments. In this context, data interlinking requires specific approaches in order to tackle cross-lingualism. We proposed a general framework for interlinking RDF data in different languages and implemented two approaches: one approach is based on machine translation, the other one takes advantage of multilingual references, such as BabelNet.

This year, we evaluated machine translation for interlinking concepts, i.e., generic entities named with a common noun or term, as opposed to individual entities. In previous work, the evaluated method has been applied on named entities. We conducted two experiments involving different thesauri in different languages. The first experiment involved concepts from the TheSoz multilingual thesaurus in three languages: English, French and German. The second experiment involved concepts from the EuroVoc and AGROVOC thesauri in English and Chinese respectively. We demonstrated that machine translation can be beneficial for cross-lingual thesauri interlining independently of a dataset structure [12].

This work has been part of the PhD of Tatiana Lesnikova [6] developed in the LINDICLE project (§7.1.1).

6.2.2. Uncertainty-sensitive reasoning for inferring sameAs facts in Linked Data

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

A major challenge in data interlinking is to design tools that effectively deal with incomplete and noisy data, and exploit uncertain knowledge. We modelled data interlinking as a reasoning problem with uncertainty. For that purpose, we introduced a probabilistic framework for modelling and reasoning over uncertain RDF facts and rules that is based on the semantics of probabilistic Datalog. We have designed an algorithm, ProbFR, based on this framework. Experiments on real-world datasets have shown the usefulness and effectiveness of our approach for data linkage and disambiguation [9].

This work was carried out in collaboration with Mustafa Al-Bakri and Marie-Christine Rousset (LIG).

6.2.3. Tableau extensions for reasoning with link keys

Participants: Manuel Atencia Arcas [Correspondent], Jérôme Euzenat, Maroua Gmati.

Link keys allow for generating links across datasets expressed in different ontologies (see §3.3). But they can also be thought of as axioms in a description logic. As such, they can contribute to infer ABox axioms, such as links, or terminological axioms and other link keys. Yet, no reasoning support existed for link keys. We extended the tableau method designed for ALC to take link keys into account [10]. We showed how this extension enables combining link keys with classical terminological reasoning with and without ABox and TBox and generate non trivial link keys.

GRAPHIK Project-Team

6. New Results

6.1. Logics and Graph-Based Languages for Ontology-Mediated Query Answering

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

Ontolology-mediated query answering (and more generally *Ontology-Based Data Access*, *OBDA*) is a recent paradigm in data management, which takes into account inferences enabled by an ontology when querying data. In other words, the notion of a database is replaced by that of a knowledge base, composed of data (also called facts) and of an ontology. Two families of formalisms for representing and reasoning with the ontological component are considered in this context: *description logics* and the more recent *existential rule* framework. Until last year, the team has mainly investigated existential rules. This expressive formalism generalizes most lightweight description logics used in OBDA (such as $\mathcal{EL}$ and DL-Lite, on which OWL 2 tractable profiles are based) on the one hand, and Datalog, the language of deductive databases, on the other hand. With the arrival of Meghyn Bienvenu, description logics have joined the core formalisms studied by the team. Compared to existential rules, the description logics considered for OBDA lead to lower complexity classes and specific algorithmic techniques. Studying both formalisms is scientifically highly relevant, specially in the context of OBDA.

We have also broadened this research line by starting investigating ontological languages for non-relational data, an issue that has barely been considered yet.

Before presenting this year' results, we recall 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 initial facts 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 rules). The backward chaining process can be divided into two steps: first, the initial query is rewritten using the rules into a first-order query (typically a union of conjunctive queries, UCQ); then the rewritten query is evaluated against the initial factbase (again, as in a classical database system). Note that forward and backward processes do not halt for all kinds of existential rules nor all lightweight description logics.

6.1.1. New Results in the Description Logics Framework

When using Description Logics (DL) ontologies to access relational data, mappings are used to link the relational schema to the vocabulary of the ontology (which uses only unary and binary predicates). In order to debug and optimize DL-based OBDA systems, it is important to be able to analyze and compare ontology-mapping pairs, called OBDA specifications. Prior work in this direction compared specifications using classical notions of equivalence and entailment.

- We have explored an alternative approach in which two specifications are deemed equivalent if they give the same answers to the considered query or class of queries for all possible data sources. After formally defining such query-based notions of entailment and equivalence of OBDA specifications, we investigated the complexity of the resulting analysis tasks when the ontology is formulated in (fragments of) DL-Lite$_R$, which forms the basis for the Semantic Web ontological language OWL 2 QL.
  - *KR'16 [28]*

- We consider a range of Horn DLs for which query answering has polynomial data complexity, but which do not guarantee the existence of First-Order(FO)-rewritings of all queries. In order to extend the applicability of the FO-rewriting technique, a key task is to be able to identify specific ontology-query pairs that admit an FO-rewriting. This led us to study FO-rewritability of conjunctive queries in the presence of ontologies formulated in DLs ranging between $\mathcal{EL}$ and Horn-$\mathcal{SHIF}$, along with related query containment problems. Apart from providing characterizations, we established complexity results ranging from EXPTIME via NEXPTIME to 2EXPTIME, pointing out several interesting effects. In particular, FO-rewriting is more complex for conjunctive queries than for atomic queries when inverse roles are present, but not otherwise.
  - *IJCAI'16 [27]*

6.1.2. New Results in the Existential Rule Framework

Several new theoretical results have been obtained on ontology-mediated query answering with existential rules:

- While most work in the area of ontology-mediated query answering focuses on conjunctive queries, navigational queries are gaining increasing attention. In collaboration with Michael Thomazo (Inria CEDAR), we took a step towards a better understanding of the combination of navigational query languages and existential rules by pinpointing the (data and combined) complexities of evaluating path queries (more precisely, two-way regular path queries) over knowledge bases whose ontology is composed of linear existential rules (a class of rules that can be seen as a natural generalisation of the description logic DL-Lite$_{\mathcal{R}}$). We extended an algorithm tailored for DL-Lite$_{\mathcal{R}}$ and showed that, despite an exponential blow-up with respect to the maximum predicate arity, our algorithm was worst-case optimal.
  - *RR'16 (Best paper award) [29]*
- Boundedness is an important notion for optimizing the processing of rule languages, as it ensures that materialisation can be performed in a predefined number of steps, independently from the size of any factbase. We are currently studying several boundedness notions for existential rules that extend the well-known boundedness notion of Datalog, and investigate their relationships with properties ensuring the finiteness of the chase or query rewriting. One of our first results is that, for a natural notion of boundedness, bounded existential rules are exactly those at the intersection of finite expansion sets (which ensure that any factbase has a finite sound and complete saturation) and finite unification sets (which ensure that any conjunctive query can be finitely rewritten into a sound and complete union of conjunctive queries).
  - *DL'16 [36]*
- Finally, Swan Rocher's PhD thesis deepened the study of the decidability and complexity of conjunctive query answering for classes of existential rules added with transitivity rules (previous results were presented at IJCAI 2015 [48])

6.1.3. Querying NoSQL databases (Key-value stores)

Over the last decade, research efforts to develop algorithms for OBDA have built on the assumption that data conforms to relational structures (including RDF) and that the paradigm can be deployed on top of relational databases with conjunctive queries at the core (e.g., in SQL or SPARQL). However, this is not the prominent way on which data is today stored and exchanged, especially in the Web. Whether OBDA can be developed for non-relational structures, like those shared by increasingly popular NOSQL languages sustaining Big-Data analytics, is still an open question. In collaboration with Marie-Christine Rousset (University of Grenoble, LIG), we proposed the first framework for studying the problem of answering ontology-mediated queries on top of NOSQL key-value stores. More precisely, we formalized the core data model and basic queries of these systems and introduced a rule language (NO-RL) to express lightweight ontologies on top of data. We defined a sound and complete query rewriting technique and studied the decidability and data complexity of answering ontology-mediated queries depending on considered the rule fragment.

- *AAAI'16 [39]; DL'16 [40]*

6.2. Representing and Processing Imperfect Information

Participants: Abdallah Arioua, Jean-François Baget, Meghyn Bienvenu, Pierre Bisquert, Patrice Buche, Madalina Croitoru, Jérôme Fortin, Fabien Garreau, Abdelraouf Hecham, Marie-Laure Mugnier, Odile Papini, Swan Rocher, Rallou Thomopoulos, Bruno Yun.

Inconsistency-Tolerant Query Answering is one of the challenging problems that received a lot of attention in recent years as inconsistency may arise in practical applications due to several reasons: merging, integration, revision. In the context of Ontology-Based Data Access (OBDA), where the ontological knowledge is assumed to be coherent and fully reliable, inconsistency comes from the data, i.e., occurs when some assertional facts contradict some constraints imposed by the ontological knowledge. Existing works in this area have studied different inconsistency-tolerant query answering techniques, called “semantics”: some examples include Brave, IAR, ICR, AR etc..These proposals are closely related to works on querying inconsistent databases, or inference from inconsistent propositional logic knowledge bases.

The work of this year on inconsistency-tolerant query answering techniques for Ontology Based Data Access focused on (i) new results about different kinds of semantics or (ii) the user interaction with such semantics (we investigated the notion of repair based explanation or argumentation based explanation). We have investigated the interest of inconsistency-tolerant semantics in general and argumentation techniques for the agrifood chain in particular.

6.2.1. Inconsistency-Tolerant Semantics for Query Answering

In all approaches considered here, a knowledge base can have, in opposition to the logics studied in 6.1 , several incompatible “minimal” models. Those models can correspond to possible repairs of an inconsistent knowledge base or can be the models generated by a non-monotonic logic. The questions we address here are linked to the semantics (how to define those models, how to define preferences on those models), while trying to preserve a satisfying trade-off between expressivity and computational complexity of the querying mechanism.

- We proposed a new inconsistency-tolerant inference relation, called non-objection inference, where a query is considered as valid if it is entailed by at least one repair and it is consistent with all the other repairs. The main salient points of the newly introduced semantics is its efficiency (query answering with non-objection inference is achieved in polynomial time) and the fact that the inferences are strictly more productive than universal inference while preserving the consistency of its set of conclusions. The intuition behind is that no repair has an objection veto to the acceptance of the query. If query entailment from repairs is seen as posing a vote for, against or abstaining to a query then, in this semantics, some repairs are “voting” for a query (i.e., the query is entailed) and the rest of the repairs are not against (i.e., the query body atoms together with the atoms in the repair are consistent with the terminology) then the query is accepted without any objection. In addition, two variants of non-objection inference based on a selection of repairs (that can be against a query) are also considered.
  - *IJCAI’16 [31]*
- We provided a dialectical characterization of the Brave and IAR semantics. We proposed an argumentation dialogue system that considers a turn taking game between a proponent and an opponent. We defined the concept of participant’s profile and depending on these profiles we were able to give necessary and sufficient conditions for the Brave and IAR semantics. We further proposed a new TPI-like dialectical proof theory (a procedure where two players exchange arguments (moves) until one of them cannot play) for universal acceptance (i.e.,AR semantics). We limit the scope of the work to finite and coherent logic-based argumentation frameworks that correspond to the OBDA instantiation we consider in practical applications.
  - *ECAI’16 [18]; FLAIRS’16 [19]*
- We proposed a unifying framework for inconsistency-tolerant query answering within existential rule setting. In this framework, an inconsistency-tolerant semantics is seen as a pair composed of a

modifier, which produces consistent subsets of the data, and an inference strategy, which evaluates queries on the selected subsets. We systematically compared the productivity and the complexity of the obtained semantics.

  – *KR'16 [22]; JELIA'16 [23]*

- We studied the relationships between our unifying repair framework and stable model semantics. In particular, we provided a generic encoding for most semantics defined in that framework using Answer Set Programming.
  – *SUM'16 [24]*

6.2.2. Practical Applicability of Inconsistency-Tolerant Semantics and Argumentation

- Several inconsistency-tolerant semantics have been introduced for querying inconsistent knowledge bases. In order for users to be able to understand the query results, it is crucial to be able to explain why a tuple is a (non-)answer to a query under such semantics. We defined explanations for positive and negative answers under the brave, AR and IAR semantics. We then studied the computational properties of explanations in the lightweight description logic DL-Lite$_R$. For each type of explanation, we analyzed the data complexity of recognizing (preferred) explanations and deciding if a given assertion is relevant or necessary. We established tight connections between intractable explanation problems and variants of propositional satisfiability (SAT), enabling us to generate explanations by exploiting solvers for Boolean satisfaction and optimization problems. Finally, we empirically studied the efficiency of our explanation framework using the well-established LUBM benchmark.
  – *AAAI'16 [25]*
- We considered the problem of query-driven repairing of inconsistent DL-Lite knowledge bases: query answers are computed under inconsistency-tolerant semantics, and the user provides feedback about which answers are erroneous or missing. The aim is to find a set of data modifications (deletions and additions), called a repair plan, that addresses as many of the defects as possible. After formalizing this problem and introducing different notions of optimality, we investigated the computational complexity of reasoning about optimal repair plans and proposed interactive algorithms for computing such plans. For deletion-only repair plans, we also presented a prototype implementation of the core components of the algorithm.
  – *IJCAI'16 [26]*
- Based on the equivalent use of inconsistency-tolerant semantics for OBDA and logical instantiation of argumentation with existential rules, we highlighted some of the practical advantages that come from the interplay of the two techniques. More generally, we focussed on the generic problem of dealing with the uncertain knowledge (elicitation, representation and reasoning) involved at different levels of the food chain that model complex processes relying on numerous criteria, using various granularity of knowledge, most often inconsistent (due to the fact that complementary points of view can be expressed).
  – *IPMU'16 [32]*

  Beside, regarding the various granularity of knowledge, inspired from a hierarchical graph-based definition, we introduced the possibility of representing hierarchical knowledge using existential rules.
  – *ICCS'16 [33]*
- Agent technology and notably argumentation can optimise food supply chain operation in presence of inconsistency by employing intelligent agent applications (as shown in supply chain management case) but also facilitate reasoning with incomplete, inconsistent and missing knowledge as shown in the results presented in the previous sections. We considered two main methods of handling inconsistency: repair-based techniques and argumentation techniques. We demonstrated how to benefit from structured argumentation frameworks in practice by means of their implementations. Such implementations provide reasoning capabilities under inconsistency-tolerant semantics by means of a workflow that will enable Datalog frameworks to handle inconsistencies in knowledge bases using existing structured argumentation implementations.

- *COMMA'16 [46]*
- We provided a first implementation of the explanation based techniques using argumentation that can be used for inconsistent tolerant semantics. Such implementation served as a proof of concept of the usefulness of the interplay of the two techniques.
  - *COMMA'16 [17]*

  Furthermore, we provided an existential rule benchmark inspired from a real practical setting in the DURDUR project.
  - *MTSR'16 [16]*

  To refine this approach, we presented a generic framework of capturing reasoning errors by the interplay of strict logical rules and associative rules in knowledge bases (with the latter being elicited using a game with a purpose). We showed that such model can capture certain reasoning biases and could be eventually used as a predictive model for interacting with domain experts. We also showed empirically the difference of associations agronomy experts exhibit with respect to a random control population validated in the context of the DURDUR ANR project.
  - *ECAI'16 [18]; ICCS'16 [20]*

6.2.3. Decision Support in Agronomy

- We addressed a crucial problem for decision-making tools that are using inconsistency-handling methods (either argumentation frameworks or inconsistency-tolerant semantics) and namely the existence of multiple extensions / repairs. We placed ourselves in an applicative scenario, in the Pack4Fresh project, that investigates the best packaging for strawberries. We showed that being given a set of preferences on the initial set of facts in the existential rule knowledge base we can output meaningful (i.e., agrifood chain expert validated) extensions / repairs that will assist the decision maker.
  - *MTSR'16 [45]*
- We proposed a novel approach for decision-making that allows not only to handle symbolic data but also handle numerical RDF datasets. To deal with the numerical data, a preprocessing step is applied to convert numerical data into symbolic data. Based on the obtained symbolic classes we discover keys that are valid in this preprocessed data. We tested this approach on a dataset that describes wines with the set of numerical values representing different chemical components that give the flavour of wines. In this application setting, the discovered keys can be used to discover flavour complementarity, unknown from the experts, that allow to distinguish various wine sorts amongst themselves. We then validated the keys obtained with domain experts and discussed their interest with respect to the statistical analysis.
  - *ICCS'16 [43]*
- We presented a decision support system (DSS) which permits to compare, in a multi-criteria approach, innovative biomass transformation processes for biorefinery. Considered criteria are process extraction rate and green indicators. This DSS implements a pipeline which permits to annotate in a RDF knowledge heterogeneous textual data sources using a OWL/SKOS termino-ontological resource, to assess data source reliabilty and to compute several green indicators taking into account data reliability.
  - *CEA [13]; FUSS-IEEE'16 [37]*

6.3. Quality and interoperability of large document catalogues

Participants: Michel Chein, Madalina Croitoru, Alain Gutierrez, Michel Leclère, Clément Sipieter.

The work in this research line mainly takes place in the ANR project Qualinca (see Section 8.1), devoted to methods and tools to repair linkage errors in bibliographical databases. Within this project, we specially work with our applicative partner ABES (French Agency for Academic Libraries, http://www.abes.fr/). ABES manages several catalogues and authority bases, in particular the Sudoc, the collective catalogue of French academic libraries. ABES also provides services to libraries and end-users, as well as to other catalogue managers (e.g., OCLC for Worldcat and, in France, Adonis for the Isidore platform).

6.3.1. Evaluating the Quality of a Bibliographic Database

This year, we have focused on the specification, development and test of the application allowing to evaluate reference quality in a bibliographic database. The goal is to evaluate "same-as" links between contextual references (references to named entities provided in the context of a bibliographic notice) and authority references (references establishing an identifier for a given named entity). Our approach to solve this problem consists in two successive steps:

1. use the linkage API developed last years to compute automatically weighted links between contextual references and authority references;
2. compare those weighted links with those present in the bibliographic database in order to produce an evaluation of those links quality.

The evaluation output considers 12 different cases split in 5 major link categories: valid, almost valid, erroneous, missing, doubtful. For the 3 latter categories, we can often provide a correction or completion proposal.

We have initially implemented this application as a standalone client written in Java (see Section 5.1). We have tested it on a benchmark comprising 550 links, for which the evaluation has been done by experts. Our application has obtained very good results, since more than 70% of the links are evaluated correctly, less than 1% wrongly, and the rest consists of links for which data is insufficient to provide an evaluation.

To allow professionals from ABES to use this application, we have developed an interactive web service: the user first asks for the evaluation output on the set of links induced by a subset of contextual references; then he can validate or invalidate the proposed correction/completion. The tool can be restarted after each correction/completion to improve the evaluation with this new data. Our ABES partner is currently developing an enhanced graphical interface for Sudoc users, that will communicate with that web service, in order to use the software in production conditions.

Finally, an evaluation of the time required by our application led to numerous optimizations. We have for now concluded that the time is essentially spent by the library functions computing similarities between attributes. We consider now using map/reduce techniques to parallelize those computations.

6.3.2. Argumentation for Quality Evaluation

Beside, we studied the use of the `owl:sameAs` property (expressing that two URIs actually refer to the same thing) in practice. Many existing identity links do not reflect genuine real identity and therefore might lead to inconsistencies. We formalized explanation dialogues that use argument-based explanation based on inconsistency-tolerant semantics, and showed how to support a domain expert in discovering inconsistencies due to erroneous `SameAs` links. We implemented a prototype of the explanation dialogue that communicates with our tool Graal and provided an example of `sameAs` invalidation over real data explaining what has been obtained while running dialogues and how such results might benefit domain experts.

- *SUM'16 [21], ECAI'16 [18]*

LACODAM Team

7. New Results

7.1. Introduction

In this section, we organize our contributions over three main research topics:

- Mining different kinds of patterns, from 7.2 to 7.9
- Data mining and decision support with ASP, from 7.10 to 7.13
- Model-based diagnosis, from 7.13 to 7.15 .

7.2. Customer Purchase Signatures: a New Model in Grocery Retail Context

Participants: Clément Gautrais, Peggy Cellier [Lis], Thomas Guyet, René Quiniou, Alexandre Termier.

In the retail context, there is an increasing need for understanding individual customer behavior in order to personalize marketing actions. We propose the novel concept of customer signature, that identifies a set of important products that the customer refills regularly. Both the set of products and the refilling time periods give new insights on the customer behavior. Our approach is inspired by methods from the domain of sequence segmentation, thus benefiting from efficient exact and approximate algorithms. Experiments on a real massive retail dataset show the interest of the signatures for understanding individual customers (under submission to PAKDD 2017 conference).

This new model is used to detect and explain customer defection in a grocery retail context from the evolution of each customer basket content. It therefore provides actionable knowledge for the retailer at an individual scale. In addition, this model is able to identify customers that are likely to defect in the future months [16].

7.3. Discriminant Chronicles for Care Pathway Analysis

Participants: Yann Dauxais, Thomas Guyet, David Gross-Amblard [Druid], André Happe [Brest University Hospital].

A care pathway is a sequence of events (drugs deliveries, hospitalisation, etc) extracted from medical databases (see section 4.3 for details). In some studies, each patient is labeled by a class (*e.g.* died or not died). This information can be taken into account for the discriminant analysis of care pathways. This year, our objective was to extract discriminant patterns from a dataset of care pathways that can discriminate patients on their labels. To this end we introduced the new task of discriminant chronicle mining. Conceptually, a chronicle is a graph whose vertices are events and edges represent quantitative time constraints between events. We also proposed DCM, an algorithm dedicated to discriminant chronicles mining. This algorithm is based on rule learning methods to extract the temporal constraints. Computational performances and discriminant power of extracted chronicles are evaluated on artificial and real data.

The paper describing this work has been accepted in the french national conference on data mining (EGC 2017) [4] and is nominated for the best paper award.

7.4. Identifying Genetic Variant Combinations using Skypatterns

Participants: Alexandre Termier, Hoang-Son Pham [Genscale], Dominique Lavenier [Genscale].

Identifying variant combination association with disease is a bioinformatics challenge. This problem can be solved by discriminative pattern mining that uses a statistical function to evaluate the significance of individual biological patterns. There is a wide range of such measures. However, selecting an appropriate measure as well as a suitable threshold in some specific practical situations is a difficult task. In this work, we propose to use the skypattern technique which enables using combinations of measures to evaluate the importance of variant combinations without having to select a given measure and a fixed threshold (Pareto frontier). Experiments on several real variant datasets demonstrates that the skypattern method effectively identifies the risk variant combinations related to diseases [13].

7.5. Steady Patterns

Participants: Alexandre Termier, Willy Ugarte [UGA Grenoble], Miguel Santana [STMicroelectronics].

Skypatterns are an elegant answer to the pattern explosion issue, when a set of measures can be provided. Skypatterns for all possible measure combinations can be explored thanks to recent work on the skypattern cube. However, this leads to too many skypatterns, where it is difficult to quickly identify which ones are more important. First, we introduce a new notion of pattern steadiness [14] which measures the conservation of the skypattern property across the skypattern cube, allowing to see which are the “most universal” skypatterns. Then, we extended this notion to partitions of the dataset, and show in our experiments that this both allows to discover especially stable skypatterns, and identify interesting differences between the partitions.

7.6. Dense Bag-of-Temporal-SIFT-Words for Time Series Classification

Participants: Adeline Bailly [IRISA/Obelix], Laetitia Chapel [IRISA/Obelix], Thomas Guyet, Simon Malinowski [LinkMedia], Romain Tavenard [IRISA/Obelix].

The SIFT framework has shown to be effective in the image classification context. In [15], we designed a Bag-of-Words approach based on an adaptation of this framework to time series classification. It relies on two steps: SIFT-based features are first extracted and quantized into words; histograms of occurrences of each word are then fed into a classifier. In this work, we investigate techniques to improve the performance of Bag-of-Temporal-SIFT-Words: dense extraction of keypoints and different normalizations of Bag-of-Words histograms. Extensive experiments show that our method significantly outperforms nearly all tested standalone baseline classifiers on UCR datasets.

7.7. Comparing Symbolic and Statistical Classifiers on Energy Consumption Data

Participant: Benjamin Négrevergne.

During his Inria Carnot postdoc, Benjamin Négrevergne aimed at testing various data mining and machine learning methods on energy consumption data from the Energiency startup. Two symbolic methods developed in Lacodam were evaluated: QTempIntMiner and discriminant chronicle mining. While QTempIntMiner was shown to be ill-adapted in this setting, discriminant chronicle mining gave promising results. These results were evaluated in collaboration with our industrial partner. We also shown the interest of other methods: Hidden Markov Models and Gaussian processes. An internal report has been written to relate the results.

7.8. Detecting Strategic Moves in HearthStone Matches

Participants: Boris Doux [M1 intern], Clément Gautrais, Benjamin Negrevergne.

In this work, we demonstrate how to extract strategic knowledge from gaming data collected among players of the popular video game HearthStone. Our methodology is as follows. First we train a series of classifiers to predict the outcome of the game during a match, then we demonstrate how to spot key strategic events by tracking sudden changes in the classifier prediction. This methodology is applied to a large collection of HeathStone matches that we have collected from top ranked European players. Expert analysis shows that the events identified with this approach are both important and easy to interpret with the corresponding data [12].

7.9. Towards Visualizing Hidden Structures

Participants: Rémy Dautriche [STMicroelectronics], Alexandre Termier, Renaud Blanch [UGA Grenoble], Miguel Santana [STMicroelectronics].

There is an increasing need to quickly understand the contents of log data. A wide range of patterns can be computed and provide valuable information: for example existence of repeated sequences of events or periodic behaviors. However pattern mining techniques often produce many patterns that have to be examined one by one, which is time consuming for experts. On the other hand, visualization techniques are easier to understand, but cannot provide the in-depth understanding provided by pattern mining approaches. Our contribution is to propose a novel visual analytics method that allows to immediately visualize hidden structures such as repeated sets/sequences and periodicity, allowing to quickly gain a deep understanding of the log [3].

7.10. Knowledge-based Sequence Mining with ASP

Participants: Thomas Guyet, René Quiniou, Torsten Schaub.

We have introduced a framework for knowledge-based sequence mining, based on Answer Set Programming (ASP) [10], [5]. We begin by modeling the basic task and refine it in the sequel in several ways. First, we show how easily condensed patterns can be extracted by modular extensions of the basic approach. Second, we illustrate how ASP's preference handling capacities can be exploited for mining patterns of interest. In doing so, we demonstrate the ease of incorporating knowledge into the ASP-based mining process. To assess the trade-off in effectiveness, we provide an empirical study comparing our approach with a related sequence mining mechanism.

7.11. Packing Graphs with ASP for Landscape Simulation

Participants: Thomas Guyet, Yves Moinard, Jacques Nicolas [Dyliss], René Quiniou.

This work [6] describes an application of Answer Set Programming (ASP) to crop allocation for generating realistic landscapes. The task is to optimally cover a bare landscape, represented by its plot graph, with spatial patterns describing local arrangements of crops. This problem belongs to the hard class of graph packing problems and is modeled in the framework of ASP. The approach provides a compact and elegant solution to the basic problem and at the same time allows extensions such as a flexible integration of expert knowledge. Particular attention is paid to the treatment of symmetries, especially due to sub-graph isomorphism issues. Experiments were conducted on a database of simulated and real landscapes. Currently, the approach can process graphs of medium size, a size that enables studies on real agricultural practices.

7.12. Care Pathway Analysis with ASP Sequence Mining

Participants: Ahmed Samet, Benjamin Négrevergne, Thomas Guyet.

This line of work aims at applying our ASP encoding for sequential pattern mining to care pathway analysis (see section 4.3 for applicative objectives). This year, we proposed an approach of meaningful rare sequential pattern mining based on the declarative programming paradigm of Answer Set Programming (ASP). The setting of rare sequential pattern mining is introduced. To cope with the huge amount of meaningless rare patterns, our ASP approach provides an easy manner to encode expert constraints on expected patterns. Encodings are presented and quantitatively compared to a procedural baseline. An application on care pathways analysis illustrates the qualitative interest of expert constraints encoding.

This work has been submitted to the PAKDD 2017 conference.

7.13. ASP and Diagnosis

Participants: Christine Largouët, Laurence Rozé.

A new approach for performing diagnosis with ASP has been explored. The system is described by automata and implemented in an ASP program whose task is to find trajectories compatible with observations. The experimentation is carried out on benchmarks already used for the diagnosis problem using SAT. These benchmarks consider different levels of difficulty and number of faults (from one to twenty) and three types of observations: timestamped observations, totally ordered observations and partially ordered observations. The results were good both for dated and for totally ordered sequences of observations, whereas work needs to be still improved for the partial ordered observation case.

7.14. Searching for Cost-Optimized Strategies. Application to Temporal Planning and Agricultural System

Participants: Christine Largouët, Marie-Odile Cordier.

We consider a system modeled as a set of interacting components evolving along time according to explicit timing constraints. The decision making problem consists in selecting and organizing actions in order to reach a goal state in a limited time and in an optimal manner, assuming actions have a cost. We propose to reformulate the planning problem in terms of model-checking and controller synthesis such that the state to reach is expressed using a temporal logic. We have chosen to represent each agent using the formalism of Priced Timed Game Automata (PTGA) and a set of knowledge. PTGA is an extension of Timed Automata that allows the representation of cost on actions and the definition of a goal (to reach or to avoid). A first paper describes two algorithms designed to address the planning problem on a network of agents and proposes a practical implementation using model-checking tools that shows promising results on an agricultural application: a grassland based dairy production system [9]. Another paper describes the expressivity of this approach on the classical Transport Domain which is extended in order to include timing constraints, cost values and uncontrollable actions. This work has been implemented and performances evaluated on benchmarks [8].

7.15. Integrating Socio-Economic Drivers in an Explicit-Time, Qualitative Fisheries Model

Participant: Christine Largouët.

EcoMata is an explicit-time, qualitative modelling tool for assessing the ecosystem impacts of fishing and evaluating options for fishery management. The model is being developed further by integrating simple socio-economic drivers in the fishery system. Specifically, we have introduced a new module of automata that describes the profits associated to a specific fishing intensity and specific timing. This new module allows the evaluation of management strategies that are economically viable. The approach is illustrated on a coral-reef fishery in the Pacific that has been the focus of previous modelling work. [7].

LINKS Project-Team

7. New Results

7.1. Querying Heterogeneous Linked Data

7.1.1. Provenance

The computation of the provenance of a query answer is a classical problem in database theory. It consists in aggregating the impact of tuples of a database to a query answer. This allows to give an explanation of the query answers, that can help to judge their reliability. The computation of the provenance of a query answer is thus an aggregation problem as studied by the ANR project *Aggreg* .

P. Bourhis [20] showed at **PODS** — the top conference on database theory — that the lineage of MSO queries on treelike database instances is tractable, but not on other instances. This work was in cooperation with Telecom ParisTech and ENS Paris. As a first application, he can show that MSO query evaluation on probabilistic databases is tractable for tree like database instances, but not otherwise.

P. Bourhis applied in cooperation with Tel Aviv, provenance problems to recommendation systems. This allows to explain the end result by summarising with similar data without changing significantly results obtained in general by aggregation on the data. The corresponding tool was demonstrated at **EDBT** [32].

7.1.2. Certain Query Answering and Access Control

The problem of certain query answering consists in finding which are the certain answer of a query in a database with incomplete data, and a set of constraints representing available the knowledge on the incomplete data.

P. Bourhis [24] presented at **LICS** — the top conference in logic in computer science — a general framework for querying databases with visible and invisible relations. This work was done in cooperation with Oxford, Santa Cruez, and Bordeaux. His framework is motivated by the problem of access control for relational databases, i.e. of data leakage in relational views, but generalizes at the same time the problem of certain query answering. Invisible relations are subject to the open world assumption possibly under constraints as usual in certain query answering, while visible relations are subject to the closed world assumption. Bourhis then show that it is decidable, whether a conjunctive has an answer in this framework, when given the visible relation, the constraints, and the query as inputs. He also studies the complexity of this problem. It turns out the complexity increases from polynomial to doubly exponential, compared to certain query answering, since adding visible relations subject to the closed world assumption.

P. Bourhis studied at **IJCAI** [19] certain query answering with some transitive closure constraints, which allow to define a constraints with recursion. This work was done in collaboration with Oxford and Telecom ParisTech.

The problem of ontological query containment consists in establishing whether the certain answers of two queries subject to an ontology are included in each other. P. Bourhis [26] studied at **KR** this problem for several closely related formalisms: monadic disjunctive Datalog (MDDLog), MMSNP (a logical generalization of constraint satisfaction problems) and ontology-mediated queries (OMQs). This work was done in cooperation with Bremen.

7.1.3. Recursive Queries

At **LICS** [21] again, P. Bourhis showed in collaboration with Oxford how to lift a major restriction on decidable fixpoint logics that can define recursive queries (such as C2RPQs), specifically on guarded logic. This allows to improve significantly expressiveness of decidable fixpoint logics.

A. Lemay contributed at **TKDE** [14] the *gMark* benchmark, a tool to generate large size graph database and an associated set of queries. This work was done in cooperation with Eindhoven and previous members of Links that are now in Lyon and Clérmont-Ferrant. The tool was also demonstrated at **VLDB** [13]. Its main interest is a great flexibility (the generation of the graph can be done from a simple schema, but can also incorporate elaborate a parameters), an ability to generate recursive queries, and the possibility to generate large sets of queries of a desired selectivity. This benchmark allowed for instance to highlight difficulties for the existing query engines to deal with recursive queries of high selectivity.

7.1.4. Data Integration

P. Bourhis and S. Staworko in cooperation with Bordeaux and Oxford presented at **TODS** [17] their work on bounded repairability for regular tree languages, which is a study on whether a tree document (typically XML) can be repaired to fit a given target tree language within a bounded amount of tree editing operations. The article studies the complexity of different classes of tree languages such as non-recursive DTDs, recursive DTDs, or languages by arbitrary bottom-up tree automaton.

J.M. Lozano started his PhD project under the supervision of I. Boneva and S. Staworko. His topic subscribes the ANR project *Datacert* on data integration and certification.

7.1.5. Schema Validation

A. Boiret, V. Hugot and J. Niehren studied schemas for JSON documents in **Information and Computation** [15]. This work was done in collaboration with Paris 7. A JSON document is an unordered data trees, so schemas for such documents are best seen as automata for unordered data trees. The paper generalizes several previous formalisms for automata on unordered trees in a uniform framework. Whether the equivalence of two schemas can be tested in P-time is studied for various instances of the framework.

This work subscribes to the ANR project *Colis* where unranked data trees are used as models of linux file systems. In this context, N. Bacquey started his postdoc on the verification of linux installation scripts.

7.2. Managing Dynamic Linked Data

7.2.1. Complex Event Processing

Complex event processing can be seen as the problem is to answer queries on data graphs, for graphs that arrive on streams. These queries may contain aggregates, so this work subscribes to the ANR project *Aggreg*.

In his PhD thesis, T. Sebastian [12] developed with his supervisor J. Niehren streaming algorithms covering all of XPath 3.0 queries on `XML` streams. For this, they proposed a higher-order query language λXP, showed how to give a formal semantics of all of XPath 3.0 by compilation to λXP, and then how to evaluate λXP queries on `XML` streams. These algorithms were implemented in the QuiXPath tool.

At **SOFSEM**, they proposed a new technique to speed up the evaluation of navigational XPath queries on XML streams based on document projection. The idea is to skip those parts of the stream that are irrelevant for the query. This speeds up the evaluation of navigation XPath queries by a factor of 4 in usual Xpath benchmarks.

M. Sakho started his PhD project on hyperstreaming query answering algorithms for graphs under the supervision of J. Niehren and I. Boneva. Part of this work will be continued with out visitor D. Vrgoc from Santiago di Chili.

7.2.2. Data Centric Workflows

Data-centric workflows are complex programs that can query and update a database. The usage of data-centric workflows for crowd sourcing is the topic of the ANR Project *HeadWork*.

In collaboration with ENS Cachan and San Diego, P. Bourhis presented at **ICDT** [18] techniques on collaborative access control in a distributed query and data exchange language (Webdamlog). The goal of this work was to provide a semantic to data exchange rules defined by Webdamlog. It also allowed to prove that it is possible to formally verify whether there are data leakages.

P. Bourhis with Tel Aviv defined at **ICDE** [25] a notion of provenance for data-centric workflows, and proved that it can be used to explain the provenance of fact in the final instance of an execution. This provenance is used to answer three main questions: *why* does a specific tuple appear in the answer of a query, *what if* the initial database is changed (Revision problem), and *how to* change the query to obtain a missing tuple.

7.3. Linking Data Graphs

7.3.1. Learning Transformations

We consider the problem to learn queries and query-based transformations on semi-structured data from examples.

A. Boiret obtained his PhD for his work on the "Normalization and Learning of Transducers on Trees and Words" under the supervision of J. Niehren and A. Lemay. In this year, he showed how to learn top-down tree transformations with regular schema restrictions [31], [33], [34]. At **LATA** [22], he deepened a result of a previous PhD student of Links on learning sequential tree-to-word transducers (with output concatenation), by showing who to find normal forms for less restrictive linear tree-to-word transducers. At **DLT** [23], he could show in cooperation with Munich, that the equivalence problem of this class of transducers is in polynomial time, even though their normal forms may be of exponential size.

In the context of learning RDF graph transformations, S. Staworko presented a cooperation with Edinburg at **VLDB** [27]. Using bisimulation technique, he aims at aligning datas of two RDF Graphs that takes into account blank velues, changes in ontology and small differences in data values and in the structure of the graph. the alignement of graphs is an important first step for the inference of transformations.

7.3.2. Learning Join Queries

S. Staworko published in **TODS** an article [16] on learning join queries from user examples in collaboration with Universities of Lyon and Clermont-Ferrand that present techniques that allow the automatic construction of a join query through interaction with a user that simply labels sets of tuples to indicate whether the tuple is in the target query or not.

MAGNET Project-Team

7. New Results

7.1. Decentralized and Private Learning

In [13], we address the problem of decentralized minimization of pairwise functions of the data points, where these points are distributed over the nodes of a graph defining the communication topology of the network. This general problem finds applications in ranking, distance metric learning and graph inference, among others. We propose new gossip algorithms based on dual averaging which aims at solving such problems both in synchronous and asynchronous settings. The proposed framework is flexible enough to deal with constrained and regularized variants of the optimization problem. Our theoretical analysis reveals that the proposed algorithms preserve the convergence rate of centralized dual averaging up to an additive bias term. We present numerical simulations on Area Under the ROC Curve (AUC) maximization and metric learning problems which illustrate the practical interest of our approach.

In [19], we consider a set of learning agents in a collaborative peer-to-peer network, where each agent learns a *personalized model* according to its own learning objective. The question addressed in this paper is: how can agents improve upon their locally trained model by communicating with other agents that have similar objectives? We introduce and analyze two asynchronous gossip algorithms running in a fully decentralized manner. Our first approach, inspired from label propagation, aims to smooth pre-trained local models over the network while accounting for the confidence that each agent has in its initial model. In our second approach, agents jointly learn and propagate their model by making iterative updates based on both their local dataset and the behavior of their neighbors. Our algorithm for solving this challenging optimization problem relies on the Alternating Direction Method for Multipliers (ADMM).

In [20], we propose a decentralized protocol for a large set of users to privately compute averages over their joint data, which can later be used to learn more complex models. Our protocol can find a solution of arbitrary accuracy, does not rely on a trusted third party and preserves the privacy of users throughout the execution in both the honest-but-curious and malicious adversary models. 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.

7.2. Natural Language Processing

In [12], we introduce a simple semi-supervised approach to improve implicit discourse relation identification. This approach harnesses large amounts of automatically extracted discourse connectives along with their arguments to construct new distributional word representations. Specifically, we represent words in the space of discourse connectives as a way to directly encode their rhetorical function. Experiments on the Penn Discourse Treebank demonstrate the effectiveness of these task-tailored representations in predicting implicit discourse relations. Our results indeed show that, despite their simplicity, these connective-based representations outperform various off-the-shelf word embeddings, and achieve state-of-the-art performance on this problem.

Along the PhD thesis of THIBAULT LIÉTARD, we are working on learning a similarity between text entities for the task of coreference resolution. Unlike indirect classification criteria often used in the literature, the similarity function naturally operates on pairs of mentions and several relevant objectives can be considered. For instance, we can learn the parameters of the similarity function such that the similarity of a given mention to its closest antecedent coreferent mention is larger than to any closer non-coreferential antecedent candidate. The resulting similarity scores can then be plugged into a greedy clustering procedure, or used to build a weighted graph of mentions to be clustered by spectral algorithms. For the representations of (pairs of) mentions on which the similarity function is learned, we consider both traditional linguistic features as well as external information about the general context of occurrence of the mentions using word embeddings.

Along the PhD thesis of MATHIEU DEHOUCK, we study the problem of cross-lingual dependency parsing, aiming at leveraging training data from different source languages to learn a parser in a target language. Specifically, this approach first constructs word vector representations that exploit structural (i.e., dependency-based) contexts but only considering the morpho-syntactic information associated with each word and its contexts. These delexicalized word embeddings, which can be trained on any set of languages and capture features shared accross languages are then used in combination with standard language-specific features to train a lexicalized parser in the target language. We evaluate our approach through experiments on a set of eight different languages that are part the Universal Dependencies Project. Our main results show that using such embeddings (monolingual or multilingual) achieves significant improvements over monolingual baselines. The work is submitted.

7.3. Edge Prediction in Networks

In [18] we address the problem of classifying the links of signed social networks given their full structural topology. In the problem of edge sign prediction, we are given a directed graph (representing a social network), and our task is to predict the binary labels of the edges (i.e., the positive or negative nature of the social relationships). Many successful heuristics for this problem are based on the troll-trust features, estimating at each node the fraction of outgoing and incoming positive/negative edges. We show that these heuristics can be understood, and rigorously analyzed, as approximators to the Bayes optimal classifier for a simple probabilistic model of the edge labels. We then show that the maximum likelihood estimator for this model approximately corresponds to the predictions of a label propagation algorithm run on a transformed version of the original social graph. Extensive experiments on a number of real-world datasets show that this algorithm is competitive against state-of-the-art classifiers in terms of both accuracy and scalability. Finally, we show that troll-trust features can also be used to derive online learning algorithms which have theoretical guarantees even when edges are adversarially labeled.

In [16], we address the problem of predicting connections between a set of data points. We focus on the *graph reconstruction* problem, where the prediction rule is obtained by minimizing the average error over all $n(n-1)/2$ possible pairs of the n nodes of a training graph. Our first contribution is to derive learning rates of order $O(\log n/n)$ for this problem, significantly improving upon the slow rates of order $O(1/\sqrt{n})$ established in the seminal work of [27]. Strikingly, these fast rates are universal, in contrast to similar results known for other statistical learning problems (e.g., classification, density level set estimation, ranking, clustering) which require strong assumptions on the distribution of the data. Motivated by applications to large graphs, our second contribution deals with the computational complexity of graph reconstruction. Specifically, we investigate to which extent the learning rates can be preserved when replacing the empirical reconstruction risk by a computationally cheaper Monte-Carlo version, obtained by sampling with replacement $B \ll n^2$ pairs of nodes. Finally, we illustrate our theoretical results by numerical experiments on synthetic and real graphs.

7.4. Mining Geotagged Social Data

Data generated on location-based social networks provide rich information on the whereabouts of urban dwellers. Specifically, such data reveal who spends time where, when, and on what type of activity (e.g., shopping at a mall, or dining at a restaurant). That information can, in turn, be used to describe city regions in terms of activity that takes place therein. For example, the data might reveal that citizens visit one region mainly for shopping in the morning, while another for dining in the evening. Furthermore, once such a description is available, one can ask more elaborate questions. For example, one might ask what features distinguish one region from another – some regions might be different in terms of the type of venues they host and others in terms of the visitors they attract. As another example, one might ask which regions are similar across cities. In [11], we present a method to answer such questions using publicly shared Foursquare data. Our analysis makes use of a probabilistic model, the features of which include the exact location of activity, the users who participate in the activity, as well as the time of the day and day of week the activity takes place. Compared to previous approaches to similar tasks, our probabilistic modeling approach allows us to make minimal assumptions about the data – which relieves us from having to set arbitrary parameters in our analysis (e.g., regarding the granularity of discovered regions or the importance of different features). We demonstrate how the model learned with our method can be used to identify the most likely and distinctive features of a geographical area, quantify the importance features used in the model, and discover similar regions across different cities. Finally, we perform an empirical comparison with previous work and discuss insights obtained through our findings. Our results were also presented through an interactive demo at the 25th World Wide Web Conference [21].

7.5. Learning from Non-iid Data

In [14] we deal with the generalization ability of classifiers trained from non-iid evolutionary-related data in which all training and testing examples correspond to leaves of a phylogenetic tree. For the realizable case, we prove PAC-type upper and lower bounds based on symmetries and matchings in such trees.

In [9], we studied learning problems where the performance criterion consists of an average over tuples (e.g., pairs or triplets) of observations rather than over individual observations, as in many learning problems involving networked data (e.g., link prediction), but also in metric learning and ranking. In this setting, the empirical risk to be optimized takes the form of a U-statistic, and its terms are highly dependent and thus violate the classic i.i.d. assumption. From a computational perspective, the calculation of such statistics is highly expensive even for a moderate sample size n, as it requires averaging $O(n^d)$ terms. We show that, strikingly, such empirical risks can be replaced by drastically computationally simpler Monte-Carlo estimates based on $O(n)$ terms only, usually referred to as incomplete U-statistics, without damaging the $O(1/\sqrt{n})$ learning rate of Empirical Risk Minimization (ERM) procedures. For this purpose, we establish uniform deviation results describing the error made when approximating a U-process by its incomplete version under appropriate complexity assumptions. Extensions to model selection, fast rate situations and various sampling techniques are also considered , as well as an application to stochastic gradient descent for ERM. Finally, numerical examples are displayed in order to provide strong empirical evidence that the approach we promote largely surpasses more naive subsampling techniques.

7.6. Adaptive Graph Construction

The efficiency of graph-based semi-supervised algorithms depends on the graph of instances on which they are applied. The instances are often in a vectorial form before a graph linking them is built. The construction of the graph relies on a metric over the vectorial space that help define the weight of the connection between entities. The classic choice for this metric is usually a distance measure or a similarity measure based on the euclidean norm. We claim that in some cases the euclidean norm on the initial vectorial space might not be the more appropriate to solve the task efficiently. In the work [17], we propose an algorithm that aims at learning the most appropriate vectorial representation for building a graph on which the task at hand is solved efficiently. In addition to experimental results showing the interest of such an approach, we define initial conditions under which the graph-based classification is ensured to perform optimally.

ORPAILLEUR Project-Team

7. New Results

7.1. The Mining of Complex Data

Participants: Quentin Brabant, Miguel Couceiro, Adrien Coulet, Esther Galbrun, Nicolas Jay, Nyoman Juniarta, Florence Le Ber, Joël Legrand, Pierre Monnin, Amedeo Napoli, Justine Reynaud, Chedy Raïssi, Mohsen Sayed, My Thao Tang, Yannick Toussaint.

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

Pattern mining and Formal Concept Analysis are suitable symbolic methods for KDDK, that may be used for real-sized applications. Global improvements are carried out on the scope of applicability, the ease of use, the efficiency of the methods, and on the ability to fit evolving situations. Accordingly, the team is extending these symbolic data mining methods for working on complex data (e.g. textual documents, biological, chemical or medical data), involving objects with multi-valued attributes (e.g. domains or intervals), n-ary relations, sequences, trees and graphs.

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 [79]) 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 [90] 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 designed new information retrieval methods based on FCA [72], text classification and heterogeneous pattern structures [71], pattern structures for structured attribute sets [67], and also a quasi-polynomial algorithm for mining top patterns w.r.t. measures satisfying special properties in a FCA framework [70]. We developed also a whole line of work on pattern structures for the discovery of functional dependencies [33], text classification and heterogeneous pattern structures [71], and fuzzy FCA as well [31]. Finally, we also proposed new visualization techniques and tools able to display important and useful information (e.g. stable concepts) from large concept lattices [28].

7.1.2. Text Mining

The thesis work of My Thao Tang [11] proposes a process where software and humans agents cooperate in knowledge discovery from different source textual types for extending a knowledge base. One challenge is that, on the one hand, knowledge discovery methods (software agents) can be run in accordance with background knowledge (or expert knowledge), at any step of the KDD process. On the other hand, human agents should be able to correct or to extend the current knowledge base. FCA is used for discovering a "class schema" (or "representation model") within textual resources which can be either a set of attribute implications or a concept lattice. However, such a schema does not necessarily fit the point of view of a domain expert for different reasons, e.g. noise, errors or exceptions in the data. Thus, a bridge filling the possible gap between the representation model based on a concept lattice and the representation model of a domain expert is studied in [44]. The background knowledge is encoded as a set of attribute dependencies or constraints which is "aligned" with the set of implications associated with the concept lattice. Such an alignment may lead to modifications in the original concept lattice. This method can be generalized for generating lattices satisfying some constraints based on attribute dependencies in using the so-called "extensional projections". It also allows experts to keep a trace of the changes occurring in the original lattice and the revised version, and to assess how concepts in practice are related to concepts discovered in the data.

In the framework of the Hybride ANR project (see 8.2.1.1), Mohsen Sayed proposes an original machine learning approach for identifying in literature disease phenotypes that are not yet represented within existing ontologies. The process is based on graph patterns extracted from sentences represented as dependency graphs. Phenotypes are usually expressed by complex noun phrases while traditional gazetteers recognize them only partially. The strength of graph patterns is to preserve the linguistic component bounds and to enable the identification of the complete phenotype formulation. A specific publication is currently in preparation.

7.1.3. Mining Sequences and Trajectories

Nowadays data sets are available in very complex and heterogeneous ways. Mining of such data collections is essential to support many real-world applications ranging from healthcare to marketing. This year, we completed a research work on the analysis of “complex sequential data” by means of interesting sequential patterns [13]. We approach the problem using FCA and pattern structures, where the subsumption relation ordering patterns is defined w.r.t. the partial order on sequences.

7.1.4. 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 [89]. 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.

In recent work [40], we focused on the problem of pattern selection, developing a method for filtering a set of redescription to identify a non-redundant, interesting subset to present to the analyst. Also, we showcased the usability of redescription mining on an application in the political domain [50]. More specifically, we applied redescription mining to the exploratory analysis of the profiles and opinions of candidates to the parliamentary elections in Finland in 2011 and 2015.

We presented an introductory tutorial on redescription mining at ECML-PKDD in September 2016 to help foster the research on these techniques and widen their use (http://siren.mpi-inf.mpg.de/tutorial/main/).

7.1.5. E-sports analytics and subgroup discovery based on 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) [87] 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 [77]. 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. In [58], we argue that the greedy nature of most of these approaches produce pattern sets that lack of diversity. Consequently, 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 recent 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.

7.1.6. Data Privacy: Online link disclosure strategies for social networks

Online social networks are transforming our culture and world. While online social networks have become an important channel for social interactions, they also raise ethical and privacy issues. A well known fact is that social networks leak information, that may be sensitive, about users. However, performing accurate real world online privacy attacks in a reasonable time frame remains a challenging task. In [57], [26] (this work is done in cooperation with the Pesto Inria Team), we address the problem of rapidly disclosing many friendship links using only legitimate queries (i.e., queries and tools provided by the targeted social network). Our study sheds new light on the intrinsic relation between communities (usually represented as groups) and friendships between individuals. To develop an efficient attack we analyzed group distributions, densities and visibility parameters from a large sample of a social network. By effectively exploring the target group network, our proposed algorithm is able to perform friendship and mutual-friend attacks along a strategy that minimizes the number of queries. The results of attacks performed on a major social network profiles show that 5 different friendship links are disclosed in average for each single legitimate query in the best cases.

7.1.7. Aggregation

Aggregation theory is the study of 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 [82]. The need to aggregate objects in a meaningful way appeared naturally in classical topics such as mathematics, statistics, physics and computer science, but it became increasingly emergent in applied areas such as social and decision sciences, artificial intelligence and machine learning, biology and medicine.

We are working on a theoretical basis of a unified theory of consensus and to set up a general machinery for the choice and use of aggregation functions. This choice depends on properties specified by users or decision makers, the nature of the objects to aggregate as well as computational limitations due to prohibitive algorithmic complexity. This problem demands an exhaustive study of aggregation functions that requires an axiomatic treatment and classification of aggregation procedures as well as a deep understanding of their structural behavior. It also requires a representation formalism for knowledge, in our case decision rules, as well as methods for extracting them. Typical approaches include rough-set and FCA approaches, that we aim to extend in order to increase expressivity, applicability and readability of results. Direct applications of these efforts are expected in the context of two multidisciplinary projects, namely the “Fight Heart Failure” and the European H2020 “CrossCult” project.

In our recent work, 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, we provided axiomatizations of noteworthy classes of lattice-based aggregation functions, which were then used to model preferences and to provide their logical description [14]. In this qualitative setting, we also tackled the problem of computing version spaces (with explicit descriptions of all models compatible with a given dataset) and proved a dichotomy theorem showing that the problem is NP-complete for preferences over at least 4 attributes whereas it is solvable in polynomial time otherwise [61].

Finding consensual structures among different classifications or metrics is again a challenging task, especially, for large and multi-source data, and its importance becomes apparent since algorithmic approaches are often heuristic on such datasets and they rarely produce the same output. The difficulty in extracting such consensual structures is then to find appropriate and meaningful aggregation rules, and their impossibility is often revealed by Arrow type impossibility results. In the current year, we focused on median structures [19], [21] that include several relational structures (trees, graphs, lattices) and allow several consensus procedures.

7.2. Knowledge Discovery in Healthcare and Life Sciences

Participants: Miguel Couceiro, Adrien Coulet, Kévin Dalleau, Joël Legrand, Pierre Monnin, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, Malika Smaïl-Tabbone, Yannick Toussaint.

Life Sciences constitute a challenging domain for KDDK. Biological data are complex from many points of views, e.g. voluminous, high-dimensional and deeply inter-connected. Analyzing such data is a crucial issue in healthcare, environment and agronomy. Besides, many bio-ontologies are available and can be used to enhance the knowledge discovery process. Accordingly, the research work of the Orpailleur team in KDDK applied to Life Sciences is in concern with mining biological data, data integration, information retrieval, and use of bio-ontologies and linked data for resource annotation.

7.2.1. *Ontology-based Clustering of Biological Linked Open Data*

Increasing amounts of biomedical data provided as Linked Open Data (LOD) offer novel opportunities for knowledge discovery in bio-medicine. We proposed an approach for selecting, integrating, and mining LOD with the goal of discovering genes responsible for a disease [88]. We are currently working on the integration of LOD about known phenotypes and genes responsible for diseases along with relevant bio-ontologies. We are also defining a corpus-based semantic distance. One possible application of this work is to build and compare possible diseaseomes, i.e. global graphs representing all diseases connected according to their pairwise similarity values.

7.2.2. *Biological Data Aggregation for Knowledge Discovery*

Two multi-disciplinary projects were initiated in 2016, in collaboration with the Capsid Team, with a group of clinicians from the Regional University Hospital (CHU Nancy) and bio-statisticians from the Maths Lab (IECL). The first project is entitled ITM2P [0] and depends on the so-called CPER 2015–2020 framework. We are involved in the design of the SMEC platform as a support for “Simulation, Modeling and Knowledge Extraction from Bio-Medical Data”.

The second project is a RHU [0] project entitled *Fight Heart Failure* (FHF), where we are in charge of a workpackage about “data aggregation” mechanisms. Accordingly, we are working on the definition of multidimensional similarity measure for comparing and clustering sets of patients. Each cluster should correspond to a bioprofile, i.e. a subgroup of patients sharing the same form of the disease and thus the same diagnosis and care strategy.

The first results were presented at the International Symposium on Aggregation and Structures (ISAS 2016) [36]. We propose the GABS for “Graph Aggregation Based Similarity” approach for complex graph aggregation resulting in a similarity graph between a subset of nodes. Indeed the initial graph contains two types of nodes, i.e. individuals and attributes. The pairwise similarity between individuals is derived from the various paths in the initial graph. This setting allows the integration of domain knowledge in the initial graph (corresponding to domain ontologies, norms...). Another advantage of the GABS approach is to generate a similarity graph which can be used as input for various clustering algorithms (graph-based ones as well as those working on similarity/distance matrix).

The next question will be to build a prediction model for each bioprofile/subgroup (once validated by the clinicians) for a decision support system. Thus, we are investigating SRL (“Statistical Relational Learning”) methods which combine symbolic and probabilistic methods for improving expressivity (through logical or relational languages) and for dealing with uncertainty.

7.2.3. *Suggesting Valid Pharmacogenes by Mining Linked Open Data and Electronic Health Records*

A standard task in pharmacogenomics research is identifying genes that may be involved in drug response variability and called “pharmacogenes”. As genomic experiments in this domain tend to generate many false positives, computational approaches based on background knowledge may generate more valuable results. Until now, the later have only used molecular networks databases or biomedical literature. We are

[0]“Innovations Technologiques, Modélisation et Médecine Personnalisée”
[0]“Recherche Hospitalo-Universitaire”

studying a new method taking advantage of various linked data sources to validate uncertain drug-gene relationships, i.e. pharmacogenes [75]. One advantage relies on the standard implementation of linked data that facilitates the joint use of various sources and makes easier the consideration of features of various origins. Accordingly, we selected, formatted, interconnected and published an initial set of linked data sources relevant to pharmacogenomics. We applied numerical classification methods for extracting drug-gene pairs that can become validated pharmacogene candidates.

The ANR project "PractiKPharma" initiated in 2016 relies on similar ideas, having the motivation of validating state-of-the-art knowledge in pharmacogenomics (http://practikpharma.loria.fr/). The originality of "PractiKPharma" is to use Electronic Health Records (EHRs) to constitute cohorts of patients that can be mined for validating extracted pharmacogenomics knowledge units.

7.2.4. Analysis of biomedical data annotated with ontologies

In the context of the Snowflake Inria Associate team, Gabin Personeni, who is a PhD student co-supervised by Marie-Dominique Devignes (Capsid EPI) and Adrien Coulet (Orpailleur EPI) spent four months at the Stanford University in 2016. After this internship, 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 [51], [49]. In this case, 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 the first research work considering such representations to mine rules between frequently associated ADRs. We illustrated the generality of the approach on two distinct 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 and analyzed and may be used as a basis of a recommendation system.

7.3. Knowledge Engineering and Web of Data

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

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

7.3.1. Around the Taaable Research Project

The Taaable project was originally created as a challenger of the Computer Cooking Contest (ICCBR Conference) [73]. Beyond its participation to the CCC challenges, the Taaable project aims at federating various research themes: case-based reasoning (CBR), information retrieval, knowledge acquisition and extraction, knowledge representation, belief change theory, ontology engineering, semantic wikis, text-mining, etc. CBR performs adaptation of recipes w.r.t. user constraints. The reasoning process is based on a cooking domain ontology (especially hierarchies of classes) and adaptation rules. The knowledge base is encoded within a semantic wiki containing the recipes, the domain ontology and adaptation rules.

As acquiring knowledge from experts is costly, a new approach was proposed to allow a CBR system using partially reliable, non expert, knowledge from the web for reasoning. This approach is based on notions such as belief, trust, reputation and quality, as well as their relationships and rules to manage the knowledge reliability. The reliability estimation is used to filter knowledge with high reliability as well as to rank the results produced by the CBR system. Performing CBR with knowledge resulting from an e-community is improved by taking into account the knowledge reliability [10]. In the same way, another study shows how the case retrieval of a CBR system can be improved using typicality. Typicality discriminates subclasses of a class in the domain ontology depending of how a subclass is a good example for its class. An approach has been proposed to partition the subclasses of some classes into atypical, normal and typical subclasses in order to refine the domain ontology. The refined ontology allows a finer-grained generalization of the query during the retrieval process, improving at the same time the final results of the CBR system.

The Taaable system also includes a module for adapting textual preparations (from a source recipe text to an adapted recipe text, through a formal representation in the qualitative algebra INDU). The evaluation of this module as a whole thanks to users has been carried out and has shown its efficiency (w.r.t. text quality and recipe quality), when compared with another approach to textual adaptation [76].

FCA allows the classification of objects according to the properties they share into a concept lattice. A lattice has been built on a large set a cooking recipes according to the ingredients they use, producing a hierarchy of ingredient combinations. When a recipe R has to be adapted, this lattice can be used to search the best ingredient combinations in the concepts that are the closest to the concept representing R.

Minimal change theory and belief revision can be used as tools to support adaptation in CBR, i.e. the source case is modified to be consistent with the target problem using a revision operator. Belief revision was applied to Taaable to adjust the ingredient quantities using specific inference engines.

Another approach to adaptation based on the principles of analogical transfer applied to the formalism RDFS has been developed [41]. It is based on the problem-solution dependency represented as an RDFS graph: this dependency within the source case is modified so that it fits the context of the target problem. The application problem that has guided this research addresses the issue of cocktail name adaptation: given a cocktail recipe, the name of this cocktail and the ingredient substitution that produces a new cocktail, how could the new cocktail be called?

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, as, with the increased interest in machine processable data, more and more data is now published in RDF (Resource Description Framework) format. The popularization and 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. Particularly, we are interested in the completeness of the data and the their potential to provide concept definitions in terms of necessary and sufficient conditions [66]. We have proposed a novel technique based on Formal Concept Analysis which organizes subsets of RDF data into a concept lattice [43]. This allows data exploration as well as the discovery of implication rules which are used to automatically detect missing information and then to complete RDF data and to provide definitions. Moreover, this is also 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.

7.4. Advances in Graph Theory

Participants: Aurore Alcolei, Rémi de Joannis de Verclos, François Pirot, Jean-Sébastien Sereni.

Keywords: graph theory, graph coloring, extremal graph theory, chromatic number, two-mode data networks

Motivated by notions brought forward by sociology, we confirm a conjecture by Everett, Sinclair, and Dankelmann [Some Centrality results new and old, J. Math. Sociology 28 (2004), 215–227] regarding the problem of maximizing closeness centralization in two-mode data, where the number of data of each type is fixed. Intuitively, our result states that among all networks obtainable via two-mode data, the largest closeness is achieved by simply locally maximizing the closeness of a node. Mathematically, our study concerns bipartite graphs with fixed size bipartitions, and we show that the extremal configuration is a rooted tree of depth 2, where neighbors of the root have an equal or almost equal number of children [24].

Using recently introduced techniques based on entropy compression, we proved that the acyclic chromatic number of a graph with maximum degree Δ is less than $2.835\Delta^{4/3} + \Delta$. This improved the previous upper bound, which was $50\Delta^{4/3}$ (see [91] which is now published in Journal of Combinatorics, 7(4):725–737, 2016).

SMIS Project-Team

6. New Results

6.1. Embedded Data Management

Participants: Nicolas Anciaux, Saliha Lallali, Philippe Pucheral, Iulian Sandu Popa [correspondent].

Embedded keyword indexing: In this work, we revisit the traditional problem of information retrieval queries over large collections of files in an embedded context. A file can be any form of document, picture or data stream, associated with a set of terms. A query can be any form of keyword search using a ranking function (e.g., TF-IDF) identifying the top-k most relevant files. The proposed search engine can be used in sensors to search for relevant objects in their surroundings, in cameras to search pictures by using tags, in personal smart dongles to secure the querying of documents and files hosted in an untrusted Cloud, or in a personal cloud securely managed using a tamper resistant smart object. A search engine is usually based on a (large) inverted index and queries are traditionally evaluated by allocating one container in RAM per document to aggregate its score, making the RAM consumption linear with the size of the document corpus. To tackle this issue, we designed a new form of inverted index which can be accessed in a pure pipeline manner to evaluate search queries without materializing any intermediate result. Successive index partitions are written once in Flash and maintained in the background by timely triggering merge operations while files are inserted or deleted from the index. This work was initially published at VLDB'15 [5] and demonstrated at SIGMOD'15 [38]. It constitutes the main contribution of the PhD thesis of Saliha Lallali defended in January 2016. In 2016, we extended this work to demonstrate at EDBT'16 [22] its applicability to set up a secure distributed search engine for the Personal Cloud. We also complemented this work with (1) a thorough analysis of the RAM consumption linked to the main algorithms implementing the solution, (2) the support of conditional top-k queries in a personal Cloud context that we consider as a killer application domain today and (3) new performance measurements with a real dataset (ENRON), representative of this personal Cloud context. These new contributions have been submitted to Information Systems journal.

6.2. Secure Global Computing on Asymmetric Architecture

Participants: Benjamin Nguyen [correspondent], Axel Michel, Philippe Pucheral, Iulian Sandu Popa.

Asymmetric Architecture Computing: This research direction studies the secure execution of various algorithms on data stored in an unstructured network of Trusted Cells (i.e., personal trusted device) so that each user can keep control over her data. The data could be stored locally in a trusted cell or encrypted on some external cloud. Execution takes place on a specific infrastructure called the Asymmetric Architecture (AA): the network of trusted cells, supported by an untrusted cloud supporting IaaS or PaaS. Our objective is to show that many different algorithms and computing paradigms can be executed on AA, thus achieving secure and private computation. Our first contribution in this area was to study the execution of Privacy Preserving Data Publishing algorithms on such an architecture (T. Allard's PhD Thesis). Then we studied general SQL queries in this same execution context. We concentrated on the subset of SQL queries without joins, but including Group By and aggregates, and show how to secure their execution in the presence of honest-but-curious attackers. This work, named SQL-AA and notably published at EDBT'14 [8] and demonstrated at VLDB'15, was part of Quoc-Cuong To's Ph.D defended in 2015. We have extended this framework through a collaboration with INSA Centre Val de Loire, LIFO Lab and University of Paris Nord, LIPN lab and have shown in CoopIS'15 [9] that it is possible to achieve seamless integration of distributed MapReduce processing using trusted cells, while maintaining reasonable performance. In 2016, we added three novel contributions to SQL-AA: (i) an extended privacy analysis in which we consider stronger adversaries with more background knowledge, (ii) an extended threat model in which we consider malicious attacker and propose safety properties to prevent malicious attacks and (iii) we tackled practical issues like exchanging securely shared keys among trusted cells and Querier (GKE protocol) and enforcing access control at query execution time. These new contributions have been published in TODS'16 [15]. In parallel, we are starting a new study in the line of our previous work on Privacy Preserving Data Publishing (PPDP) with the objective to inject individualized privacy requirements in the PPDP protocol. A preliminary contribution has been published at BDA'16 [25] to compute SQL aggregate queries under k-anonymity constraints where each individual contributing to the query may define her own k constraint, thereby letting each one weighting differently the sensitiveness of a given piece of information according to her own situation.

Secure spatio-temporal distributed processing: Mobile participatory sensing could be used in many applications such as vehicular traffic monitoring, pollution tracking, or even health surveying (e.g., to allow measuring in real-time the individual exposure to environmental risk factors or the propagation of an epidemic). However, its success depends on finding a solution for querying a large number of users which protects user location privacy and works in real-time [10]. We addressed these issues and proposed 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, perform distributed query processing, while preventing users from accessing other users' data. Secure probes exchange data in encrypted form with help from an untrusted supporting server infrastructure. PAMPAS uses two efficient, parallel, and privacy-aware protocols for location-based aggregation and adaptive spatial partitioning of secure probes. Our experimental results and security analysis demonstrate that these protocols are able to collect, aggregate and share statistics or derived data in real-time, without any privacy leakage. This work is part of Dai Hai Ton That's Ph.D. thesis defended in January 2016, co-supervised by Iulian Sandu Popa. The system implementation was demonstrated in [41], while two papers describing the technical details of the system have been published in 2016 [23], [16].

6.3. Personal Cloud

Participants: Nicolas Anciaux [correspondent], Luc Bouganim, Julien Loudet, Benjamin Nguyen, Philippe Pucheral, Iulian Sandu Popa, Guillaume Scerri, Paul Tran Van.

We are witnessing an exponential increase in the acquisition of personal data about the individuals or produced by them. Today, this information is managed using Web applications, centralizing this data in cloud data servers, under the control of few Web majors [2]. However, it has now become clear that (1) centralizing millions of personal records exposes the data to very sophisticated attacks, linked to a very high potential benefit in case of success (millions of records being revealed), and (2) delegating the management of personal records without any tangible guarantee for the individuals leads to privacy violations, the data being potentially made accessible to other organizations (e.g., governments, commercial partners) and being subject to lucrative secondary usages (not advertised to the individuals). To face this situation, many recent initiatives push towards the emergence of the Personal Cloud paradigm. A personal cloud can be viewed as a personal server, owned by a given individual, which gives to its owner the ability to store her complete digital environment, synchronize it among various devices and share it with other individuals and applications under control. In the SMIS team, we claim the need of a Secure Personal Cloud, and promote the introduction of a secure (tamper resistant) data engine in the architecture [30]. On this basis, we investigate new data sharing and dissemination models, where usage and access control rules endorsed by the individuals could be enforced and have presented this vision at EDBT'14 [6] and at ADBIS'15 [31]. We have started a cooperation with the startup CozyCloud at the end of 2014. A contract was signed at the end of 2014 to integrate PlugDB in a CozyCloud instance and two CIFRE PhD thesis have been launched so far. Paul Tran Van's PhD thesis explores a new data sharing paradigm dedicated to the personal cloud context. This paradigm, called SWYSWYK (Share What You See with Who You Know), allows to automatically derive intuitive sharing rules from a personal cloud content, to share rules among a community of users and to let each user physically visualize the net effects of these rules on her own Personal Cloud. We propose a reference architecture providing the users with tangible guarantees about the enforcement of SWYSWYK policies and demonstrate through a performance evaluation conducted on a real personal cloud platform that the approach is practical. This work constitutes the core of Paul Tran Van's thesis and is being submitted for publication at VLDB. Preliminary ideas related to this work are presented in ERCIM news'16 [27]. Julien Loudet's PhD thesis is just starting with the objective to explore privacy-preserving distributed computations over personal clouds.

More generally, the personal cloud context gain in importance in our research work. It is even at the heart of our future project-team named PETRUS (PErsonal and TRUSted cloud). PETRUS is expected to take over from the SMIS team beginning of 2017.

6.4. Interdisciplinary study on Privacy-by-Design

Participants: Nicolas Anciaux, Luc Bouganim [correspondent], Athanasia Katsouraki, Benjamin Nguyen, Philippe Pucheral.

The objective of this research action is to study the reciprocal entanglements between economic, legal, societal and technological aspects of the management and exploitation of personal data. Indeed, devising new ways of protecting data privacy cannot be done in isolation; it requires also identifying alternative economic models that are both viable and regulatory compliant. We started an interdisciplinary research work with economists (RITM Lab) and jurists (CERDI and DANTE labs) in the privacy axis of ISN (Institut de la Société Numérique) and plan to pursue it in two projects in preparation: the Convergence Institute I2DRIVE (Interdisciplinary Institute for Data Research: Intelligence, Values and Ethics) and the CNRS Federation SIHS (Sciences Informatiques, Humaines et Sociales) at UVSQ. A first interdisciplinary work conducted in 2016 concerns the design of a privacy preserving platform needed to conduct privacy studies “in vivo”. Such platforms are required to validate the effectiveness of privacy preserving solutions, in terms of technical feasibility, lawfulness, acceptability and benefits. To this end, we have designed a privacy preserving mobile lab, derived from the personal cloud platform developed by the team (see ‘Software’ section). In her PhD thesis, Athanasia Katsouraki developed a beta-version of that platform and used it to perform a pre-experimentation in the context of online form based survey, targeting 140 students. The goal was threefold: (1) to test the effectiveness of the proposed platform, (2) to test the adequation of the questionnaire and experimentation protocol (a result for the experimental economist), and (3) to check the impact of the use of a secure platform on the student’s answers. The pre-experimentation showed several improvement axis and led to the actual design of the privacy preserving mobile lab described in the Software section.

Another joint work is related to the design of technical means to help individuals perceive how their personal life is exposed compared to others and to make appropriate protection choices. This work led to the definition of a new principle called Privacy-by-Using [20], that we introduced to try to circumvent the limits of the privacy-by-design principle promoted by the regulator. The confrontation of the Privacy-by-Using principle with Big Data processing [26] has also been studied with jurists and economists.

Finally, we conducted a scientific expertise on behalf of DGCCRF (Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes) and of the European Council regarding the draft proposal of "Directive of the European Parliament and of the Council on certain aspects concerning contracts for the online and other distance sales of goods" regulating the payment of numeric goods and services by means of personal data. This led us to a cross-analysis, with researchers in Law and computer scientists, of technical, societal and economic issues linked to the smart disclosure principle, that is, under which conditions and formats individuals can get their data back from service providers [17], [19], [18].

6.5. Formal guarantees

Participant: Guillaume Scerri.

The aim of the action is to investigate the changes required for the PlugDB architecture to be amenable to formal security proofs.

More precisely we started exploring the precise formal guarantees that are desirable for a personal data server. Following work started in Bristol [7], relating to formal guarantees provided by secure hardware, we started studying how one could leverage the low level guarantees provided by secure hardware (PlugDB for example) to cover the more complex operations and guarantees required of a personal data server. The first finding of the action is that a modular architecture is required for formal proofs to be obtainable. This is reflected in the architectural concerns presented in the PETRUS project.

Additionally, we started studying how to leverage secure hardware guarantees in order to perform secure computations on distributed data. A first result in this direction is presented in [33], and submitted to Financial Cryptography 2017.

TYREX Project-Team

6. New Results

6.1. Experimental evaluation of attitude estimation algorithms for smartphones

- **Context:** Pervasive applications on smartphones increasingly rely on techniques for estimating attitude. Attitude is the orientation of the smartphone with respect to Earth's local frame.

  Modern smartphones embed sensors such as accelerometer, gyroscope, and magnetometer which make it possible to leverage existing attitude estimation algorithms.
- **Contribution:** We investigated the precision of attitude estimation algorithms in the context of commodity smartphones carried by pedestrians. We considered eight typical motions (such as texting, phoning, running, etc.) with various impacts on external accelerations, as well as the presence/absence of magnetic perturbations typically found in indoor environments. We systematically analyzed, compared and evaluated eight state-of-the-art algorithms (and their variants). We precisely quantified the attitude estimation error obtained with each technique, owing to the use of a precise ground truth obtained with a motion capture system (the Inria Kinovis platform). We made our benchmark available (see Sec. 5.1 above) and payed attention to the reproducibility of results. We analyzed and discussed the obtained results and reported on lessons learned [7] [17]. We also presented a new technique which helps in improving precision by limiting the effect of magnetic perturbations with all considered algorithms.

6.2. Efficient Distributed Evaluation of SPARQL Queries

- **Context:** SPARQL is the standard query language for retrieving and manipulating data represented in the Resource Description Framework (RDF). SPARQL constitutes one key technology of the semantic web and has become very popular since it became an official W3C recommendation.

  The construction of efficient SPARQL query evaluators faces several challenges. First, RDF datasets are increasingly large, with some already containing more than a billion triples. To handle efficiently this growing amount of data, we need systems to be distributed and to scale. Furthermore, semantic data often have the characteristic of being dynamic (frequently updated). Thus being able to answer quickly after a change in the input data constitutes a very desirable property for a SPARQL evaluator.
- **Contributions:** First of all, to constitute a common basis of comparative analysis, we evaluated on the same cluster of machines various SPARQL evaluation systems from the literature [15]. These experiments led us to point several observations: (i) the solutions have very different behaviors; (ii) most of the benchmarks only use temporal metrics and forget other ones e.g. network traffic. That is why we proposed a larger set of metrics; and thanks to a new reading grid based on 5 features, we proposed new perspectives which should be considered when developing distributed SPARQL evaluators.

  Second, we developed and shared several distributed SPARQL evaluators which take into account these new considerations we introduced:

  - A SPARQL evaluator named SPARQLGX (see Sec. 5.6): an implementation of a distributed RDF datastore based on Apache Spark. SPARQLGX is designed to leverage existing Hadoop infrastructures for evaluating SPARQL queries. It relies on a translation of SPARQL queries into executable Spark code that adopts evaluation strategies according to the storage method used and statistics on data.

    In [12], [11], [8], [13], we showed that SPARQLGX makes it possible to evaluate SPARQL queries on billions of triples distributed across multiple nodes, while providing attractive performance figures. We reported on experiments which show how SPARQLGX compares to related state-of-the-art implementations and we showed that our approach scales better than these systems in terms of supported dataset size. With its simple design, SPARQLGX represents an interesting alternative in several scenarios.

– Two SPARQL direct evaluators i.e. without a preprocessing phase: SDE (stands for Sparqlgx Direct Evaluator) lays on the same strategy than SPARQLGX but the translation process is modified in order to take the orign data files as argument. RDFHive (see Sec. 5.3) evaluates translated SPARQL queries on top of Apache Hive which is a distributed relational data warehouse based on Apache Hadoop.

6.3. An Efficient Translation from a modal μ-Calculus with Converse to Tree Automata

In [16], we presented a direct translation from a sub-logic of μ-calculus to non-deterministic binary automata of finite trees. The logic is an alternation-free modal μ-calculus, restricted to finite trees and where formulae are cycle-free. This logic is expressive enough to encode significant fragments of query languages (such as Regular XPath). The size of the generated automaton (the number of transitions) is bounded by 2^n where n is the size of a Fischer-Ladner closure of the formula. This is an improvement over previous translations in 2^{n^2}. We have implemented our translation. In practice, our prototype effectively decides static analysis problems that were beyond reach, such as the XPath containment problem with DTDs of significant size.

6.4. SPARQL Query Containment with ShEx Constraints

ShEx (Shape Expressions) is a language for expressing constraints on RDF graphs. In [14], we considered the problem of SPARQL query containment in the presence of ShEx constraints. We first investigated the complexity of the problem according to the fragments considered for SPARQL queries and for ShEx constraints. In particular, we showed that the complexity of SPARQL query containment remains the same with or without ShEx constraints. We developed two radically different approaches for solving the problem and we evaluated them. The first approach relies on the joint use of a ShEx validator and a tool for checking query containment without constraints. In a second approach, we showed how the problem can be solved by a reduction to a fragment of first-order logic with two variables. This alternative approach allows to take advantage of any of the many existing FOL theorem provers in this context. We evaluated how the two approaches compare experimentally, and reported on lessons learned. To the best of our knowledge, this is the first work addressing SPARQL query containment in the presence of ShEx constraints.

6.5. XQuery Static Type-Checking

In the context of our ongoing work on XQuery static type-checking [3], we extended our type system and improved the associated software accordingly (see Sec. 5.5 and 5.4). The type language it is based on is now a subset of RelaxNG+Schematron (instead of DTDs), which is novel in the context of static typing: Schematron is normally used to validate a document after it has been generated, whereas our system is able to ensure statically that a program will always generate a valid document.

Schematron constraints present the advantage of describing some properties in a very concise way compared to schema languages based on regular tree types, e.g. it allows writing in one line that nested anchors are forbidden in HTML, a constraint which appears in the specification but not in the formal DTD schema because of the verbosity it would involve.

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: Patrice Pena, Alain Giboin.

This work is done in collaboration with Karima Boudaoud, SPARKS, I3S. In the context of the PadDOC FUI project, we elaborated 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 is 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 [71].

7.1.2. User Modeling of Collaborative Ontology Editors/Environments

Participant: Alain Giboin.

To demonstrate the importance of an in-depth modeling of users in the design of collaborative ontologies editors or environments (COEs), we began a study on the evolution of the user modeling techniques and the resulting user models from the origins of the design of COEs.

7.1.3. Recommendation of Pedagogical Resources Adapted to User Profile and Context

Participants: Oscar Rodríguez Rocha, Catherine Faron-Zucker.

In the continuation of the Semantic Educloud project, we constructed an ontology and associated thesaurus to represent an official standard of knowledge and skills. We proposed a process to extract knowledge and skills from the official texts describing the French educational program and to automatically populate our ontology with the knowledge extracted from the official texts which we further enrich by aligning it with the Web of Data. This work has been presented at the EKM 2016 workshop [49].

Together with researchers from DUIN (Italy), we worked on the design of a recommendation algorithm based on Linked Data, that could be used to recommend pedagogical resources. The algorithm exploits existing relationships between resources by dynamically analyzing both the categories to which they belong to and their explicit references to other resources. The algorithm has been applied in a mobile application to recommend movies by relying on DBpedia. This work has been presented at the RecSys workshop [50]

7.1.4. Requirements Analysis

Participant: Isabelle Mirbel.

Requirements representation in agile methods is often done on the basis of User Stories (US) which are short sentences relating a WHO, WHAT and (possibly) WHY dimension. They are by nature very operational and simple to understand thus very efficient. Previous research allowed to build a unified model for US templates associating semantics to a set of keywords based on templates collected over the Web and scientific literature. Since the semantics associated to these keywords is mostly issued of the i* framework [0], we overviewed in this work how to build a custom rationale diagram on the basis of a US set tagged using that unified template. The rationale diagram is strictly speaking not an i* strategic rationale diagram but uses parts of its constructs and visual notation to build various trees of relating US elements in a single project. Indeed, the benefits of editing such a rationale diagram is to identify depending US, identifying EPIC ones (EPIC: large User Story) and group them around common Themes. The results of this research have been published in [51].

[0] http://www.cs.toronto.edu/km/istar/

7.1.5. Design of a User-Centered Evaluation Method of Exploratory Search Systems Based on a Model of the Exploratory Search Process

Participants: Emilie Palagi, Alain Giboin, Fabien Gandon.

This work was undertaken in the context of the PhD of Emilie Palagi, in cooperation with with Raphaël Troncy (Eurecom). Our method takes into account users's Exploratory Search (ES) behavior and will be based on a cognitive model of an ES task. We will specially work on Discovery Hub and 3cixty 4 (EURECOM project) ESSs.

During the first year of the PhD, we were looking for a model of ES process on which the method will be based. To achieve this objective, several models of information seeking process were analyzed and we compared them with our own grid of the typical characteristics of exploratory search activities. The chosen model will fill the grid as much as possible with suitable adaptations if needed. It is an on-going work and we are actually designing an ES search model. We also performed a comparative analysis of 15 ESSs in order to identify the relevant functionalities supporting an exploratory search. We want to associate these functionalities to our grid of exploratory search characteristics. We will select some of these systems to test and validate the future method.

Contrary to lookup search engines that help users to retrieve specific items (e.g., names, numbers, short statements, or specific documents), Exploratory Search Systems (ESSs) are search engines that help users to explore a topic of interest. ES tasks are open-ended, multi-faceted, and iterative like learning or topic investigation [77], [80]. Currently, the evaluation methods of ESSs are not entirely adapted to the special features of ES tasks, and do not effectively assess that ESSs support users in performing those tasks. Our research goal is to elaborate methods that effectively lead to this assessment.

7.2. Communities and Social Interactions Analysis

7.2.1. Ontologies-Based Platform for Sociocultural Knowledge Management

Participants: Papa Fary Diallo, Olivier Corby, Isabelle Mirbel.

This work is done in the PhD Thesis of P. F. Diallo †. We designed a sociocultural platform aiming at persevering and capitalizing sociocultural events in Senegal. This platform relies on Semantic Web technologies. We provided two ontologies to support our platform: an upper level sociocultural ontology (USCO) and a human time ontology (HuTO). To build our upper level ontology we proposed a methodology based on the theory of Russian psychologist Lev Vygotsky called "Vygotskian Framework". We designed the Human Time Ontology [0] (HuTO) of which major contributions are (i) the modeling of non convex intervals (repetitive interval) like every Monday, (ii) the representation of deictic temporal expressions (e.g. *today*) which form specific relations with time speech and (iii) qualitative temporal notions which are temporal notions relative to a culture or a geographical position. The platform allows Senegalese communities to share and co-construct their sociocultural knowledge. This work was published in the Journal of Data Semantics [14].

7.2.2. SMILK - Social Media Intelligence and Linked Knowledge

Participants: Farhad Nooralahzadeh, Elena Cabrio, Molka Dhouib, Fabien Gandon.

Automated Natural Language Processing (NLP), Web Open Data (Linked Open Data) and social networks are the three topics of the SMILK ANR LabCom including their coupling studied in three ways: texts and Linked Data, Linked Data and social resources, texts and social resources. It is a Joint laboratory between the Inria research institute and the VISEO company to develop research and technologies, retrieve, analyze, and reason about linking data from textual Web resources and other to use open Web data taking into account the social structures and interactions in order to improve the analysis and understanding of textual resources.

[0] http://ns.inria.fr/huto

In this context, we have developed an entity discovery tool by adopting the semantic spreading activation, and we integrated it in the SMILK framework. The goal of such a tool is to semantically enrich the data by linking the mentions of named entities in the text to the corresponding known entities in knowledge bases. In our approach multiple aspects are considered: the prior knowledge of an entity in Wikipedia (i.e. the keyphraseness and commonness features that can be precomputed by crawling the Wikipedia dump), a set of features extracted from the input text and from the knowledge base, along with the correlation/relevancy among the resources in Linked Data. More precisely, this work explores the *collective ranking approach* formalized as a weighted graph model, in which the mentions in the input text and the candidate entities from knowledge bases are linked using the local compatibility and the global relatedness. Experiments on the datasets of the Open Knowledge Extraction (OKE) [0] challenge with different configurations of our approach in each phase of the linking pipeline reveal its optimum mode. We investigate the notion of semantic relatedness between two entities represented as sets of neighbors in Linked Open Data that relies on an associative retrieval algorithm, with consideration of common neighborhood. This measure improves the performance of prior link-based models and outperforms the explicit inter-link relevancy measure among entities (mostly Wikipedia-centric). Thus, our approach is resilient to non-existent or sparse links among related entities.

In parallel, an approach to automatically annotate texts in the cosmetics field with the vocabularies ProVoc and GoodRelations in RDF has been proposed, resulting in a knowledge base in the format of the Semantic Web that can be used in various applications. Given the entity linking tool described before (that allows to link named entities in a text with entities in the LOD), we focused on the extraction of relations between these entities (in French texts). In the extraction process, particular attention is given to the contribution of syntactic rules, in order to improve accuracy with respect to existing systems.

7.2.3. Community Detection and Interest Labeling

Participants: Zide Meng, Fabien Gandon, Catherine Faron-Zucker.

7.2.3.1. Temporal Analysis of User and Topic

Based on previous work on overlapping community detection in Question-Answer sites, we proposed an approach to jointly model topic, expertise, activity and trends, we were able to retrieve many meaningful latent information from the user generated contents. We proposed a method to track the dynamics of topics and users. It can also track the dynamics with a specific granularity of time level such as, yearly, monthly, daily and hourly. Besides, the model can overcome a comparison problem of LDA (Latent Dirichlet Allocation) based model by modeling the reverse distribution. This work has been published in IEEE/WIC/ACM Web Intelligence [62].

7.2.3.2. Topic labeling

The output of topic model is normally a bag of words. Each topic consists of closely related words. An interesting question is to assign one or more topic label to this set in order to indicate the general meaning of a bag of words. By integrating the original dataset with linked open data sources, we are now planning to propose a generic method to automatically label the detected topics.

7.2.4. Default Knowledge based on the Analysis of Natural Language

Participants: Elena Cabrio, Valerio Basile, Fabien Gandon.

In the context of the ALOOF project, we developed new methods to build repositories of default knowledge based on the analysis of natural language. The first efforts are aimed at extracting information about common objects, in particular their location and their typical usage [24].

One of the methods to extract general knowledge from text is implemented in the KNEWS pipeline, of which a demo was presented at ECAI [25]. At the same conference, we also presented the results of another system that helps robots identifying unknown objects based on their proximity with known objects observed at the scene [52]. KNEWS was also used to automatically build a large collection of text aligned with RDF semantics representation of its meaning. The first envisioned application of such resource is to provide a basis for robust natural language generation from RDF triples using statistical methods [22].

[0] https://github.com/anuzzolese/oke-challenge

We also explored the application of distributional semantics to the general knowledge extraction problem. We computed vector-based models of objects and used supervised statistical models to predict their typical locations (e.g. knife-kitchen, printer-office) [27]. Once our models were successfully tested experimentally against a gold standard of human judgments, we were able to build a large knowledge base of object locations freely available [0].

7.2.5. Semantic Modeling of Social, Spatiotemporal and Dedicated Networks

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

During the academic year 2015/2016, we have been working partially on validating the model we proposed in [72]. A long version of this former paper, entitled *An Agent-based Architecture for Personalized Recommendations* will be published in January 2017 in the LNCS series published by Springer. For this purpose, we proposed in [29] a multi-agent based simulation on NetLogo environment in order to illustrate the usefulness and feasibility of the proposed framework in a realistic scenario. For that purpose, we evaluated the performance of the agent behaviors adopting two different strategies:

- Selfish agents: agents do not communicate with each others;
- Social agents: agents communicate and try to influence each other's to adopt some beliefs or desires.

Results show that agents achieve a better performance collectively when they are in "communities", i.e., agents with shared interests (thus similar to each other), than when they are acting as solitary agents. We believe that the issues of trust and recommendation are tightly related. For that reason, we analyzed the behavior of social agents with and without a trust model. Results show that exchanging beliefs or desires with trustworthy agents can improve the whole performance of agents.

We have been also working on extending the proposed model with spatial and temporal reasoning. A spatio-temporal belief or desire is considered as an event that is defined as a spatial relation holding in a temporal interval. For reasoning with such information, we propose to combine the Region Connection Calculus (RCC-8) formalism with Allen's intervals algebra. Spatio-temporal data is often affected by imprecision and vagueness. To tackle this problem we believe that a fuzzy set, because its ability to represent a degree of membership, is more suitable for modeling spatio-temporal data. A fuzzy version of RCC-8 and Allen's interval is proposed. Then we combined both approaches in order to represent and reason about imprecise spatio-temporal beliefs and desires. We worked also in validating this approach in a real-world scenario.

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

7.3.1. Semantic Web Technologies and Natural Language

Participants: Serena Villata, Elena Cabrio.

Together with Sara Tonelli (FBK, Italy) and Mauro Dragoni (FBK, Italy), we have presented the integration, enrichment and interlinking activities of metadata from a small collection of verbo-visual artworks in the context of the Verbo-Visual-Virtual project. We investigate how to exploit Semantic Web technologies and languages combined with natural language processing methods to transform and boost the access to documents providing cultural information, i.e., artist descriptions, collection notices, information about technique. We also discuss the open challenges raised by working with a small collection including little-known artists and information gaps, for which additional data can be hardly retrieved from the Web. The results of this research have been published at the ESWC conference [37].

[0] https://project.inria.fr/aloof/data/

Together with Vijay Ingalalli (LIRMM), Dino Ienco (IRSTEA), Pascal Poncelet (LIRMM), we have introduced AMbER (Attributed Multigraph Based Engine for RDF querying), a novel RDF query engine specifically designed to optimize the computation of complex queries. AMbER leverages subgraph matching techniques and extends them to tackle the SPARQL query problem. AMbER exploits structural properties of the query multigraph as well as the proposed indexes, in order to tackle the problem of subgraph homomorphism. The performance of AMbER, in comparison with state-of-the-art systems, has been extensively evaluated over several RDF benchmarks. The results of this research have been published at the EDBT conference [39].

7.3.2. Semantic Web Languages and Techniques for Digital Humanities

Participants: Catherine Faron-Zucker, Franck Michel, Konstantina Poulida, Safaa Rziou, Andrea Tettamanzi.

In the framework of the Zoomathia project, we conducted three complementary works, with the ultimate goal of exploiting semantic metadata to help historians in their studies of knowledge transmission through texts. First, together with Olivier Gargominy and other MNHN researchers, and Johan Montagnat (I3S, UNS), we continued a work initiated last year on the construction of a SKOS (Simple Knowledge Organization System) thesaurus based on the TAXREF taxonomical reference, designed to support studies in Conservation Biology [73]. We deployed the Corese Semantic Web factory as a backend to publish this SKOS thesaurus on the Web of Linked Open Data. This work was presented at the SemWeb.Pro 2016 conference.

Second, together with Irene Pajon (UNS) and Arnaud Zucker (UNS), we continued a work initiated last year on the construction of a SKOS thesaurus capturing zoological specialities (ethology, anatomy, physiology, psychology, zootechnique, etc.). This thesaurus was constructed while manually annotating books VIII-XI of Pliny the Elder's Natural History, chosen as a reference dataset to elicit the concepts to be integrated in the Zoomathia thesaurus. This work has been published in the ALMA journal [79] (Archivum Latinitatis Medii Aevi).

Third, together with Arnaud Zucker (UNS), we developed an approach of knowledge extraction from ancient texts consisting in semantically categorizating text segments based on machine learning methods applied to a representation of segments built by processing their translations in modern languages with Natural Language Processing (NLP) methods and by exploiting the above described thesaurus of zoology-related concepts. We applied it to categorize Pliny the Elder's Natural History segments. The above describe manually annotated dataset served us as goldstandard evaluate our approach. This work has been presented at the ESWC 2016 workshop on Semantic Web for Scientific Heritage [38].

Relatedly, together with Emmanuelle Kuhry (UNS) and Arnaud Zucker (UNS), we developed an approach which originates in seeing copying as a special kind of ”virtuous” plagiarism and consists in paradoxically using plagiarism detection tools in order to measure *distances* between texts, rather than similarities. We first applied it to the *Compendium Philosophie*'s tradition, whose manuscript tradition is well studied and mostly understood and can therefore be considered as a gold standard. Then we applied the validated and calibrated method to investigate the *Physiologus latinus*'s tradition, which is a complex manuscript tradition for which our knowledge is much less sure, with the aim of supporting the elaboration of stemmatological hypotheses.

7.3.3. Argumentation Theory and Multiagent Systems

Participants: Andrea Tettamanzi, Serena Villata.

Together with Célia da Costa Pereira (I3S, UNS) we have proposed a formal framework to support belief revision based on a cognitive model of credibility and trust. In this framework, the acceptance of information coming from a source depends on *(i)* the agent's goals and beliefs about the source's goals, *(ii)* the credibility, for the agent, of incoming information, and *(iii)* the agent's beliefs about the context in which it operates. This makes it possible to approach belief revision in a setting where new incoming information is associated with an acceptance degree. In particular, such degree may be used as input weight for any possibilistic conditioning operator with uncertain input (i.e., weighted belief revision operator). The results of this research have been published at the SUM conference [56].

Moreover, together with Célia da Costa Pereira (UNS) and Mauro Dragoni (FBK, Italy), we have provided an experimental validation of the fuzzy labeling algorithm proposed by da Costa Pereira et al. at IJCAI-2011 with the aim of carrying out an empirical evaluation of its performance on a benchmark of argumentation graphs. Results show the satisfactory performance of our algorithm, even on complex graph structures as those present in our benchmark. The results of this research have been published at the SUM conference [55].

Serena Villata, together with the other organizers, has also reported about the results of the first Computational Argumentation Challenge (ICCMA) in a AI Magazine paper [17].

7.3.4. RDF Mining

Participants: Amel Ben Othmane, Tran Duc Minh.

In collaboration with Claudia d'Amato of the University of Bari, Italy, we have carried on our investigation about extracting knowledge from RDF data, by proposing a level-wise generate-and-test [53] and an evolutionary [54] approach to discovering multi-relational rules from ontological knowledge bases which exploits the services of an OWL reasoner.

7.3.5. LDScript Linked Data Script Language

Participants: Olivier Corby, Catherine Faron-Zucker, Fabien Gandon.

We design and develop LDScript, a Linked Data Script Language [68]. It is a DSL (domain-specific programming language) the objects of which are RDF terms, triples and graphs as well as SPARQL query results. Its main characteristic is to be designed on top of SPARQL filter language in such a way that SPARQL filter expressions are LDScript expressions. Mainly speaking, it introduces a function definition statement into SPARQL filter language. The main use case of LDScript is the definition of SPARQL extension functions and custom aggregates. With LDScript, we were able to develop a W3C DataShape SHACL [0] validator using STTL and we provide a Web service [0].

7.3.6. Ontology-based Workflow Management Systems

Participants: Tuan-Anh Pham, Nhan Le Thanh.

The main objective of the PhD work is to improve Coloured Petri Nets (CPNs) and Ontology engineering to support the development of business process and business workflow definitions of the various fields and to develop a Shared Workflow Management System (SWMS) using the ontology engineering. Everybody can share a semi-complete workflow which is called ”Workflow template”, and other people can modify and complete it to use in their system. This customized workflow is called ”Personalized workflow”. The challenges of a SWMS is to be simple, easy to use, friendly with the user and not too heavy. But it must have all functions of a WMS. There are three major challenges in this work: How to allow the users to customize the workflow template to correspond to their requirements, but with their changes compliant with the predefined rules in the workflow template? How to build an execution model to evaluate step by step a personalized workflow?

7.3.7. A Service Infrastructure Providing Access to Variables and Heterogeneous Resources

Participants: The-Cân Do, Nhan Le Thanh.

This work is done together with Gaëtan Rey (I3S, PhD co-director). The aim of this PhD work is to develop an adaptation of applications to their context. However, in view of the difficulties of context management in its entirety, we choose to approach the problem by decomposing context management from different points of views (or contextual concerns). A concern (or point of view) may be the business process of the application, security, etc. or any other cross-functionality. In addition to simplifying the context management, sharing between different experts the analysis to be performed, this approach aims to allow the reuse of specifications of each point of view between different applications. Finally, because of the independence of points of view (from their specification to implementation), it is easily conceivable to add and/or delete dynamically points

[0] https://www.w3.org/TR/shacl/
[0] http://corese.inria.fr

of view during the execution of the application we want to adapt. The scientific challenge of this thesis is based on the automatic resolution of conflicts between the points of view made to the adaptation of the target application. Of course, this must be done at runtime.

7.3.8. DBpdia.fr & DBpedia Historic

Participants: Raphaël Boyer, Fabien Gandon, Olivier Corby, Alexandre Monnin.

A new version of the DBpedia historic extractor has been developed and the database is publicly accessible on a dedicated Web server footnote http://dbpedia-historique.inria.fr/sparql. We redesigned the DBpedia Live mechanism from the international DBpedia community to deploy a DBpedia live instance that is able to update itself in near real time by following the edition notification feed from Wikipedia; it is available on our server [0].

We also designed a new DBpedia extractor materializing the editing history of Wikipedia pages as linked data to support queries and indicators on the history [61], [60]. An example of application supported by this service is showed in figure 1 where we provide a Web portal based on STTL [18] crossing linked data from DBpedia.fr and DBpedia Historic to detect events concerning artists.

Finally, we redesigned the DBpedia.fr Web site with a responsive interface, a modern design and a technical documentation. The Web site is also available in English because internationalizing the document allows a wider audience8 to use the data extracted.

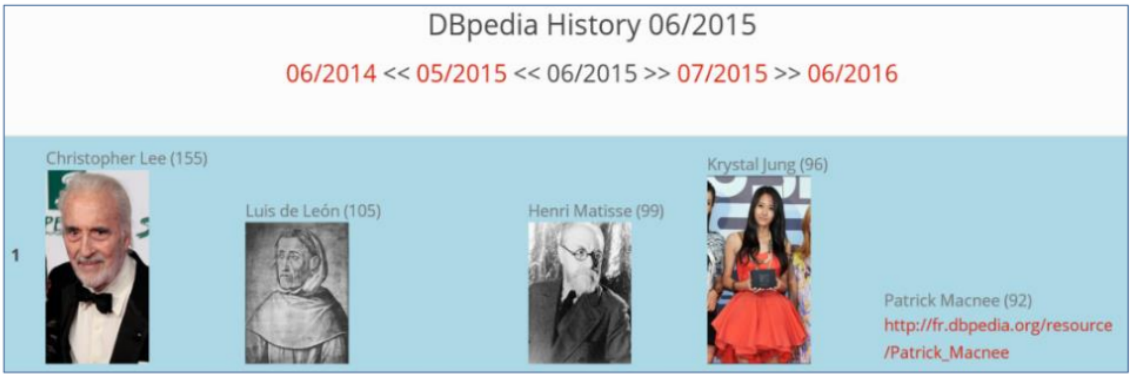


Figure 1. DBpedia Artist category with edition history

7.3.9. Provoc Ontology from SMILK

Participants: Fabien Gandon, Elena Cabrio.

ProVoc [0] (Product Vocabulary) is a vocabulary that can be used to represent information about Products and manipulate them through the Web. This ontology reflects: the basic hierarchy of a company (Group/Company, Divisions of a Group, Brand names attached to a Division or a Group) and the production of a company (products, ranges of products, attached to a Brand, the composition of a product, packages of products, etc.).

7.4. Analyzing and Reasoning on Heterogeneous Semantic Graphs

7.4.1. SPARQL Template Transformation Language

Participants: Olivier Corby, Catherine Faron-Zucker, Raphaël Gazzotti.

[0] http://dbpedia-live.inria.fr/sparql
[0] http://ns.inria.fr/provoc

In the continuation of our work on the design of the STTL SPARQL Template Transformation Language [18], we showed that it can be used as a constraint language for RDF and we applied our approach to implement the semantics of OWL 2 profiles, each viewed as a set of constraints to be validated: we defined an STTL transformation to represent each of the three OWL 2 profiles (OWL RL, OWL QL and OWL EL). The application of one of these STTL transformations to an ontology (in OWL/RDF syntax) enables users to validate it against the OWL 2 profile this transformation represents. This work has been presented at the RR 2016 conference [34].

7.4.2. Exposing Heterogeneous Data Sources on the Web of Linked Open Data

Participants: Catherine Faron-Zucker, Franck Michel.

While the emerging Web of Data continuously grows as data sets are published as Linked Open Data, data is produced ever faster in data silos where it often remains locked. In particular, NoSQL systems have gained a remarkable success during recent years. Consequently, harnessing the data available in NoSQL databases to populate the Web of Data, and more generally achieving RDF-based data integration and SPARQL querying of NoSQL databases, are timely questions.

Together with Johan Montagnat (I3S, UNS), we previously proposed a generic mapping language, xR2RML, able to describe the mapping of most common types of databases into an arbitrary RDF representation [78]. In the continuation of this work, we developed a two-step approach to execute SPARQL queries over heterogeneous databases based on the xR2RML mapping of the database to RDF. We demonstrated the effectiveness of this approach by providing SPARQL access over MongoDB, the popular NoSQL document store. This work was undertaken in the context of the PhD of Franck Michel, and was published in the WebIST 2016 conference [43], and in the DEXA 2016 conference [44].

7.4.3. Combining Argumentation Theory and Natural Language Processing

Participants: Serena Villata, Valerio Basile, Elena Cabrio, Andrea Tettamanzi, Tom Bosc.

We have proposed a new approach to text exploration combining argumentation theory and natural language processing. They define bipolar entailment graphs, i.e., graphs whose nodes are text fragments and the edges represent the entailment or non entailment relations. They adopt abstract dialectical frameworks to define acceptance conditions for the nodes such that the resulting framework returns us relevant information for the text exploration task. The results of this research have been published at the ICAART conference [33].

Moreover, we have proposed a new approach to argument mining for Twitter data. The proposed approach consists first in detecting argumentative tweets from a stream of tweets, and second, starting from this set of argument-tweets, in predicting the relations, i.e., attack and support, holding between two argument-tweets. The annotated corpus resulting from this research line has been described in a paper published at the LREC conference [30], while the results of the argument mining task have been published at the COMMA conference [31].

Following a novel research direction, we investigated the relationship between the emotions displayed by the participants to our experiments and the sentiment expressed in the natural language of their arguments. We ran state-of-the-art sentiment analysis software on the transcriptions of the debates and compared the result with the output of the emotion reading systems. The results of our analysis were presented at the Artificial Intelligence and Cognition Workshop [26] and at the Italian Conference on Computational Linguistics [23].

Finally, together with Celia da Costa Pereira (UNS) and Mauro Dragoni (FBK, Italy), we have proposed an opinion summary application built on top of an argumentation framework, used to exchange, communicate and resolve possibly conflicting viewpoints in distributed scenarios. They show how this application is able to extract relevant and debated opinions from a set of documents containing user-generated content from online commercial Web sites. The result of this research has been published as a short paper at the IJCAI conference [35], and an extended version has been submitted to the AI Comm. journal and it is currently under review.

7.4.4. Opinion Mining

Participants: Andrea Tettamanzi, Serena Villata.

Together with Célia da Costa Pereira of I3S and Mauro Dragoni of FBK, Trento, who visited our team for three months from April to June 2014, we have proposed DRANZIERA, an evaluation protocol for the evaluation of multi-domain opinion mining methods [36] and an argumentation framework for opinion mining [35].

7.4.5. SMILK - Automatic Generation of Quizzes through Semantic Web Technologies

Participant: Oscar Rodríguez Rocha.

The research work focuses on the automatic generation of quizzes using Semantic Web technologies. It takes inspiration from the existing research works about automatic generation of multi choice questions from domain ontologies and aims to apply such existing techniques and contibute to its extension, in order to semantically generate statements that allow to describe the content of a given Web ontology. This research work is carried out in the context of SMILK. SMILK (Social Media Intelligence and Linked Knowledge) is a joint laboratory (LabCom, 2013-2016) between the Wimmics team and the Research and Innovation unit of VISEO (Grenoble). Natural Language Processing, Linked Open Data and Social Networks as well as the links between them are at the core of this LabCom. The purpose of SMILK is both to develop research and technologies in order to retrieve, analyze, and reason on textual data coming from Web sources, and to make use of LOD, social networks structures and interaction in order to improve the analysis and understanding of textual resources. Topics covered by SMILK also include: use of data and vocabularies published on the Web in order to search, analyze, disambiguate and structure textual knowledge in a smart way, but also to feed internal information sources; reasoning on the combination of internal and public data and schemes, query and presentation of data and inferences in natural formats.

7.4.6. Event Identification & Tracking

Participants: Amosse Edouard, Elena Cabrio, Nhan Le Thanh.

In the past year, we have been working on approaches for detecting, classifying and tracking events on Twitter. In the context of social media, an event is considered as "An occurrence causing change in the volume of text data that discusses the associated topic at a specific time. This occurrence is characterized by topic and time, and often associated with entities such as people and location". This definition shows that Named Entities (NE) play a key role in events on social medias and particularly on Twitter. In our approaches we exploit the NE in tweets to analyse events on Twitter.

7.4.6.1. Event Identification and Classification

We developed an approach that exploit occurrences of Named Entities in tweets to train a supervised model for two purposes:

- To classify tweets as either related or not related to events.
- To classify tweets related to events as event categories such as Economy, Politics or Sport.

We combined techniques from Natural Language Processing, Linked Open Data and Machine Learning to build a supervised model for classifying tweets. More specifically, we replaced the NE in tweets by their related class in ontologies (e.g DBpedia or YAGO) and used the modified content to train machine learning algorithms (e.g. SVM, Naive Bayes and Neural Network). Our experiments on two gold standard datasets shown that the NER mechanism helped in reducing overfitting on the output of classifiers.

7.4.6.2. Event Tracking

More recently, we started to work on an approach for tracking planned events on Twitter. In this work, we were particularly interested in tracking the evolution of existing events over time. For example, important actions in a soccer game (goal, yellow/red cards). We proposed an unsupervised approach based on NE in tweets and graph analysis to process the Twitter stream in real time. In this approach, we dynamically update a local gazetteer with actors involved in the events such as player and team names as well as terms that describe the actions of interests (e.g. goal, yellow card for football). The preliminary evaluations are quite promising since we are able to track the most important events in a soccer game as well as the player or teams involved in the actions.

7.4.7. Software and Hardware Architecture of EMOTICA: an Emotions Detection System

Participant: Nhan Le Thanh.

This work is performed with Chaka Kone (3rd year PhD student - LEAT, UNS) and Cécile Belleudy (Thesis Director - LEAT, UNS). The aim of this PhD work is to propose a complete low power system for the recognition of emotions satisfying all application constraints such as energy consumption, size and positioning of sensors. To achieve this goal, our work focuses on two main axes: the detection of emotions and the architectural exploration of objects communicating for health, with particular emphasis on the energy consumption of such systems.

7.4.8. Conversational Agent Assistant

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

This CIFRE PhD Thesis is performed in collaboration with SynchroNext, a company located in Nice. As part of this thesis, we will be interested in setting up an ECA (Embodied Conversational Agents) for FAQs to advisers. The ECA will need to integrate a question and answer system to address the most common issue types without human intervention [76], [81]. For this purpose, it must be able to understand the questions asked in natural language by the users and to reason with the knowledge acquired. Beyond such a system of questions and answers, the ECA must be able to reopen the conversation with the Internet user according to the nature of his requests or the sequence of questions formulated. The objective is to reduce the dropout rate of Internet users on FAQs and to reduce the number of incoming calls and e-mails. This will enable to customer advisers to focus on more difficult questions.

ZENITH Project-Team

7. New Results

7.1. Data Integration

7.1.1. CloudMdsQL, a query language for heterogeneous data stores

Participants: Carlyna Bondiombouy, Boyan Kolev, Oleksandra Levchenko, Patrick Valduriez.

In the context of the CoherentPaaS European project, we have developped the Cloud Multi-datastore Query Language (CloudMdsQL), and its query engine. CloudMdsQL is a functional SQL-like language, capable of querying multiple heterogeneous data stores, e.g. relational, NoSQL or HDFS) [21]. The major innovation is that a CloudMdsQL query can exploit the full power of the local data stores, by simply allowing some local data store native queries to be called as functions, and at the same time be optimized. In [42], we demonstrate CloudMdsQL on two use cases each involving four diverse data stores (graph, document, relational, and key-value) with its corresponding CloudMdsQL queries. The query execution flows are visualized by an embedded real-time monitoring subsystem. In [17], we extend CloudMdsQL to allowing the ad-hoc usage of user defined map/filter/reduce operators in combination with traditional SQL statements, to integrate relational data and big data stored in HDFS and accessed by a data processing framework like Spark. Our experimental validation with several different data stores and representative queries [43] demonstrates the usability of the query language and the benefits from query optimization.

7.1.2. Agronomic Linked Data

Participant: Pierre Larmande.

Agronomic Linked Data (AgroLD) [30], [55], [54] is a knowledge system that exploits Semantic Web technology to integrate information on plant species widely studied by the agronomic research community. The objective is to provide the community with a platform for domain specific knowledge, capable of answering complex biological questions and thus facilitating the formulation of new hypotheses. The conceptual framework is based on well-established ontologies in plant sciences such as Gene Ontology, Sequence Ontology, Plant Ontology and Plant Environment Ontology. AgroLD version 1 consists of 50 million knowledge statements (i.e. RDF triples), which will grow in the subsequent versions to provide the required critical mass for hypotheses generation.

AgroLD relyes on AgroPortal [40], a reference ontology repository for the agronomi domain that features ontology hosting and search visualization with services for semantically annotating data with the ontologies. We used the AgroPortal Annotator web service to annotate more than 50 datasets and produced 22% additional triples validated manually. We also developed a dedicated AgroLD vocabulary that bridges the gap between these references ontologies and formalizes their mappings.

7.2. Data Management

7.2.1. Scalable Query Processing with Big Data

Participants: Reza Akbarinia, Patrick Valduriez.

In [22], we extend the popular Hadoop framework to deal efficiently with skewed MapReduce jobs. We extend the MapReduce programming model to allow the collaboration of reduce workers on processing the values of an intermediate key, without affecting the correctness of the final results. In FP-Hadoop, the reduce function is replaced by two functions: intermediate reduce and final reduce. There are three phases, each phase corresponding to one of the functions: map, intermediate reduce and final reduce phases. In the intermediate reduce phase, the function, which usually includes the main load of reducing in MapReduce jobs, is executed by reduce workers in a collaborative way, even if all values belong to only one intermediate key. This allows performing a big part of the reducing work by using the computing resources of all workers, even in case of highly skewed data. We implemented a prototype of FP-Hadoop by modifying Hadoop's code, and conducted extensive experiments over synthetic and real datasets. The results show that FP-Hadoop makes MapReduce job processing much faster and more parallel, and can efficiently deal with skewed data. We achieve excellent performance gains compared to native Hadoop, e.g. more than 10 times in reduce time and 5 times in total execution time.

7.2.2. Management of Simulation Data

Participant: Patrick Valduriez.

Supported by increasingly efficient HPC infrastructures , numerical simulations are rapidly expanding to fields such as oil and gas, medicine and meteorology. As simulations become more precise and cover longer periods of time, they may produce files with terabytes of data that need to be efficiently analyzed. In [24], we investigate techniques for managing such data using an array DBMS. We take advantage of multidimensional arrays that nicely models the dimensions and variables used in numerical simulations.We propose efficient techniques to map coordinate values in numerical simulations to evenly distributed cells in array chunks with the use of equi-depth histograms and space-filling curves. We implemented our techniques in SciDB and, through experiments over real-world data, compared them with two other approaches: row-store and column-store DBMS. The results indicate that multidimensional arrays and column-stores are much faster than a traditional row-store system for queries over a larger amount of simulation data. They also help identifying the scenarios where array DBMSs are most efficient, and those where they are outperformed by column-stores.

7.3. Scientific Workflows

7.3.1. A Scientific Workflow Infrastructure for Plant Phenomics

Participants: Didier Parigot, Patrick Valduriez.

Plant phenotyping consists in the observation of physical and biochemical traits of plant genotypes in response to environmental conditions. There are many challenges, in particular in the context of climate change and food security. High-throughput platforms have been introduced to observe the dynamic growth of a large number of plants in different environmental conditions. Instead of considering a few genotypes at a time (as it is the case when phenomic traits are measured manually), such platforms make it possible to use completely new kinds of approaches. However, the datasets produced by such widely instrumented platforms are huge, constantly augmenting and produced by increasingly complex experiments, reaching a point where distributed computation is mandatory to extract knowledge from data.

In[25], we introduce InfraPhenoGrid,an infrastructure to efficiently manage datasets produced by the PhenoArch plant phenomics platform in the context of the French Phenome Project. Our solution consists in deploying scientific workflows on a grid using a middle-ware to pilot workflow executions. Our approach is user-friendly in the sense that despite the intrinsic complexity of the infrastructure, running scientific workflows and understanding results obtained (using provenance information) is kept as simple as possible for end-users.

7.3.2. Managing Scientific Workflows in Multisite Cloud

Participants: Ji Liu, Esther Pacitti, Patrick Valduriez.

A cloud is typically made of several sites (or data centers), each with its own resources and data. Thus, it becomes important to be able to execute big scientific workflows at multiple cloud sites because of geographical distribution of data or available resources. Therefore, a major problem is how to execute a SWf in a multisite cloud, while reducing execution time and monetary cost. In [23], we propose a general solution based on multi-objective scheduling in order to execute SWfs in a multisite cloud. The solution includes a multi-objective cost model with execution time and monetary cost, a Single Site Virtual Machine (VM) Provisioning approach (SSVP) and ActGreedy, a multisite scheduling approach. We present an experimental evaluation, based on the execution of the SciEvol workflow in Microsoft Azure cloud. The results reveal that our scheduling approach significantly outperforms two adapted baseline algorithms and the scheduling time is reasonable compared with genetic and brute-force algorithms.

In [46], we present a hybrid decentralized/distributed model for handling frequently accessed metadata in a multisite cloud. We couple our model with a scientific workflow management system (SWfMS) to validate and tune its applicability to different real-life scientific scenarios. We show that efficient management of hot metadata improves the performance of SWfMS, reducing the workflow execution time up to 50% for highly parallel jobs and avoiding unnecessary cold metadata operations.

7.3.3. Online Input Data Reduction in Scientific Workflows

Participant: Patrick Valduriez.

Many scientific workflows are data-intensive and must be iteratively executed for large input sets of data elements. Reducing input data is a powerful way to reduce overall execution time in such workflows. When this is accomplished online (i.e., without requiring users to stop execution to reduce the data and resume execution), it can save much time and user interactions can integrate within workflow execution. Then, a major problem is to determine which subset of the input data should be removed. Other related problems include guaranteeing that the workflow system will maintain execution and data consistent after reduction, and keeping track of how users interacted with execution. In [48], we adopt the approach “human-in-the-loop” for scientific workflows by enabling users to steer the workflow execution and reduce input elements from datasets at runtime. We propose an adaptive monitoring approach that combines workflow provenance monitoring and computational steering to support users in analyzing the evolution of key parameters and determining which subset of the data should be removed. We also extend a provenance data model to keep track of user interactions when users reduce data at runtime. In our experimental validation, we develop a test case from the oil and gas industry, using a 936-cores cluster. The results on our parameter sweep test case show that the user interactions for online data reduction yield a 37% reduction of execution time.

7.4. Data Analytics

7.4.1. Parallel Mining of Maximally Informative k-Itemsets

Participants: Saber Salah, 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 [26], [52], we address the problem of parallel mining of maximally informative k-itemsets (miki) based on joint entropy. We propose PHIKS (Parallel Highly Informative K-itemSets) a highly scalable, parallel mining algorithm. PHIKS renders the mining process of large scale databases (up to terabytes of data) succinct and effective. Its mining process is made up of only two compact, yet efficient parallel jobs. PHIKS uses a clever heuristic approach to efficiently estimate the joint entropies of miki having different sizes with very low upper bound error rate, which dramatically reduces the runtime process. PHIKS has been extensively evaluated using massive, real-world datasets. Our experimental results confirm the effectiveness of our approach by the significant scale-up obtained with high featuresets length and hundreds of millions of objects.

7.4.2. Chiaroscuro

Participants: Tristan Allard, Florent Masseglia, Esther Pacitti.

New personal data fields are currently emerging due to the proliferation of on-body/at-home sensors connected to personal devices. However, strong privacy concerns prevent individuals to benefit from large-scale analytics that could be performed on this fine-grain highly sensitive wealth of data. In [32] we propose a demonstration of Chiaroscuro, a complete solution for clustering massively-distributed sensitive personal data while guaranteeing their privacy. The demonstration scenario highlights the affordability of the *privacy vs. quality* and *privacy vs. performance* tradeoffs by dissecting the inner working of Chiaroscuro, exposing the results obtained by the individuals participating in the clustering process, and illustrating possible uses.

7.5. Data Search

7.5.1. Spatially Localized Visual Dictionary Learning

Participants: Valentin Leveau, Alexis Joly, Patrick Valduriez.

In [44], we devise new representation learning algorithms that overcome the lack of interpretability of classical visual models. We introduce a new recursive visual patch selection technique built on top of a Shared Nearest Neighbors embedding method. The main contribution is to drastically reduce the high-dimensionality of such over-complete representation using a recursive feature elimination method. We show that the number of spatial atoms of the representation can be reduced by up to two orders of magnitude without degrading much the encoded information. The resulting representations are shown to provide competitive image classification performance with the state-of-the-art while enabling to learn highly interpretable visual models. This contribution was the last one in Valentin Leveau's PhD on Nearest Neighbor Representations [13].

7.5.2. Crowdsourcing Biodiversity Monitoring

Participants: Alexis Joly, Julien Champ, Herve Goeau, Jean-Christophe Lombardo.

Large scale biodiversity monitoring is essential for sustainable development (earth stewardship). With the recent advances in computer vision, we see the emergence of more and more effective identification tools, thus allowing large-scale data collection platforms such as the popular Pl@ntNet initiative to reuse interaction data. Although it covers only a fraction of the world flora, this platform has been used by more than 300K people who produce tens of thousands of validated plant observations each year. This explicitly shared and validated data is only the tip of the iceberg. The real potential relies on the millions of raw image queries submitted by the users of the mobile application for which there is no human validation. People make such requests to get information on a plant along a hike or something they find in their garden but do not know anything about. Allowing the exploitation of such contents in a fully automatic way could scale up the world-wide collection of implicit plant observations by several orders of magnitude, thus complementing the explicit monitoring efforts.

In [37], we first survey existing automated plant identification systems through a five-year synthesis of the PlantCLEF benchmark and an impact study of the Pl@ntNet platform. We then focus on the implicit monitoring scenario and discuss related research challenges at the frontier of computer science and biodiversity studies. Finally, we discuss the results of a preliminary study focused on implicit monitoring of invasive species in mobile search logs. We show that the results are promising while there is room for improvement before being able to automatically share implicit observations within international platforms.

7.5.3. Unsupervised Individual Whales Identification

Participants: Alexis Joly, Jean-Christophe Lombardo.

Identifying organisms is critical in accessing information related to the ecology of species. Unfortunately, this is difficult to achieve due to the level of expertise necessary to correctly identify and record living organisms. To bridge this gap, a lot of work has been done on the development of automated species identification tools such as image-based plant identification or audio recordings-based bird identification. Yet, for some groups, it is preferable to monitor the organisms at the individual level rather than at the species level. The automation of this problem has received much less attention than species identification.

In [39], we address the specific scenario of discovering humpack whale individuals in a large collection of pictures collected by nature observers. The process is initiated from scratch, without any knowledge on the number of individuals and without any training samples of these individuals. Thus, the problem is entirely unsupervised. To address it, we set up and experimented a scalable fine-grained matching system, which allows discovering small rigid visual patterns in highly cluttered backgrounds. The evaluation was conducted in blind in the context of the LifeCLEF evaluation campaign. Results show that the proposed system provides very promising results with regard to the difficulty of the task but that there is still room for improvements to reach higher recall and precision in the future. This work was done in collaboration with the Cetamada NGO.

7.5.4. Evaluation of Biodiversity Identification and Search Techniques

Participants: Alexis Joly, Herve Goeau, Jean-Christophe Lombardo.

We ran a new edition of the LifeCLEF evaluation campaign in the context of the CLEF international research forum. We did share a new subset of the data produced by the Pl@ntNet platform and set up three new challenges: one related to the identification of plant images in open-world data streams, one related to bird sounds identification in soundscapes and one related to the visual-based identification of fish species and whales individuals. More than 150 research groups registered to at least one of the challenges and about 15 of them crossed the finish lines by running their system on the final test data. A synthesis of the results is published in the LifeCLEF 2016 overview paper [38] and more detailed analyses are provided in research reports for the plant task [35] and the bird task [36].

7.5.5. Crowdsourcing Thousands of Specialized Labels using a Bayesian Approach

Participants: Maximilien Servajean, Alexis Joly, Dennis Shasha, Julien Champ, Esther Pacitti.

Large-scale annotated corpora are often at the basis of huge performance gaps in machine learning based content analysis. However, the availability of such datasets has only been made possible thanks to the great amount of human labeling efforts leveraged by popular crowdsourcing and social media platforms. When the labels correspond to well known concepts, it is straightforward to train the annotators by giving a few examples with known answers. It is also straightforward to judge the quality of their labels. But neither is true with thousands of complex domain specific labels. Training on all labels is infeasible and the quality of an annotator's judgements may be vastly different for some subsets of labels than for others. This paper proposes a set of data-driven algorithms to (i) train annotators on how to disambiguate automatically labelled images, (ii) evaluate the quality of annotators' answers on new test items and (iii) weight predictions. The algorithms adapt to the skills of each annotator both in the questions asked and the weights given to their answers. The underlying judgements are Bayesian, based on adaptive priors. We measure the benefits of these algorithms by a live user experiment related to image-based plant identification involving around 1,000 people [47] (at the origin of ThePlantGame, see Software section). The proposed methods yield huge gains in annotation accuracy. While a standard user could correctly label around 2% of our data, this goes up to 80% with machine learning assisted training and almost 90% when doing a weighted combination of several annotators' labels.

ALICE Project-Team

7. New Results

7.1. Hexahedral-dominant Remeshing

Participants: Dmitry Sokolov, Nicolas Ray, Bruno Lévy, Maxence Reberol

Representing the geometry of complex objects in a computer is usually achieved by a mesh: the object is decomposed in cells that have a simple geometry. Each cell is defined by a set of facets. The simplest choice is to use meshes with tetrahedral cells that are relatively easy to produce and to work with. However, some applications involving numerical simulations better work with hexahedral cells. Such hexahedral meshes are very difficult to produce, even when it is completely done by a designer. Our objective is to relax the intrinsic difficulties of full hexahedral remeshing by allowing the process to generate a few tetrahedra in the hexahedral mesh (hexahedral-dominant meshes). Our approach is to produce as many hexahedra as possible by filling most of the volume with a deformed 3D grid, and to rely on more classic meshing techniques everywhere else. We also develop tools to evaluate how our remeshing impacts results of FEM simulations.

7.1.1. Generation of Hexahedral-dominant Meshes

The traditional approach to produce hexahedral dominant meshes is by advancing front: first hexahedra are produced near the object boundary, then additionals hexahedra are attached to them. An alternative solution is to consider an hexahedral mesh as a deformed 3D grid: the hexahedral remeshing problem is then restated as finding the (geometric) deformation that will transform the hexahedral mesh into the regular grid. This approach is able to generate very good hexhaderal meshes, but it is often impossible to entirely remesh the input object.

Our objective is to produce hexahedra from the mapping approach, then complete the mesh with traditional approaches that may leave some tetrahedra. We proposed a first solution [9]: we compute a mapping, extract the vertices of the deformed 3D grid, generate a tetrahedral mesh having these vertices, then merge sets of tetrahedra into hexahedra with an extension of [25]. Using the mapping as a heuristic made this solution very competitive with other hexahedral dominant methods. We are now developing a software pipeline that makes it easy for different algorithms (frame field, mapping and classic remeshing) to work together. With a simple implementation of each step, we already observe better performances than previous works, and we foresee many opportunities to improve it.

7.1.2. Impact on FEM Performance

It is admitted by our scientific community that hexahedral meshes are better than tetrahedral meshes for some FEM simulation. We would like to demonstrate evidence of this belief, including fair comparisons with equal running time and/or result accuracy, with the best function basis for each case. For hexahedral dominant meshes, we have developed a new specific function basis devoted to properly link tetrahedral and hexahedral elements. Using a combination of tri-linear and quadratic hexahedra, we can build an approximation space made of continuous functions, even at non-conforming interfaces between hexahedra and tetrahedra. But in practice, hexahedral-dominant meshes are mainly useful to mesh complicated 3D domains. In such cases, there are no analytical solutions of partial differential equations and thus it is not straightforward to evaluate the accuracy of a new numerical method. To measure the differences between finite element solutions defined on different meshes of the same 3D model, we are developing a fast and efficient sampling method which exploits GPU hardware. These topics are addressed in the (ongoing) Ph.D. thesis of Maxence Reberol.

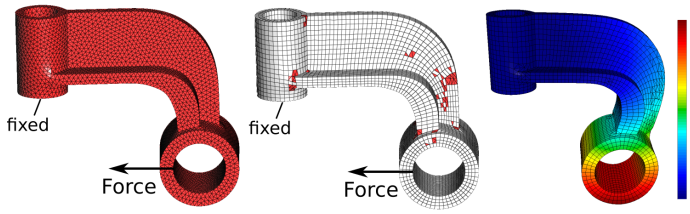


Figure 2. Mechanical problem on the Hanger 3D model. (left) Sandard tetrahedral mesh. (center) Our hex-dominant mesh, hex in gray and tet in red. (right) Solution of the problem with mixed hexahedral-tetrahedral finite elements, color is the amplitude of the displacement.

7.2. Optimal transport

Participant: Bruno Lévy

7.2.1. Optimal transport:

Optimal Transport is not only a fundamental problem with a rich structure, but also a new computational tool, with many possible applications. To name but a few, applications of Optimal Transport comprise image registration, reflector and refractor design, histogram interpolation, artificial intelligence. In astrophysics, it is used by Early Universe Reconstruction, a difficult inverse problem that reconstructs the time evolution of the universe from the observed current state. It can be also used in meteorology, to simulate certain phenomena (semi-geostrophic currents). It is also the main component of solvers for certain equations, based on a variational formulation that leads to a gradient flow. All these applications and future developments depend on a single component: an efficient solver for the Monge-Ampère equation. We developed a new algorithm that overcome by several order of magnitude the speed of the classical "auction algorithm" (that solves in $O(n \log(n))$ a discrete version of the problem). The *semi-discrete* version of the problem that we study can be solved by extremizing a smooth objective function, thus a significantly faster speed is obtained as compared to the previous combinatorial algorithm. This year we improved our Quasi-Newton solver and replaced it with a Full-Newton solver, that gains one additional order of magnitude in speed, and we can solve semi-discrete problems with 1 million Dirac masses in a matter of minutes. We also experimented with applications of this solver to fluid simulation. Last winter (December 2015) Wenping Wang visited Nancy, and we discussed several ideas on Optimal Transport. We proposed together this year (2016) a new method to sample a surface with a power diagram [31]. The positions of the samples are optimized by a criterion similar to centroidal Voronoi tessellations, and the associated weights are used to control the areas of the power cells with prescribed values. We give the expressions of the derivatives of the combined objective function, and propose a quasi-Newton algorithm to optimize it. We describe several applications of the algorithm.

7.3. Spectral Clustering of Plant Units From 3D Point Clouds

Participant: Dobrina Boltcheva

High-resolution terrestrial Light Detection And Ranging (tLiDAR), a 3-D remote sensing technique, has recently been applied for measuring the 3-D characteristics of vegetation from grass to forest plant species. The resulting data are known as a point cloud which shows the 3-D position of all the hits by the laser beam giving a raw sketch of the spatial distribution of plant elements in 3-D, but without explicit information on their geometry and connectivity. We have developed a new approach based on a delineation algorithm that clusters

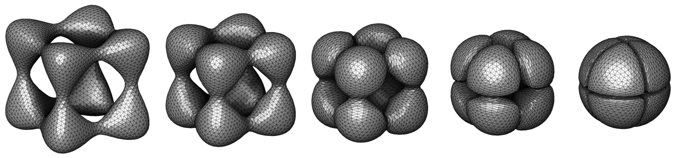


Figure 3. Semi-discrete optimal transport between a shape and a sphere, computed by our algorithm

a point cloud into elementary plant units such as internodes, petioles and leaves. The algorithm creates a graph (points + edges) to recover plausible neighboring relationships between the points and embeds this graph in a spectral space in order to segment the point-cloud into meaningful elementary plant units. This work has been published in the International Journal of Remote Sensing [6].

7.4. Surface Reconstruction From 3D Point Clouds

Participants: Dobrina Boltcheva, Bruno Lévy

We have developed a practical reconstruction algorithm that generates a surface triangulation from an input pointset. In the result, the input points appear as vertices of the generated triangulation. The algorithm has several desirable properties: it is very simple to implement, it is time and memory efficient, and it is trivially parallelized. On a standard hardware (core i7, 16Gb) it takes less than 10 seconds to reconstruct a surface from 1 million points, and scales-up to 36 million points (then it takes 350 seconds). On a smartphone (ARMV7 Neon, quad core), it takes 55 seconds to reconstruct a surface from 900K points. The algorithm computes the Delaunay triangulation of the input pointset restricted to a "thickening" of the pointset (similarly to several existing methods, like alpha-shapes, crust and co-cone). By considering the problem from the Voronoi point of view (rather than Delaunay), we use a simple observation (radius of security) that makes the problem simpler. The Delaunay triangulation data structure and associated algorithms are replaced by simpler ones (KD-Tree and convex clipping) while the same set of triangles is provably obtained. The restricted Delaunay triangulation can thus be computed by an algorithm that is not longer than 200 lines of code, memory efficient and parallel. The so-computed restricted Delaunay triangulation is finally post-processed to remove the non-manifold triangles that appear in regions where the sampling was not regular/dense enough. Sensitivity to outliers and noise is not addressed here. Noisy inputs need to be pre-processed with a pointset filtering method. In the presented experimental results, we are using two iterations of projection onto the best approximating plane of the 30 nearest neighbors (more sophisticated ones may be used if the input pointset has many outliers).

This work has been published in the Research Report [13] and is currently in revision for Eurographics 2017.

7.5. Microstructures for additive manufacturing

Participants: Jonas Martinez, Sylvain Lefebvre

Nowadays, there is a big interest in the functional optimization of microstructures for additive manufacturing, as reflected by the high number of recent publications on the subject. This also comes not only from research but also industry, as controlling the macroscopic elasticity of materials has a wide range of industrial applications. For instance, to fabricate flexible prosthetic body parts, or to produce rigid but porous prosthetics for surgery. In particular, controlling material elasticity will enable the design of lightweight and resistant materials, and in turn, reduce material consumption.

Most of the existing software either optimize for periodic tilings of microstructures, or generate random microstructures without precise control of their functionality. We recently introduced a method [7] to generate stochastic structures (figure 4) while having unique computational advantages, and precisely controlling their functionality. Our optimization approach of stochastic porous structures deviates significantly from both the periodic tiling of microstructures and the optimization of macrostructures, by making a link between microstructures, and procedural solid textures with controlled statistics in Computer Graphics. We believe there are many other such structures left to be discovered, and hope our work will spark further interest in procedurally generated, stochastic microstructures.

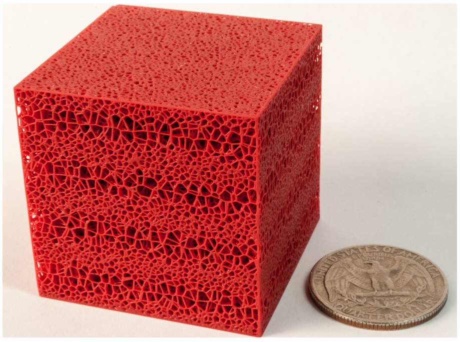


Figure 4. Anisotric microstructures

7.6. Towards Zero-Waste Furniture Design

Participants: Bongjin Koo, Jean Hergel, Sylvain Lefebvre, Niloy J. Mitra.

This project considered the optimization of parametric models of furniture to reduce the wastage of material used to fabricate the model. Our approach uses a 2D packing algorithm to pack the different parts of the furniture in a wooden plank. Then we optimize locally the wastage by editing smoothly the parameters with only moving smoothly the parts in the packing space. We produced full size objects with laser cutter to prove the efficiency of our method. This work has been accepted in Transaction on Visualization and Computer Graphics.

7.7. Anti-aliasing for fused filament deposition

Participants: Hai-Chuan Song, Nicolas Ray, Dmitry Sokolov, Sylvain Lefebvre

Layered manufacturing inherently suffers from staircase defects along surfaces that are gently sloped with respect to the build direction. Reducing the slice thickness improves the situation but never resolves it completely as flat layers remain a poor approximation of the true surface in these regions. In addition, reducing the slice thickness largely increases the print time. In this project we focus on a simple yet effective technique to improve the print accuracy for layered manufacturing by filament deposition. Our method [16] works with standard three-axis 3D filament printers (e.g. the typical, widely available 3D printers), using standard extrusion nozzles. It better reproduces the geometry of sloped surfaces without increasing the print time. Our key idea is to perform a local anti-aliasing, working at a sub-layer accuracy to produce slightly curved deposition paths and reduce approximation errors. We show that the necessary deviation in height compared to standard slicing is bounded by half the layer thickness. Therefore, the height changes remain small and plastic deposition remains reliable. We further split and order paths to minimize defects due to the extruder nozzle shape, avoiding any change in the existing hardware. We apply and analyze our approach on 3D printed examples, showing that our technique greatly improves surface accuracy and silhouette quality while keeping the print time nearly identical.

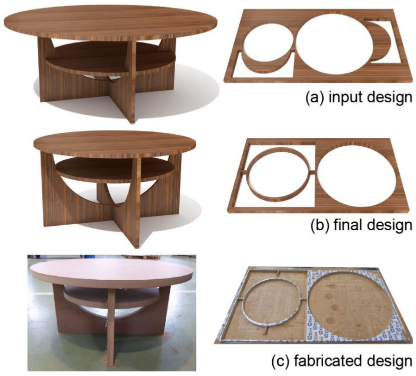


Figure 5. Our technique modifies the parameters of the input design (a) to improve the packing and waste less material (b). The produced furniture is shown in (c).

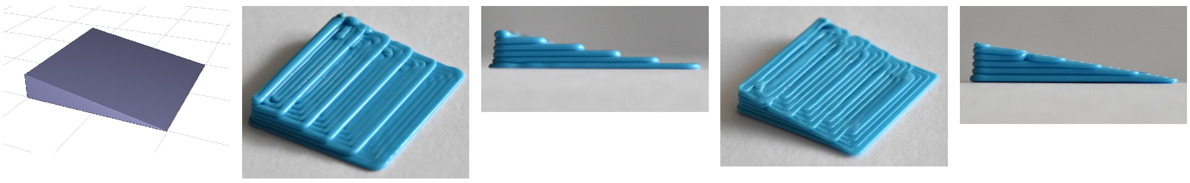


Figure 6. Printing a wedge model clearly reveals the staircase defects that plague 3D printing. (a) Input 3D model; the bottom edge length is 20mm and the angle of the incline plane is 10. (b) Global view and (c) side view of a standard, flat layer printed result. (d) Global view and (e) side view of our anti-aliased printed result, revealing the improvement in surface accuracy and silhouette smoothness

AVIZ Project-Team

7. New Results

7.1. Swarm User Interfaces

Participants: Mathieu Le Goc [correspondant], Lawrence Kim, Ali Parsaei, Jean-Daniel Fekete, Pierre Dragicevic, Sean Follmer.

We introduce swarm user interfaces (Fig 3), a new class of human-computer interfaces comprised of many autonomous robots that handle both display and interaction. We describe the design of Zooids, a hardware and software system: a small wheel-propelled robot with position and touch sensing capabilities that can be freely arranged and repositioned on any horizontal surface, both through user manipulation and computer control. Zooids is an open-source open-hardware platform for developing tabletop swarm interfaces. We illustrate the potential of tabletop swarm user interfaces through a set of application scenarios developed with Zooids, and discuss general design considerations unique to swarm user interfaces.

More on the project Web page: http://www.aviz.fr/swarmui.

7.2. A Systematic Review of Experimental Studies on Data Glyphs

Participants: Johannes Fuchs, Petra Isenberg [correspondant], Anastasia Bezerianos, Daniel Keim.

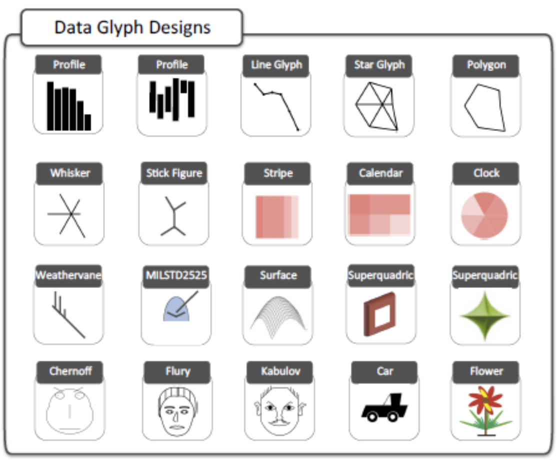


Figure 9. Overview of the glyphs reviewed in the study.

We systematically reviewed 64 user-study papers on data glyphs to help researchers and practitioners gain an informed understanding of tradeoffs in the glyph design space. The glyphs we considered were individual representations of multi-dimensional data points, often meant to be shown in small-multiple settings. Over the past 60 years many different glyph designs were proposed and many of these designs have been subjected to perceptual or comparative evaluations. Yet, a systematic overview of the types of glyphs and design variations tested, the tasks under which they were analyzed, or even the study goals and results did not yet exist. We provide such an overview by systematically sampling and tabulating the literature on data glyph studies, listing their designs, questions, data, and tasks. In addition we present a concise overview of the types of glyphs and their design characteristics analyzed by researchers in the past, and a synthesis of the study results. Based on our meta analysis of all results we further contribute a set of design implications and a discussion on open research directions.

7.3. Towards an Understanding of Mobile Touch Navigation in a Stereoscopic Viewing Environment for 3D Data Exploration

Participants: David López, Lora Oehlberg, Candemir Doger, Tobias Isenberg [correspondant].

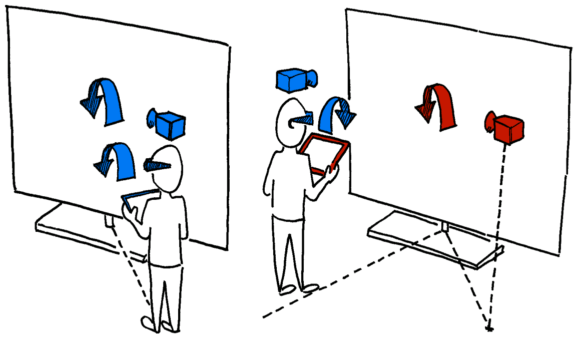


Figure 10. Illustration of the problem of mobility within a virtual environment, while interacting with a view on a tablet.

We discuss touch-based navigation of 3D visualizations in a combined monoscopic and stereoscopic viewing environment. We identify a set of interaction modes, and a workflow that helps users transition between these modes to improve their interaction experience. In our discussion we analyze, in particular, the control-display space mapping between the different reference frames of the stereoscopic and monoscopic displays. We show how this mapping supports interactive data exploration, but may also lead to conflicts between the stereoscopic and monoscopic views due to users' movement in space; we resolve these problems through synchronization. To support our discussion, we present results from an exploratory observational evaluation with domain experts in fluid mechanics and structural biology. These experts explored domain-specific datasets using variations of a system that embodies the interaction modes and workflows; we report on their interactions and qualitative feedback on the system and its workflow.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Lopez2016TUM.

7.4. CAST: Effective and Efficient User Interaction for Context-Aware Selection in 3D Particle Clouds

Participants: Lingyun Yu, Konstantinos Efstathiou, Petra Isenberg, Tobias Isenberg [correspondant].

We present a family of three interactive Context-Aware Selection Techniques (CAST) for the analysis of large 3D particle datasets. For these datasets, spatial selection is an essential prerequisite to many other analysis tasks. Traditionally, such interactive target selection has been particularly challenging when the data subsets of interest were implicitly defined in the form of complicated structures of thousands of particles. Our new techniques SpaceCast, TraceCast, and PointCast improve usability and speed of spatial selection in point clouds through novel context-aware algorithms. They are able to infer a user's subtle selection intention from gestural input, can deal with complex situations such as partially occluded point clusters or multiple cluster layers, and can all be fine-tuned after the selection interaction has been completed. Together, they provide an effective and efficient tool set for the fast exploratory analysis of large datasets. In addition to presenting Cast, we report on a formal user study that compares our new techniques not only to each other but also to existing state-of-the-art selection methods. Our results show that Cast family members are virtually always faster than existing methods without tradeoffs in accuracy. In addition, qualitative feedback shows that PointCast and TraceCast were strongly favored by our participants for intuitiveness and efficiency.

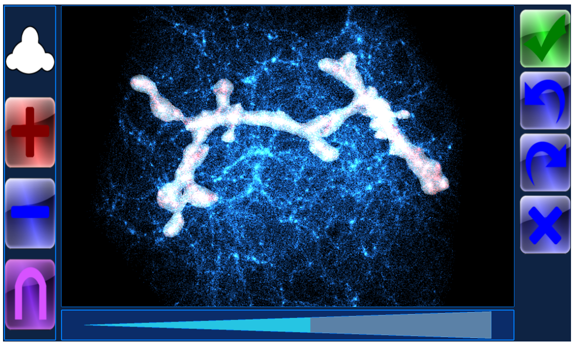


Figure 11. Illustration of the complexity metrics used in the study.

More on the project Web page: https://tobias.isenberg.cc/VideosAndDemos/Yu2016CEE.

7.5. Hybrid Tactile/Tangible Interaction for 3D Data Exploration

Participants: Lonni Besançon [correspondant], Paul Issartel, Mehdi Ammi, Tobias Isenberg.

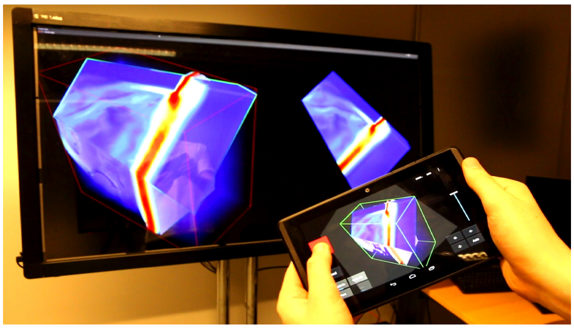


Figure 12. Picture of the hybrid interaction system.

We present the design and evaluation of an interface that combines tactile and tangible paradigms for 3D visualization. While studies have demonstrated that both tactile and tangible input can be efficient for a subset of 3D manipulation tasks, we reflect here on the possibility to combine the two complementary input types. Based on a field study and follow-up interviews, we present a conceptual framework of the use of these different interaction modalities for visualization both separately and combined—focusing on free exploration as well as precise control. We present a prototypical application of a subset of these combined mappings for fluid dynamics data visualization using a portable, position-aware device which offers both tactile input and tangible sensing. We evaluate our approach with domain experts and report on their qualitative feedback.

More on the project Web page: http://lonni.besancon.pagesperso-orange.fr/Projects/HybridInteraction/HybridInteraction.html.

7.6. A Tangible Volume for Portable 3D Interaction

Participants: Paul Issartel, Lonni Besançon [correspondant], Tobias Isenberg, Mehdi Ammi.

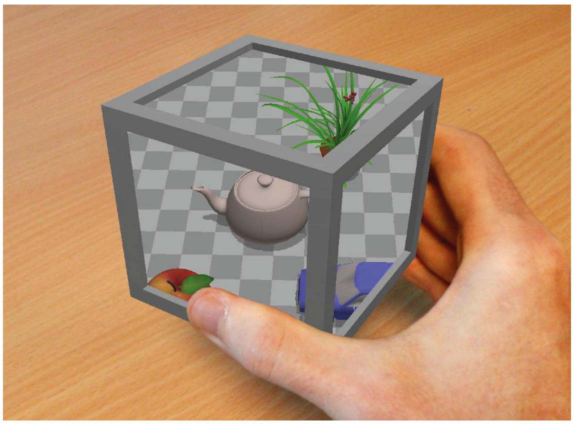


Figure 13. Image of the Cube.

We present a new approach to achieve tangible object manipulation with a single, fully portable and self-contained device. Our solution is based on the concept of a “tangible volume”. We turn a tangible object into a handheld fish-tank display. The tangible volume represents a volume of space that can be freely manipulated within a virtual scene. This volume can be positioned onto virtual objects to directly grasp them, and to manipulate them in 3D space. We investigate this concept through two user studies. The first study evaluates the intuitiveness of using a tangible volume for grasping and manipulating virtual objects. The second study evaluates the effects of the limited field of view on spatial awareness. Finally, we present a generalization of this concept to other forms of interaction through the surface of the volume.

More on the project Web page: http://lonni.besancon.pagesperso-orange.fr/Projects/TangibleCube/TangibleCube.html.

7.7. Preference Between Allocentric and Egocentric 3D Manipulation in a Locally Coupled Configuration

Participants: Paul Issartel, Lonni Besançon [correspondant], Steven Franconeri.

We study user preference between allocentric and egocentric 3D manipulation on mobile devices, in a configuration where the motion of the device is applied to an object displayed on the device itself. We first evaluate this preference for translations and for rotations alone, then for full 6-DOF manipulation. We also investigate the role of contextual cues by performing this experiment in different 3D scenes. Finally, we look at the specific influence of each manipulation axis. Our results provide guidelines to help interface designers select an appropriate default mapping in this locally coupled configuration.

More on the project Web page: http://lonni.besancon.pagesperso-orange.fr/Projects/Mappings/Mappings.html.

7.8. Embedded Data Representations

Participants: Wesley Willett, Yvonne Jansen, Pierre Dragicevic [correspondant].

We introduces *embedded data representations* as the use of visual and physical representations of data that are deeply integrated with the physical spaces, objects, and entities to which the data refers [16]. Technologies like lightweight wireless displays, mixed reality hardware, and autonomous vehicles are making it increasingly easier to display data in-context. While researchers and artists have already begun to create embedded data representations, the benefits, trade-offs, and even the language necessary to describe and compare these approaches remain unexplored.

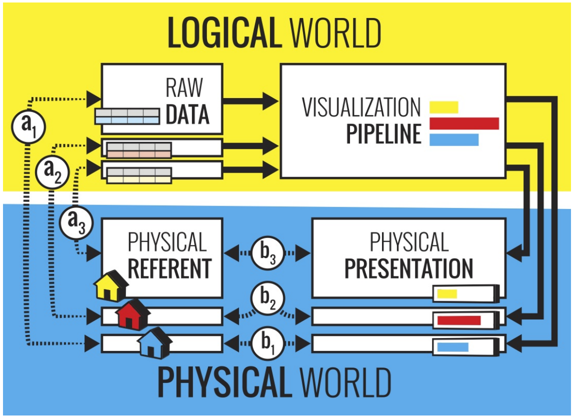


Figure 14. Conceptual model for situated and embedded data representations.

In this paper, we formalize the notion of physical data referents – the real-world entities and spaces to which data corresponds – and examine the relationship between referents and the visual and physical representations of their data. We differentiate situated representations, which display data in proximity to data referents, and embedded representations, which display data so that it spatially coincides with data referents. Drawing on examples from visualization, ubiquitous computing, and art, we explore the role of spatial indirection, scale, and interaction for embedded representations. We also examine the tradeoffs between non-situated, situated, and embedded data displays, including both visualizations and physicalizations. Based on our observations, we identify a variety of design challenges for embedded data representation, and suggest opportunities for future research and applications

7.9. Space-Time Cube Framework

Participants: Benjamin Bach, Pierre Dragicevic [correspondant], Dominique Archambault, Christophe Hurter, Sheelagh Carpendale.

We presented a descriptive model for visualizations of temporal data based on a generalized space-time cube framework [1]. Visualizations are described as operations on a conceptual space-time cube, which transform the cube's 3D shape into readable 2D visualizations. Operations include: extracting subparts of the cube, flattening it across space or time, or transforming the cube's geometry and content. We introduced a taxonomy of elementary space-time cube operations and explained how these operations can be combined and parameterized.

The generalized space-time cube has two properties: a) it is purely conceptual without the need to be implemented, and b) it applies to all datasets that can be represented in two dimensions plus time (e.g., geospatial, videos, networks, multivariate data). The proper choice of space-time cube operations depends on many factors, e.g., density or sparsity of a cube, hence we proposed a characterization of structures within space-time cubes, which allowed us to discuss strengths and limitations of operations. We also reviewed interactive systems that support multiple operations, allowing a user to customize his view on the data. With this framework, we hope to facilitate the description, criticism and comparison of temporal data visualizations, as well as encourage the exploration of new techniques and systems.

More on the project Web page: spacetimecubevis.com.

7.10. The Attraction Effect in Information Visualization

Participants: Evanthia Dimara [correspondant], Anastasia Bezerianos, Pierre Dragicevic.

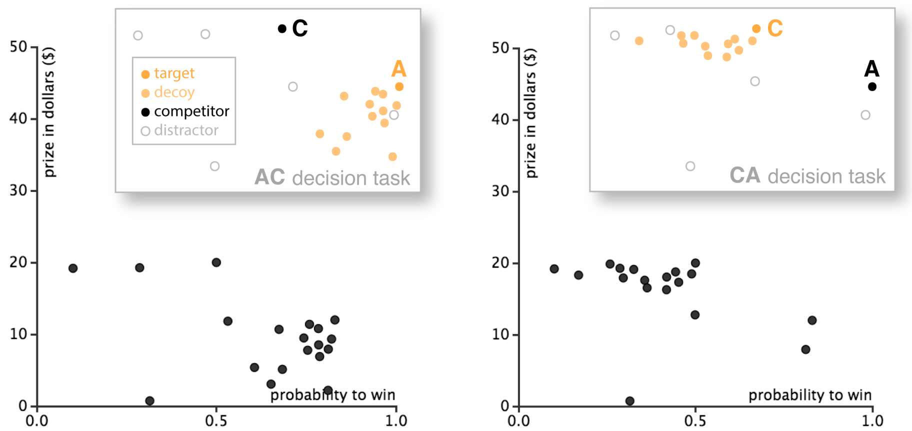


Figure 15. Illustration of the extension of the attraction effect in large datasets. Example of two matched decision tasks AC and CA in scatterplots.

The attraction effect is a well-studied cognitive bias in decision making research, where one's choice between two alternatives is influenced by the presence of an irrelevant (dominated) third alternative. We examine whether this cognitive bias, so far only tested with three alternatives and simple presentation formats such as numerical tables, text and pictures, also appears in visualizations. Since visualizations can be used to support decision making — e.g., when choosing a house to buy or an employee to hire — a systematic bias could have important implications. In a first crowdsource experiment, we indeed partially replicated the attraction effect with three alternatives presented as a numerical table, and observed similar effects when they were presented as a scatterplot. In a second experiment, we investigated if the effect extends to larger sets of alternatives, where the number of alternatives is too large for numerical tables to be practical. Our findings indicate that the bias persists for larger sets of alternatives presented as scatterplots. We discuss implications for future research on how to further study and possibly alleviate the attraction effect.

More on the project Web page: www.aviz.fr/decoy.

7.11. An Exploratory Study of Word-Scale Graphics in Data-Rich Text Documents

Participants: Pascal Goffin [correspondant], Jeremy Boy, Wesley Willett, Petra Isenberg.

We investigated the design and function of word-scale graphics and visualizations embedded in text documents. Word-scale graphics include both data-driven representations such as word-scale visualizations and sparklines, and non-data-driven visual marks. There has been little research attention on their design, function, and use so far. We present the results of an open ended exploratory study with nine graphic designers. The study resulted in a rich collection of different types of graphics, data provenance, and relationships between text, graphics, and data. Based on this corpus, we present a systematic overview of word-scale graphic designs, and examine how designers used them. We also discuss the designers' goals in creating their graphics, and characterize how they used word-scale graphics to visualize data, add emphasis, and create alternative narratives. We discuss implications for the design of authoring tools for word-scale graphics and visualizations building on these examples, and explore how new authoring environments could make it easier for designers to integrate them into documents.

Figure 16. Examples of word-scale graphics collected during the study.

EX-SITU Team

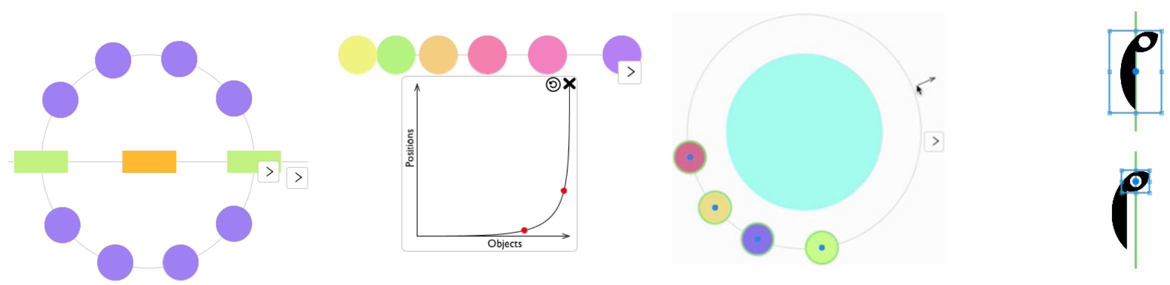
7. New Results

7.1. Fundamentals of Interaction

Participants: Sarah Fdili Alaoui, Michel Beaudouin-Lafon, Cédric Fleury, Wendy Mackay, Theophanis Tsandilas.

In order to better understand fundamental aspects of interaction, ExSitu studies interaction in extreme situations. We conduct 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.

StickyLines – Aligning and distributing graphical objects is a common, but cumbersome task. We studied graphic designers and regular users and identified three key problems with current tools: lack of persistence, unpredictability of the results, and inability to 'tweak' the layout. We created *StickyLines* [14], a tool that reifies guidelines into first-class objects: Users can create precise, predictable and persistent interactive alignment and distribution relationships, and can 'tweak' the alignment in a way that can be maintained for subsequent interactions (Figure 2). We ran a [2x2] within-participant experiment to compare *StickyLines* with standard commands and found that *StickyLines* performed up to 40% faster and required up to 50% fewer actions than traditional alignment and distribution commands for complex layouts. Finally, we gave *StickyLines* to six professional designers and found that not only did they quickly adopt it, they also identified novel uses, including creating complex compound guidelines and using them for both spatial and semantic grouping. This work demonstrate the power of reifying concepts, such as alignment and distribution, into first-class objects that can be directly manipulated and appropriated by end users.



(a) (b) (c) (d)

Figure 2. StickyLines reify alignment and distribution into first-class graphical objects that users can manipulate directly. (a) Circular and horizontal alignments. (b) Non-linear distribution. (c) Ghost guideline. (d) Tweaking an object's bounding box. .

UIST Video Browser – We created an interactive video browser that provides a rapid overview of the 30-second video previews of the ACM UIST conference papers, based on the conference schedule [16]. The web application was made available to the 600+ conference attendees, who could see an overview of upcoming talks, search by topic, and create personalized, shareable video playlists that capture the most interesting or relevant papers. Reifying playlists into first-class objects and applying instrumental interaction concepts helped create a fluid and efficient interface.

In(SITE) – We explored touch-based 3D interaction in the situation where users are immersed in a 3D virtual environment and move in front of a large multi-touch wall-sized display. We designed In(SITE) [20], a bimanual touch-based technique combined with object teleportation features which enables users to perform 3D object manipulation on a large vertical display (Figure 3). This technique was compared with a standard 3D interaction technique. The results showed that participants can reach the same level of performance for completion time and a better precision for fine adjustments with the In(SITE) technique. They also revealed that combining object teleportation with both techniques improves translation tasks in terms of ease of use, fatigue, and user preference.

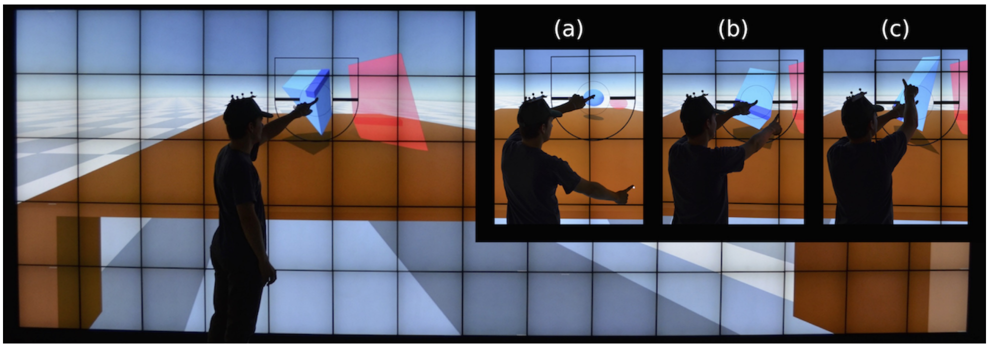


Figure 3. 3D manipulation on a multi-touch wall-sized display combining bimanual interaction and teleportation. The user is performing a xy translation (main pict.), z translation (a), roll rotation (b), and pitch & yaw rotation(c).

In collaboration with Inria Lille (MJOLNER group) and Univ. Strasbourg, we applied our design principles for instrumental interaction to create new interactive tools for the parallelization of programs, a highly specialized task that is currently done by expert developers. Current programming models, languages and tools do not help developers restructure existing programs for more effective execution. At the same time, automatic approaches are overly conservative and imprecise to achieve sufficient performance. We introduced *interactive program restructuring* [28], [11] to bridge the gap between semi-automatic program manipulation and software visualization. First, we extended a state-of-the-art polyhedral model for program representation so that it supports high-level program manipulation. Based on this model, we designed and evaluated a direct manipulation visual interface for program restructuring. This interface provides information about the program that was not immediately accessible in the code and allows to manipulate programs without rewriting code. By providing a visual and textual representation of an automatically computed program optimization that is easily modifiable and reusable by the developer, we create a sort of human-machine partnership where the developer can better take advantage of the power of the machine. An empirical study of developers using this tool showed the value of program manipulation tools based on the instrumental interaction paradigm. This work illustrates how the combination of our conceptual approaches, namely instrumental interaction and human-computer partnership, can benefit extreme users such as developers of parallel programs.

Finally, we reviewed statistical methods for the analysis of user-elicited gestural vocabularies [24]. We showed that measures currently used to assess agreement between participants of a gesture elicitation study are problematic. We discussed the problem of chance agreement and showed how it can bias results. We reviewed chance-corrected agreement coefficients that are routinely used in inter-reliability studies and showed how to apply them to gesture elicitation studies. We also discussed how to compute interval estimates for these coefficients and how to use them for statistical inference.

7.2. Partnerships

Participants: Wendy Mackay, Jessalyn Alvina, Ghita Jalal, Joseph Malloch, Nolwenn Maudet.

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 algortithm, 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 sytstems' that are discoverable, appropriable and expressive for the user. *Interactive program restructuring* [28] offers a concrete example, where expert programmers interact with dynamic visualisations of parallel programs to better understand and organize their code. Similarly, tools such as *Color Partner* generate color suggestions based the users input, helping the user guide their discovery of new color possibilities, and *Linkify* helps users create rules to define how visual properties should change under different user contexts (see Jalal's dissertation).

We hosted the 30-person *ERC CREATIV* workshop in Paris, to explore our concepts of *Co-adaptive Systems* (including human-centered machine learning); and *Instrumental Interaction* (including substrates) with prominent researchers from Stanford University, New York University, University of Aarhus, Goldsmiths College, University of Toulouse, IRCAM, University of British Columbia, UC San Diego, and UC Berkeley. Our long-term, admittedly ambitious, goal is to create a unified theory of interaction grounded in how people interact with the world. Our principles of co-adaptive systems and instrumental interaction offer a generative approach for supporting creative activities, from early exploration to implementation. The workshop launched several research projects that are currently in progress or will be published in 2017.

Human-Centred Machine Learning:

We begin by challenging some of the standard assumptions surrounding Machine Learning, clearly one of the most important and successful techniques in contemporary computer science. It involves the statistical inference of models (such as classifiers) from data. However, all too often, the focus is on impersonal algorithms that work autonomously on passively collected data, rather than on dynamic algorithms that progressively reveal their progress to support human users. We collaborated on a workshop at the CHI 2017 conference, entitled "Human-centred Machine Learning" [15] with colleagues from Ircam, Goldsmiths College, and Microsoft Research. We seek a different understanding of the 'human-in-the-loop', where the focus is less on the human user as input to an algorithm, but rather as an algorithm in service of a human user. Examining machine learning from a human-centred perspective includes explicitly recognising human work in the creation of these algorithms, as well as the situated use these algorithms by human work practices. A human-centred understanding of machine learning in human context can lead not only to more usable machine learning tools, but to new ways of framing learning computationally.

Supporting Expressivity:

We helped organize and participated in a workshop at CHI 2017 *Human Computer Interaction meets Computer Music* [27], where we described the results of the MIDWAY Equipe Associé project (with McGill University, Ex-Situ and the *MINT* EP at Inria, Lille.) We presented results of our extensive research with contemporary music composers, in particular our strategy for developing 'co-adaptive instruments'. This involves a paradigm shift, where the goal of the technology is not necessarily the accuracy of a particular result, but rather, the human user's ability to express themselves through the technology.

We also explored the idea of rethinking the use of machine learning to support human-computer partnerships for everyday interaction. We built on gesture-typing, which offers users an efficient, easy-to-learn, and error-tolerant technique for producing typed text on a soft keyboard. Our focus was not on improving recognition accuracy, which we take as a given, but rather on how to make gesture-typed output more expressive. Experiment 1 demonstrated that users vary word gestures according to instructions (accurately, quickly or creatively) as well as specific characteristics of each word, including length, angle, and letter repetition. We show that users produce highly divergent gestures, with three easily detectable characteristics: curviness, size, and speed. We created the Expressive Keyboard [10] which maps these characteristics to color variations, thus allowing users to control both the content and the color of gesture-typed words (Figure 4). Experiment 2 demonstrates that users can successfully control their gestures to produce the desired colored output, and find it easier to react to visual feedback than explicitly controlling the characteristics of each gesture. Expressive keyboards can map gestural input to any of a variety of output characteristics, such as personalized handwriting and dynamic emoticons, to let users transform gesture variation into expressivity, without sacrificing accuracy.

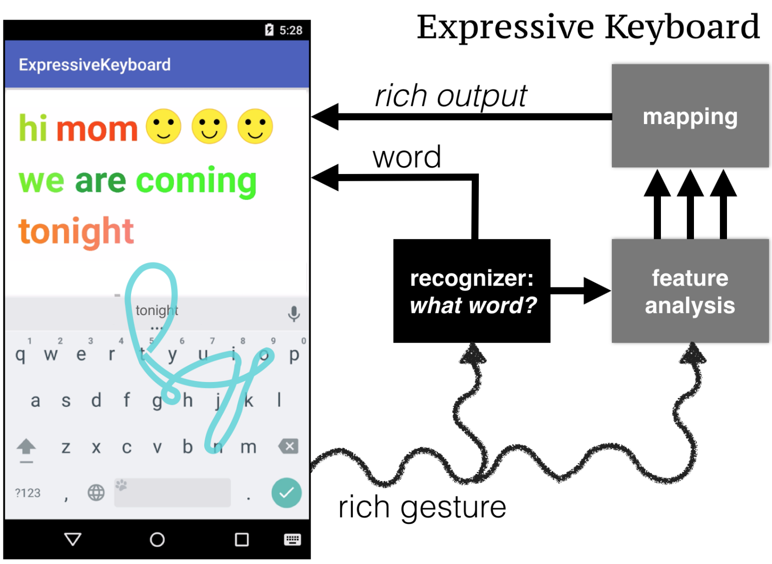


Figure 4. Expressive Keyboards produce accurate words, but also let users control multiple expressive output properties.

7.3. Creativity

Participants: Sarah Fdili Alaoui, Michel Beaudouin-Lafon, Ghita Jalal, Wendy Mackay, Joseph Malloch, Nolwenn Maudet, Michael Wessely, Theophanis Tsandilas.

ExSitu is interested in understanding the work practices of creative professionals, particularly artists, designers, and scientists, who push the limits of interactive technology.

We explore how concepts of substrate and co-adaptation can change how we design interactive technology for supporting creativity. Co-adaptation is the phenomenon in which users both adapt their behavior to the system's constraints, and appropriate the system for their own needs. We explore these concepts using participatory design studies in creative contexts with expert and non-expert users. We study structuring layouts for graphic designers, sketching movement for choreographers, expressive movements for dancers and further explore expressive gesture of non-experts on mobile devices and possible interactions on hybrid stretchable interfaces. These studies require a multi-disciplinary design team that works closely with users throughout the design process. We create situations that cause users to reflect deeply about their activities in context and work with them to articulate the design problem. The experiments, prototypes and systems that we developed and deployed are illustrated below:

Graphic design: Our studies of the creative design practices of professional graphic designers show that designers appropriate visual properties of existing tools to create their own personal 'instruments'. Unfortunately, most professional design tools make this difficult: At best, they provide only indirect access, through property sheets or dialog boxes, to visual properties, such as color and style, rather treating them as as independent interactive objects. We developed a number of composition tools that demonstrate how to explicitly reify visual properties, using the concept of co-adaptive instruments. Ghita Jalal successfully defended her doctoral dissertation on this topic (see [9]).

We also examined artists' and designers' practices as they manipulate color and create layouts in their projects. We found that artists and designers select colors from personal representations. They manipulate color in the context of its surrounding graphical elements, and combine it with other visual properties such as texture. As they create their layouts, designers establish links among visual properties such as size, position, and layering of graphical elements. They define rules for how these properties change in space, across instances of the same composition, or in time, across related compositions. We also found that designers prefer tools that provide direct access to visual properties.

Choreography: We are interested in designing choreographic support tools because choreographers rarely have access to interactive tools that are designed specifically to support their creative process [13]. In order to design for such a technology, we interviewed six contemporary choreographers about their creative practice. We found that even though each process is unique, choreographers represent their ideas by applying a set of operations onto choreographic objects. Throughout different creative phases, choreographers compose by shifting among various degrees of specificity and vary their focal points from dancers to stage, to interaction, to the whole piece. Based on our findings, we presented a framework for articulating the higher-level patterns that emerge from these complex and idiosyncratic processes. We then articulated the resulting implications for the design of interactive tools to support the choreographic practice.

On generating choreographic ideas, we developed the Choreographer's Workbench, a full-body interactive system that aims to help choreographers explore and design dance movements during the ideation phase by creating a link between past recorded movement ideas and revealing their underlying relationships. The system explores how to increase the discoverability and appropriateness of movement ideas via feedforward visualization of movement characteristics.

Figure 5. Passersby interact with the animated Père Noël, first mirrors their behavior and then shapes it.

We collaborated with the N+1 theater group on the "Grande Vitrine" art and science project, an interactive installation that takes place during the month of Christmas. It consisted of a virtual animated character with whom participants interact and a physical kinetic sculpture whose motions are triggered by participant interaction (Figure 5). The participant were expected to perform full-body movements and figure out the correct one that will help the animated character escape from the virtual screen into the physical motorized display. The installation tested the concept of ”shaping” from experimental psychology where the participant is guided to make ”successive approximations” in arriving at the correct gesture. It was installed at the theater of Évry, that has a display on the shopping mall in Évry for the entire month of December.

Finally, we collaborated with Simon Fraser University on an interactive installation called still, moving. The installation created a sonic experience that heightens self-awareness of our micro-movements in stillness. Sound created an intimate envelope that nurtures self-reflection and the experience of inward sensations. In still, moving, the audience was equipped with two Myo Armbands that capture their movements as well as their muscular activity. The physiological signals such as muscle tension and subtle accelerations were analyzed and mapped to a sound environment in order to increase perception of the inner self. The design of the relationship between movement and sound was evolving along the interaction, shifting the soundscape from reflective to challenging, guiding the audience in an exploration of novel and gradual relationship to weight and understanding of the complexity of the silent body.

Everyday creativity:

Finally, for non expert users, we developed an inexpensive method for fabricating *Stretchis*, highly stretchable interfaces that combine sensing and displaying capabilities [22]. This method enables designers and casual makers to embed transparent conductors and electroluminescence displays in stretchable PDMS substrates (Figure 6). We showed how to prototype stretchable user interfaces for a range of application scenarios by using standard design software and screen-printing techniques. Despite the use of inexpensive equipment, our results demonstrate that we can produce durable and highly stretchable sensors and displays that remain functional under strain levels of more than 100%.

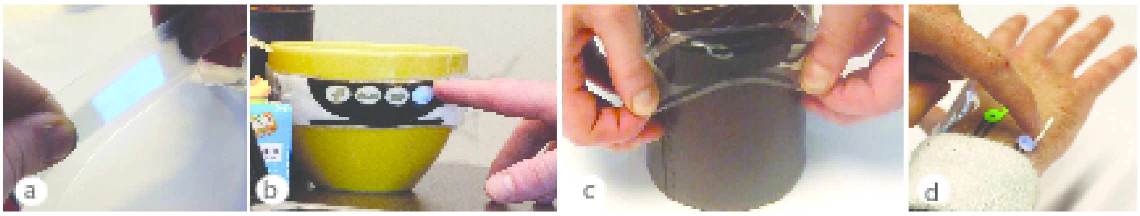


Figure 6. Stretchis are highly stretchable user interfaces that include touch and proximity sensors and electroluminescent displays (a). Stretchis are transparent (b); can be stretched to fit to the geometry of different physical objects (c); and can act as on-skin user interfaces (d).

7.4. Collaboration

Participants: Michel Beaudouin-Lafon, Cédric Fleury, Wendy Mackay, Can Liu, Ignacio Avellino Martinez.

ExSitu is interested in exploring new ways to support collaborative interaction, especially within and across large interactive spaces such as those of the Digiscope network (http://digiscope.fr/).

We studied how wall-sized displays support small groups of users working together on large amounts of data. We conducted observational studies showing that users adopt a range of collaboration styles, from loosely to closely coupled and that shared interaction techniques, in which multiple users perform a command collaboratively, support co-located collaborative work. In order to test the effect of such shared interaction techniques, we operationalize five collaborative situations with increasing levels of coupling in a data manipulation task [18]. The results show the benefits of shared interaction for close collaboration: it encourages collaborative manipulation, it is more efficient and preferred by users, and it reduces physical navigation and fatigue. We also identified the time costs caused by disruption and communication in loose collaboration and analyzed the trade-offs between parallelization and close collaboration. Altogether, these findings can inform the design of shared interaction techniques to support collaboration on wall-sized displays.

We are also interested in how to help teams of novice crafters prototype physical objects. To this end, we conducted a study [12] framed around two all-day design charrettes where novices performed a complete design process: ideation sketching, concept development and presentation, fabrication planning documentation and collaborative fabrication of hand-crafted prototypes. This structure allowed us to control key aspects of the design process while collecting rich data about creative tasks, including sketches on paper, physical models, and videos of collaboration discussions. Participants used a variety of drawing techniques to convey 3D concepts. They also extensively manipulated physical materials, such as paper, foam, and cardboard, both to support concept exploration and communication with design partners. Based on these observations, we proposed design guidelines for CAD tools targeted at novice crafters.

GRAPHDECO Project-Team

6. New Results

6.1. Computer-Assisted Design with Heterogeneous Representations

6.1.1. How Novices Sketch and Prototype Hand-Fabricated Objects

Participant: Adrien Bousseau.

We are interested in how to create digital tools to support informal sketching and prototyping of physical objects by novices. Achieving this goal first requires a deeper understanding of how non-professional designers generate, explore, and communicate design ideas with traditional tools, i.e., sketches on paper and hands-on prototyping materials. We conducted a study framed around two all-day design charrettes where participants perform a complete design process: ideation sketching, concept development and presentation, fabrication planning documentation and collaborative fabrication of hand-crafted prototypes (Figure 4). This structure allows us to control key aspects of the design process while collecting rich data about creative tasks, including sketches on paper, physical models, and videos of collaboration discussions. Participants used a variety of drawing techniques to convey 3D concepts. They also extensively manipulated physical materials, such as paper, foam, and cardboard, both to support concept exploration and communication with design partners. Based on these observations, we propose design guidelines for CAD tools targeted at novice crafters.

This work is a collaboration with Theophanis Tsandilas, Lora Oehlberg and Wendy Mackay from the ExSitu group, Inria Saclay. It has been published at ACM Conference on Human Factors in Computing Systems (CHI) 2016 [9].

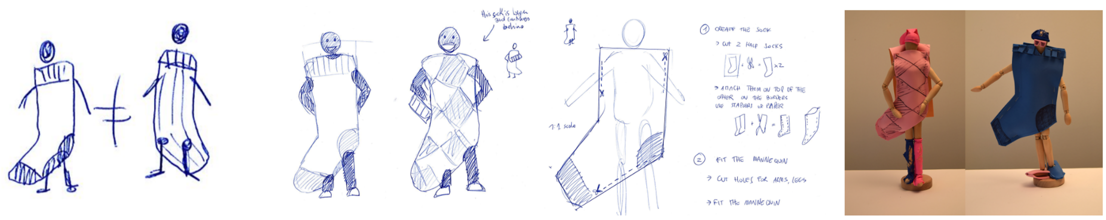


Figure 4. We asked participants to design a costume, from an initial sketch to a physical prototype.

6.1.2. Interactive Sketching of Urban Procedural Models

Participant: Adrien Bousseau.

3D modeling remains a notoriously difficult task for novices despite significant research effort to provide intuitive and automated systems. We tackle this problem by combining the strengths of two popular domains: sketch-based modeling and procedural modeling. On the one hand, sketch-based modeling exploits our ability to draw but requires detailed, unambiguous drawings to achieve complex models. On the other hand, procedural modeling automates the creation of precise and detailed geometry but requires the tedious definition and parameterization of procedural models. Our system uses a collection of simple procedural grammars, called snippets, as building blocks to turn sketches into realistic 3D models. We use a machine learning approach to solve the inverse problem of finding the procedural model that best explains a user sketch. We use non-photorealistic rendering to generate artificial data for training convolutional neural networks capable of quickly recognizing the procedural rule intended by a sketch and estimating its parameters. We integrate our algorithm in a coarse-to-fine urban modeling system that allows users to create rich buildings by successively sketching the building mass, roof, facades, windows, and ornaments (Figure 5). A user study shows that by using our approach non-expert users can generate complex buildings in just a few minutes.

This work is a collaboration with Gen Nishida, Ignacio Garcia-Dorado, Daniel G. Aliaga and Bedrich Benes from Purdue University. It has been published at ACM Transactions on Graphics (proc. SIGGRAPH) 2016 [8].

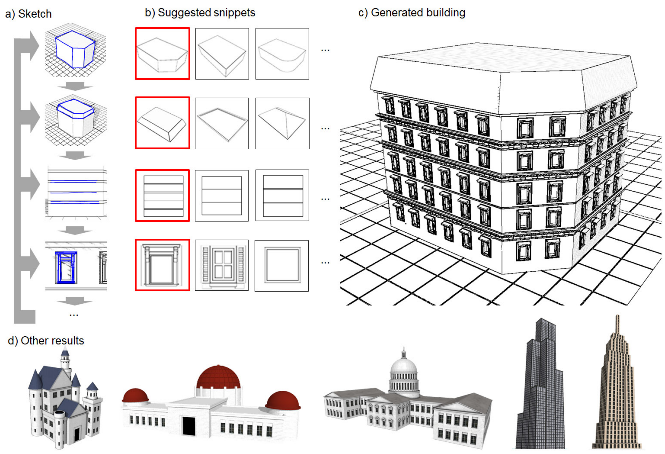


Figure 5. Our system allows novices to quickly create complex procedural 3D models of buildings by sketching.

6.1.3. Fidelity vs. Simplicity: a Global Approach to Line Drawing Vectorization

Participant: Adrien Bousseau.

Vector drawing is a popular representation in graphic design because of the precision, compactness and editability offered by parametric curves. However, prior work on line drawing vectorization focused solely on faithfully capturing input bitmaps, and largely overlooked the problem of producing a compact and editable curve network. As a result, existing algorithms tend to produce overly-complex drawings composed of many short curves and control points, especially in the presence of thick or sketchy lines that yield spurious curves at junctions. We propose the first vectorization algorithm that explicitly balances fidelity to the input bitmap with simplicity of the output, as measured by the number of curves and their degree. By casting this trade-off as a global optimization, our algorithm generates few yet accurate curves, and also disambiguates curve topology at junctions by favoring the simplest interpretations overall. We demonstrate the robustness of our algorithm on a variety of drawings, sketchy cartoons and rough design sketches (Figure 6).

The first author of this work, Jean-Dominique Favreau, is co-advised by Adrien Bousseau and Florent Lafarge (Titane team). The work was published at ACM Transactions on Graphics (proc. SIGGRAPH) 2016 [5].

6.1.4. SketchSoup: Exploratory Ideation using Design Sketches

Participant: Adrien Bousseau.

A hallmark of early stage design is a number of quick-and-dirty sketches capturing design inspirations, model variations, and alternate viewpoints of a visual concept. We present SketchSoup, a workflow that allows designers to explore the design space induced by such sketches. We take an unstructured collection of drawings as input, register them using a multi-image matching algorithm, and present them as a 2D interpolation space (Figure 7). By morphing sketches in this space, our approach produces plausible visualizations of shape and viewpoint variations despite the presence of sketch distortions that would prevent standard camera calibration and 3D reconstruction. In addition, our interpolated sketches can serve as inspiration for further drawings, which feed back into the design space as additional image inputs. SketchSoup thus fills a significant gap in the early ideation stage of conceptual design by allowing designers to make better informed choices before

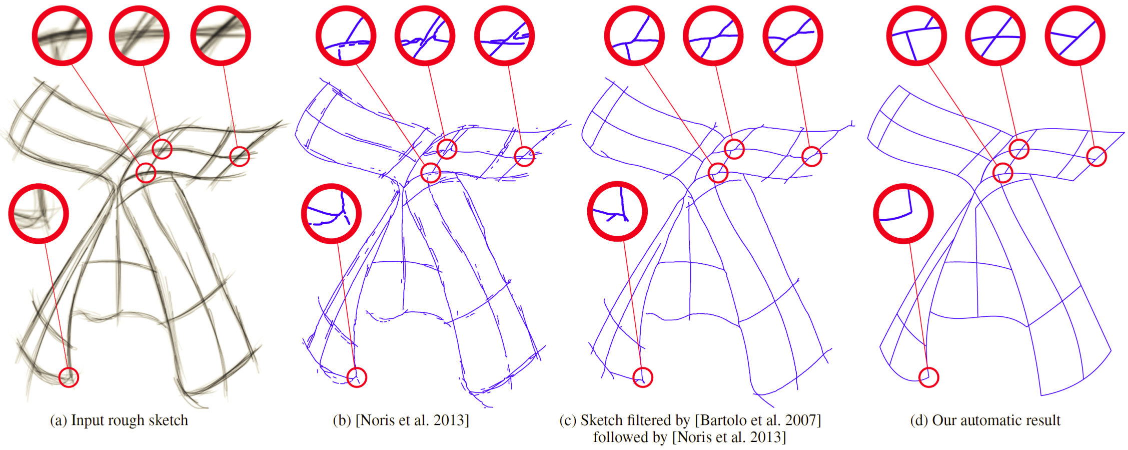


Figure 6. Rough sketches often contain overlapping strokes (a), which existing vectorization algorithms represent as multiple curves (b). Pre-filtering the drawing improves the vectorization, but produces spurious curve segments at junctions (c). Since existing algorithms analyze junctions locally, they cannot recover the proper topology of these seemingly similar line configurations. By adopting a global formulation that optimizes for both fidelity to the input sketch and simplicity of the output curve network, our algorithm recovers proper topology while significantly reducing the overall number of curves and control points. Design sketch after Sori Yanagi's "Butterfly" stool.

proceeding to more expensive 3D modeling and prototyping. From a technical standpoint, we describe an end-to-end system that judiciously combines and adapts various image processing techniques to the drawing domain – where the images are dominated not by color, shading and texture, but by sketchy stroke contours.

This work is a collaboration with Rahul Arora and Karan Singh from University of Toronto and Vinay P. Namboodiri from IIT Kampur. The project was initiated while Rahul Arora was an intern in our group. It will be published in Computer Graphics Forum in 2017.

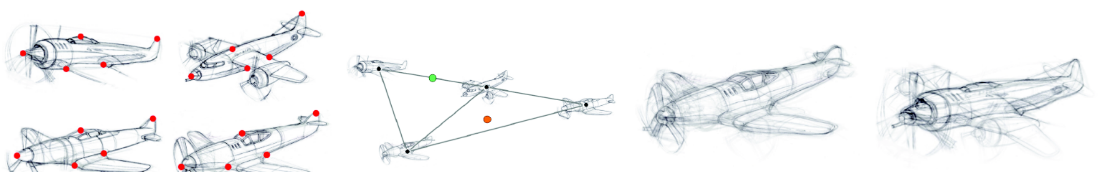


Figure 7. SketchSoup takes an unstructured set of sketches as input, along with a small number of correspondences (shown as red dots), registers the sketches and embeds them into a 2D interpolation space based on their shape differences. Users can explore the interpolation space to generate novel sketches.

6.1.5. Modeling Symmetric Developable Surfaces from a Single Photo

Participants: Amelie Fondevilla, Adrien Bousseau.

We propose to reconstruct 3D developable surfaces from a single 2D drawing traced and annotated over a side-view photo of a partially symmetrical object (Figure 8). Our reconstruction algorithm combines symmetry and orthogonality shapes cues within a unified optimization framework that solves for the 3D position of the Bézier control points of the drawn curves while being tolerant to drawing inaccuracy and perspective distortions. We then rely on existing surface optimization methods to produce a developable surface that interpolates our 3D curves. Our method is particularly well suited for the modeling and fabrication of fashion items as it converts the input drawing into flattened developable patterns ready for sewing.

This work is a collaboration with Damien Rohmer, Stefanie Hahmann and Marie-Paule Cani from the Imagine team (LJK/ Inria Grenoble Rhône Alpes). This work was presented at the AFIG French conference in November 2016, where it received the 3rd price for best student work.

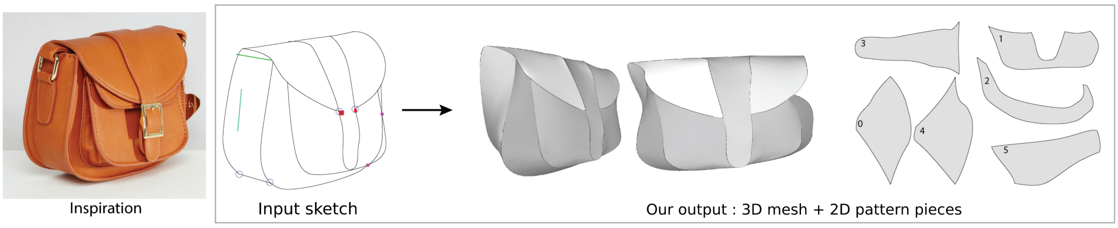


Figure 8. Our method reconstructs a 3D mesh and 2D pattern pieces of a sewed object from a single annotated drawing.

6.1.6. DeepSketch: Sketch-Based Modeling using 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. 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. It is supported by the CRISP Inria associate team.

6.2. Graphics with Uncertainty and Heterogeneous Content

6.2.1. Cotemporal Multi-View Video Segmentation

Participants: Abdelaziz Djelouah, George Drettakis.

We address the problem of multi-view video segmentation of dynamic scenes in general and outdoor environments with possibly moving cameras. Multi-view methods for dynamic scenes usually rely on geometric calibration to impose spatial shape constraints between viewpoints. In this work, we show that the calibration constraint can be relaxed while still getting competitive segmentation results using multi-view constraints. We introduce new multi-view cotemporality constraints through motion correlation cues, in addition to common appearance features used by cosegmentation methods to identify co-instances of objects. We also take advantage of learning based segmentation strategies by casting the problem as the selection of monocular proposals that satisfy multi-view constraints. This yields a fully automated method that can segment subjects of interest without any particular pre-processing stage (see Fig. 9). Results on several challenging outdoor datasets demonstrate the feasibility and robustness of our approach.

This work is a collaboration with Jean-Sébastien Franco and Edmond Boyer from Morpheo team at Inria Grenoble, and Patrick Pérez from Technicolor. The work has been published in the International Conference on 3D Vision (3DV) - 2016 [10].

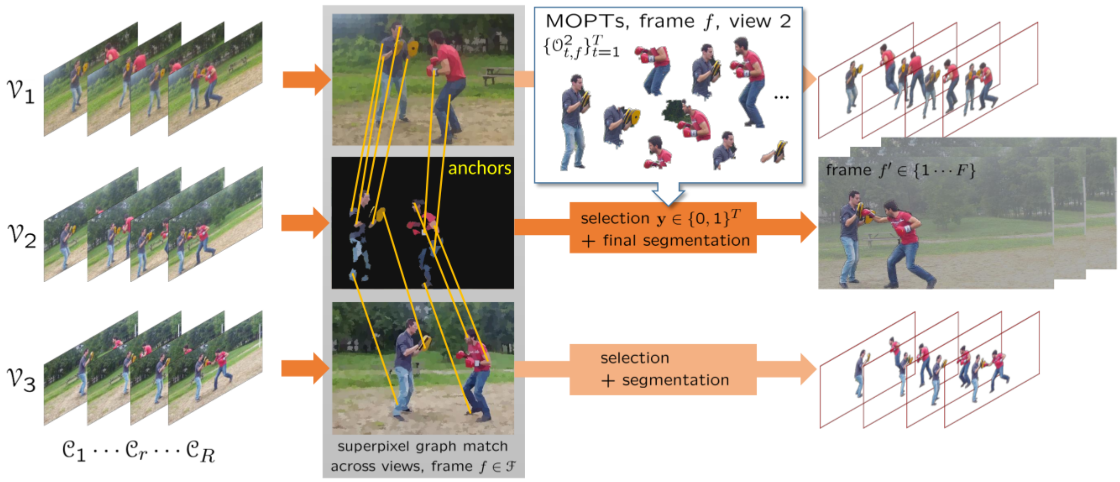


Figure 9. Synchronized videos of the same scene are partitioned into short clips. At a small number of instants where motion is sufficiently informative, cross-view correspondences between with similar appearance and motion are obtained by graph matching. In each view, matched superpixels, which are likely to lie on moving foreground objects, are used as sparse anchors to guide the selection process among a large pool of moving objects proposals extracted from all clips.

6.2.2. Automatic 3D Car Model Alignment for Mixed Image-Based Rendering

Participants: Rodrigo Ortiz Cayon, Abdelaziz Djelouah, George Drettakis.

We present a method that improves quality of Image-Based Rendering of poorly reconstructed objects. We focus on the case of reflective objects which are hard to reconstruct, such as cars. The key insight is to replace these poorly reconstructed objects with models from existing rich 3D CAD databases, and subsequently align them to the input images. We use deep learning-based algorithms to obtain the 3D model present in the databases which is closest to the object seen in the images. We formulate two optimizations using all available information to finely position and orient the model and adapt it to image contours (see Fig. 10 (1.)). Our method provides much higher quality rendering results of such objects compared to previous solutions as seen in Fig. 10 (2.(b)).

This work is a collaboration with Francisco Massa and Mathieu Aubry from École des Ponts ParisTech. The work was published in the International Conference in 3D Vision (3DV) - 2016 [12].

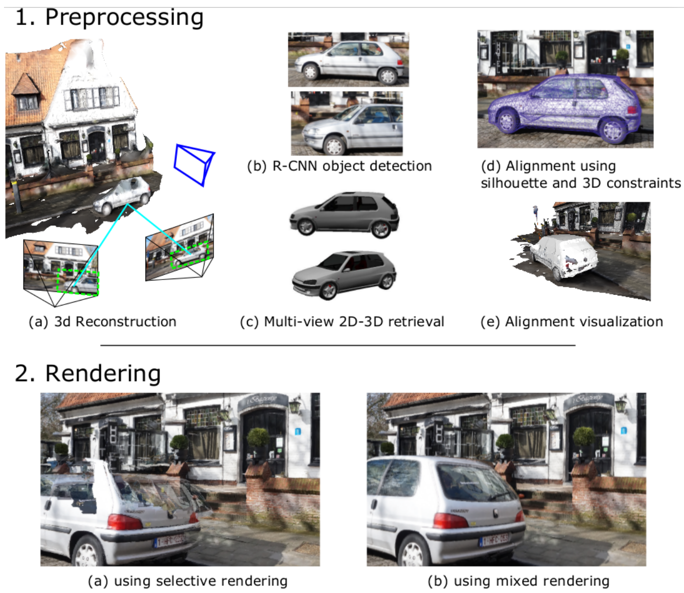


Figure 10. Overview of our pre-processing and rendering pipeline for Mixed Image-Based Rendering.

6.2.3. Multi-View Inpainting for Image-Based Scene Editing and Rendering

Participants: Theo Thonat, George Drettakis.

We propose a method to remove objects such as people and cars from multi-view urban image datasets (Figure 11), enabling free-viewpoint Image-Based Rendering (IBR) in the edited scenes. Our method combines information from multi-view 3D reconstruction with image inpainting techniques, by formulating the problem as an optimization of a global patch-based objective function. We use IBR techniques to reproject information from neighboring views, and 3D multi-view stereo reconstruction to perform multi-view coherent initialization for inpainting of pixels not filled by reprojection. Our algorithm performs multi-view consistent inpainting for color and 3D by blending reprojections with patch-based image inpainting. We run our algorithm on casually captured datasets, and Google Street View data, removing objects such as cars, people and pillars, showing that our approach produces results of sufficient quality for free-viewpoint IBR on "cleaned up" scenes, as well as IBR scene editing, such as limited displacement of real objects.

This work is a collaboration with Eli Shechtman and Sylvain Paris from Adobe Research. It has been published in the International Conference on 3D Vision (3DV) - 2016 [13].

6.2.4. Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation in Games

Participants: George Koulieris, George Drettakis.

Comfortable, high-quality 3D stereo viewing has become a requirement for interactive applications. Previous research shows that manipulating disparity can alleviate some of the discomfort caused by 3D stereo, but it is best to do this locally, around the object the user is gazing at. The main challenge is thus to develop a gaze predictor in the demanding context of real-time, heavily task-oriented applications such as games. Our key observation is that player actions are highly correlated with the present state of a game, encoded by game variables. Based on this, we trained a classifier to learn these correlations using an eye-tracker which

Figure 11. Our system takes as input a set of images from the same scene (top row). The method then removes all the vehicules in a multi-view coherent way (bottom row).

provides the ground-truth object being looked at. The classifier is used at runtime to predict object category – and thus gaze – during game play, based on the current state of game variables. We used this prediction to propose a dynamic disparity manipulation method, which provided rich and comfortable depth. We evaluated the quality of our gaze predictor numerically and experimentally, showing that it predicts gaze more accurately than previous approaches. A subjective rating study demonstrates that our localized disparity manipulation is preferred over previous methods.

This work is a collaboration with Katerina Mania from the Technical University of Crete and Douglas Cunningham from the Technical University of Cottbus. The work was presented at the IEEE conference for Virtual Reality (IEEE VR) 2016 [14].

Figure 12. We propose a gaze predictor (a), used to perform localized stereo grading (b). We compare to no stereo grading (c) and prior work (d). We provide rich and comfortable depth. Please use red/cyan anaglyph glasses.

6.2.5. A Feasibility Study with Image-Based Rendered Virtual Reality in Patients with Mild Cognitive Impairment

Participant: George Drettakis.

Virtual Reality (VR) has emerged as a promising tool in many domains of therapy and rehabilitation, and has recently attracted the attention of researchers and clinicians working with elderly people with MCI, Alzheimer's disease and related disorders. In this study we tested the feasibility of using highly realistic image-based rendered VR with patients with MCI and dementia. We designed an attentional task to train selective and sustained attention, and we tested a VR and a paper version of this task (see Fig. 13) in a single-session within-subjects design. Results showed that participants with MCI and dementia reported to be highly satisfied and interested in the task, and they reported high feelings of security, low discomfort, anxiety and fatigue. In addition, participants reported a preference for the VR condition compared to the paper condition, even if the task was more difficult. Interestingly, apathetic participants showed a preference for the VR condition stronger than that of non-apathetic participants. These findings suggest that VR-based training can be considered as an interesting tool to improve adherence to cognitive training for elderly people with cognitive impairment.

This work was a collaboration with EA CoBTek/IA, CMRR (memory center) of the CHU (University Hospital) of Nice, Disney Research and Trinity College Dublin, as part of the (completed) VERVE EU project. The work was published in the PLoS ONE journal [7].

Figure 13. The VR and paper conditions of the study

6.2.6. Scalable Inside-Out Image-Based Rendering

Participant: George Drettakis.

The goal of this project was to provide high-quality free-viewpoing rendering of indoors environments. captured with off-the-shelf equipment such as a high-quality color camera and a commodity depth sensor. Image-based Rendering (IBR) can provide the realistic imagery required at real-time speed. For indoor scenes however, two challenges are especially prominent. First, the reconstructed 3D geometry must be compact, but faithful enough to respect occlusion relationships when viewed up close. Second, man-made materials call for view-dependent texturing, but using too many input photographs reduces performance.

We customize a typical RGB-D 3D surface reconstruction pipeline to produce a coarse global 3D surface, and local, per-view geometry for each input image. Our tiled IBR preserves quality by economizing on the expected contributions that entire groups of input pixels make to a final image. The two components are designed to work together, giving real-time performance, while hardly sacrificing quality. Testing on a variety of challenging scenes shows that our inside-out IBR scales favorably with the number of input images.

This work was a collaboration with P. Hedman, G. Brostow and T. Ritschel at UCL, as part of the CR-PLAY project. It was published in ACM Transactions on Graphics (Proc. SIGGRAPH Asia) [6].

Figure 14. Images from our method rendered in 1080p at 55 Hz on an Nvidia Titan X GPU. Input is an RGB-D video and 298 high-quality photos of 'Dr Johnson's house', London. With no wheelchair access to this floor, curators were keen to have their rooms digitized.

6.2.7. Measuring Accommodation and Comfort in Head-Mounted Displays

Participants: George Koulieris, George Drettakis.

Head-mounted displays (HMDs) are rapidly becoming the preferred display for stereo viewing in virtual environments, but they often cause discomfort and even sickness. Previous studies have shown that a major cause of these adverse symptoms is the vergence-accommodation (VA) conflict. Specifically, the eyes use the distance to the screen to accommodate, while they use the distance to the fixated virtual object to converge. The VA conflict is the difference between those distances. The magnitude of the conflict is well correlated with subjective reports of discomfort. Many methods have been proposed for reducing the VA conflict and thereby reducing discomfort by making accommodation more consistent with vergence. But no one has actually measured accommodation in HMDs to see how well a given method is able to drive it to the desired distance. We built the first device for measuring accommodation in an HMD, using a modular design with off-the-shelf components, focus-adjustable lenses, and an autorefractor. We conducted experiments using the device to determine how well accommodation is driven with various combinations of HMD properties and viewing conditions: focus-adjustable lenses, depth-of-field rendering, binocular viewing, and "monovision" (setting the two eyes' focal distances to quite different values). We found that focus-adjustable lenses drive accommodation appropriately across many conditions. The other techniques were much less effective in driving accommodation. We also investigated whether the ability to drive accommodation predicts viewer comfort. We did this by conducting a discomfort study with most of the conditions in the accommodation study. We found that the ability to drive accommodation did in fact predict the amount of discomfort. Specifically, the most comfortable conditions were the ones that generated accommodation consistent with vergence. Together, these results illustrate the potential benefit of focus-adjustable lenses: They enable stimulation of accommodation and thereby comfortable viewing. In contrast, monovision neither enable accurate accommodation nor comfortable viewing.

This work is an ongoing collaboration with Martin S. Banks from UC Berkeley, in the context of the CRISP Inria associate team.

6.2.8. Beyond Gaussian Noise-Based Texture Synthesis

Participants: Kenneth Vanhoey, Georgios Kopanas, George Drettakis.

Texture synthesis methods based on noise functions have many nice properties: they are continuous (thus resolution-independent), infinite (can be evaluated at any point) and compact (only functional parameters need to be stored). A good method is also non-repetitive and aperiodic. Current techniques, like Gabor Noise, fail to produce structured content. They are limited to so-called "Gaussian textures", characterized by second-order statistics like mean and variance only. This is suitable for noise-like patterns (e.g., marble, wood veins, sand) but not for structured ones (e.g., brick wall, mountain rocks, woven yarn). Other techniques, like Local Random-Phase noise, leverage some structure but as a trade-off with repetitiveness and periodicity.

In this project, we model higher-order statistics produced by noise functions in a parametric model. Then we define an algorithm for sampling of the noise functions' parameters so as to produce a texture that meets prescribed statistics. This sampling ensures both the reproduction of higher-order visual features with high probability, like edges and ridges, and non-repetitiveness plus aperiodicity thanks to the stochastic sampling method. Moreover a (deep) learning algorithm has been established to infer the prescribed statistics from an input examplar image.

This project is a collaboration with Ian H. Jermyn (Durham University, UK, former Inria) and Mathieu Aubry (ENPC, France).

6.2.9. Fences in Image Based Rendering

Participants: Abdelaziz Djelouah, Frederic Durand, George Drettakis.

One of the key problem in Image Based Rendering (IBR) methods is the rendering of regions with incorrect 3D reconstruction. Some methods try to overcome the issue in the case of reflections and transparencies trough the estimation of two planes or the usage of 3D stock models. Fences with their thin repetitive structures are another important common source of 3D reconstruction errors but have received very little attention in the context of image based rendering. They are present in most urban pictures and represent a standard failure case for reconstruction algorithms, and state of the art rendering methods exhibit strong artifacts.

In this project, we propose to detect and segment fences in urban pictures for IBR applications. Similarly to related methods in image *de-fencing*, we use the assumptions that fences are thin repetitive structures lying on a 3D plane. To address this problem we consider the following steps: First we propose a multi-view approach to estimate the plane supporting the fences using repetition candidates. Second, we estimate image matting taking advantage of the multi-view constraints and the repetitive patterns. Finally, the estimated 3D plane and matting masks are used in a new rendering algorithm.

6.2.10. Handling reflections in Image-Based Rendering

Participants: Theo Thonat, Frederic Durand, George Drettakis.

In order to render new viewpoints, current Image Based Rendering techniques (IBR) use an approximate geometry to warp and blend photographs from close viewpoints. They assume the scene materials are diffuse, so geometry colors are independent of the viewpoint, an assumption that fails in the case of specular surfaces such as windows. Dealing with reflections in an IBR context first requires identifying what are the diffuse and the specular color layers in the input images. The challenge is then to correctly warp the specular layers since their associated geometry is not available and since the normals of the reflective surfaces might be not reliable.

HYBRID Project-Team

7. New Results

7.1. Virtual Reality and 3D Interaction

7.1.1. Perception in Virtual Environments

With the increasing demand in consumer VR applications, the need to understand how users perceive the virtual environment and their virtual self (avatar) is becoming more and more important. In particular, with the potential of virtual reality to alter and control avatars in different ways, the user representation in the virtual world does not always necessarily match the user body structure. Besides, the study of how the users perceive their surrounding environment (e.g. depth perception) is another active field of research in VR.

The role of interaction in virtual embodiment: Effects of the virtual hand representation

Participants: Ferran Argelaguet and Anatole Lécuyer

First, we have studied how people appropriate their virtual hand representation when interacting in virtual environments [14]. In order to answer this question, we conducted an experiment studying the sense of embodiment when interacting with three different virtual hand representations (see Figure 2), each one providing a different degree of visual realism but keeping the same control mechanism. The main experimental task was a Pick-and-Place task in which participants had to grasp a virtual cube and place it to an indicated position while avoiding an obstacle (brick, barbed wire or fire). Results show that the sense of agency is stronger for less realistic virtual hands which also provide less mismatch between the participant's actions and the animation of the virtual hand. In contrast, the sense of ownership is increased for the human virtual hand which provides a direct mapping between the degrees of freedom of the real and virtual hand.

This work was done in collaboration with MimeTIC team.

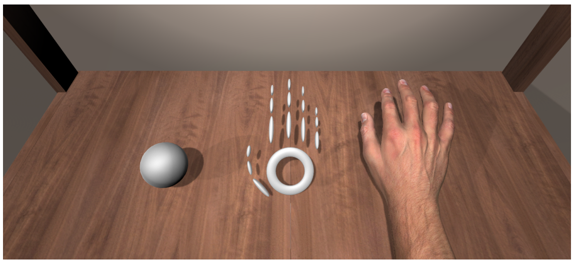


Figure 2. Evaluated virtual hand representations: abstract (left), iconic (center) and realistic virtual hands (right). Each virtual hand had its own visual feedback when the grasping operation was triggered.

Wow! I Have Six Fingers!": Would You Accept Structural Changes of Your Hand in VR?

Participants: Ferran Argelaguet and Anatole Lécuyer

In a different context, we have explored how users would accept as their own a six-digit realistic virtual hand [6]. By measuring participants' senses of ownership (i.e., the impression that the virtual hand is actually our own hand) and agency (i.e., the impression to be able to control the actions of the virtual hand), we somehow evaluate the possibility of creating a Six-Finger Illusion in VR. We measured these two dimensions of virtual embodiment in a virtual reality experiment where participants performed two tasks successively: (1) a self-manipulation task inducing visuomotor feedback, where participants mimicked finger movements presented in the virtual scene and (2) a visuotactile task inspired by Rubber Hand Illusion protocols, where an experimenter stroked the hand of the user with a brush (see Figure 3). The real and virtual brushes were synchronously stroking the participants' real and virtual hand, and in the case when the virtual brush was stroking the additional virtual digit, the real ring finger was also synchronously stroked to provide consistent tactile stimulation and elicit a sense of embodiment. Results of the experiment show that participants did experience high levels of ownership and agency of the six-digit virtual hand as a whole. These results bring preliminary insights about how avatar with structural differences can affect the senses of ownership and agency experienced by users in VR.

This work was done in collaboration with MimeTIC team.

Figure 3. The virtual six-finger hand and the participant's hand are synchronously stimulated using a virtual and a real brush respectively.

CAVE Size Matters: Effects of Screen Distance and Parallax on Distance Estimation in Large Immersive Display Setups

Participants: Ferran Argelaguet and Anatole Lécuyer

When walking within a CAVE-like system, accommodation distance, parallax, and angular resolution vary according to the distance between the user and the projection walls, which can alter spatial perception. As these systems get bigger, there is a need to assess the main factors influencing spatial perception in order to better design immersive projection systems and virtual reality applications. In this work, we performed two experiments that analyze distance perception when considering the distance toward the projection screens and parallax as main factors. Both experiments were conducted in a large immersive projection system with up to 10-meter interaction space. The first experiment showed that both the screen distance and parallax have a strong asymmetric effect on distance judgments. We observed increased underestimation for positive parallax conditions and slight distance overestimation for negative and zero parallax conditions. The second experiment further analyzed the factors contributing to these effects and confirmed the observed effects of the first experiment with a high-resolution projection setup providing twice the angular resolution and improved accommodative stimuli. In conclusion, our results suggest that space is the most important characteristic for distance perception, optimally requiring about 6- to 7-meter distance around the user, and virtual objects with high demands on accurate spatial perception should be displayed at zero or negative parallax [3].

This work was done in collaboration with MimeTIC team and the University of Hamburg.

7.1.2. 3D User Interfaces

GiAnt: stereoscopic-compliant multi-scale navigation in VEs

Participants: Ferran Argelaguet

Navigation in multi-scale virtual environments (MSVE) requires the adjustment of the navigation parameters to ensure optimal navigation experiences at each level of scale (see Figure 4). In particular, in immersive stereoscopic systems, e.g. when performing zoom-in and zoom-out operations, the navigation speed and the stereoscopic rendering parameters have to be adjusted accordingly. Although this adjustment can be done manually by the user, it can be complex, tedious and strongly depends on the virtual environment. We have proposed GiAnt (GIant/ANT) [15], a new multi-scale navigation technique which automatically and seamlessly adjusts the navigation speed and the scale factor of the virtual environment based on the user's perceived navigation speed. The adjustment ensures an almost-constant perceived navigation speed while avoiding diplopia effects or diminished depth perception due to improper stereoscopic rendering configurations. The results from the conducted user evaluation shows that GiAnt is an efficient multi-scale navigation which minimizes the changes of the scale factor of the virtual environment compared to state-of-the-art multi-scale navigation techniques.

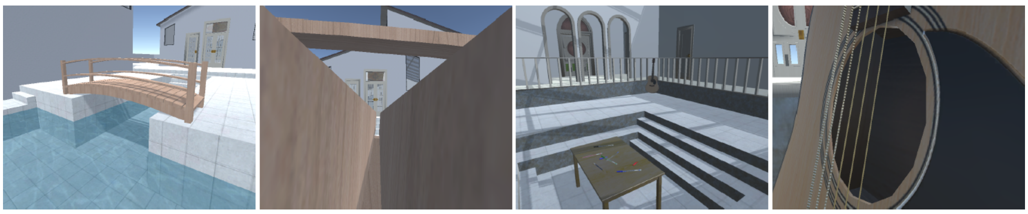


Figure 4. Multi-scale navigation sequence requiring the adaptation of the camera speed and the stereoscopic rendering parameters (e.g. parallax). GiAnt ensures that the navigation speed and the scale factor of the virtual environment are adjusted ensuring a comfortable navigation experience.

Enjoying 360° Vision with the FlyVIZ

Participants: Florian Nouviale, Maud Marchal and Anatole Lécuyer

FlyVIZ is a novel concept of wearable display device which enables to extend the human field-of-view up to 360°. With the FlyVIZ users can enjoy an artificial omnidirectional vision and see "with eyes behind their back"! We propose a novel version of our approach called the FlyVIZ v2. It is based on affordable and on the shelf components. For image acquisition, the FlyVIZ v2 relies on an iPhone4S smart-phone combined with a GoPano lens that contains a curved mirror enabling the capture of video with 360° horizontal field-of-view. For image transformation, we developed a dedicated software for iPhone that processes the video stream and transforms it into a real-time meaningful representation for the user. The "FlyVIZ_v2" was demonstrated at the ACM SIGGRAPH Emerging Technologies (2016).

Figure 5. (Left) Overview of the system. (Middle) 360° panoramic image displayed in the HMD when walking in a corridor. (Right) User grabbing an object located outside his natural field-of-view.

D3PART: A new Model for Redistribution and Plasticity of 3D User Interfaces

Participants: Jérémy Lacoche and Bruno Arnaldi

D3PART (Dynamic 3D Plastic And Redistribuable Technology) is a new model that we introduced to handle redistribution for 3D user interfaces. Redistribution consists in changing the components distribution of an interactive system across different dimensions such as platform, display and user. We extended previous plasticity models with redistribution capabilities, which lets developers create applications where 3D content and interaction tasks can be automatically redistributed across the different dimensions at runtime [21].

This work was done in collaboration with b<>com, ENIB and Telecom Bretagne.

Integration concept and model of Industry Foundation Classes (IFC) for interactive virtual environments

Participants: Anne-Solène Dris, Valérie Gouranton and Bruno Arnaldi

We defined a concept of Building Information Modeling (BIM) in combination with an integration model in order to enable interaction in Virtual Environments (see Figure 6). Such model, rich of information could be used to increase the level of abstraction of the interaction process. We proposed to explore and define how to create a BIM to ensure interoperability with the Industry Foundation Classes (IFC) model. The IFC model provides a definition of building objects, geometry, relation between objects, and other attributes such as layers, systems, link to planning, construction method, materials, domain (HVAC, Electrical, Architectural, Structure...) and quantities. The interoperability will enrich the virtual environment with the aim of creating an informed and interactive virtual environments, thus reducing the costs of applications' development. We defined a BIM modeling methodology extending the IFC interoperability to the interactive virtual environment [19].

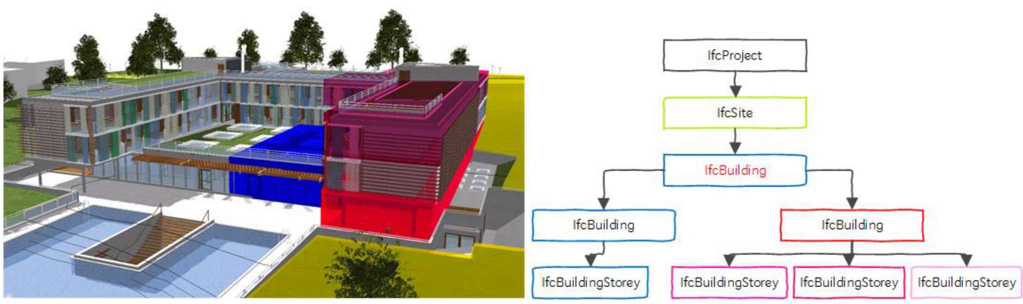


Figure 6. Interaction in virtual environments related to construction area based on BIM and a model of Industry Foundation Classes (IFC).

7.1.3. Virtual Archaeology

Digital and handcrafting processes applied to sound-studies of archaeological bone flutes

Participants: Jean-Baptiste Barreau, Ronan Gaugne, Bruno Arnaldi and Valérie Gouranton.

Bone flutes make use of a naturally hollow raw-material. As nature does not produce duplicates, each bone has its own inner cavity, and thus its own sound-potential. This morphological variation implies acoustical specificities, thus making it impossible to handcraft a true and exact sound-replica in another bone. This phenomenon has been observed in a handcrafting context and has led us to conduct two series of experiments (the first one using a handcrafting process, the second one using a 3D process) in order to investigate its exact influence on acoustics as well as on sound-interpretation based on replicas. The comparison of the results has shed light upon epistemological and methodological issues that have yet to be fully understood. This work contributes to assessing the application of digitization, 3D printing and handcrafting to flute-like sound instruments studied in the field of archaeomusicology [26].

This work was done in collaboration with MimeTIC team, ARTeHis, LBBE and Atelier El Block.

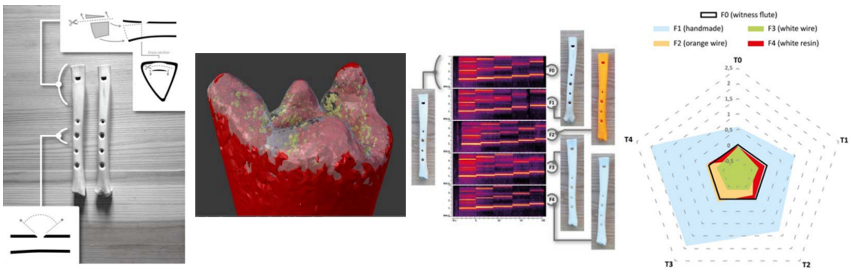


Figure 7. Sound-studies of archaeological bone flutes: (a) control flute (left) and its replica (right) both made out of goat's tibias, (b) 3D sculpted patch (transparent gray) on Blender (based on the geometry of the cloud), (c) Diagrams analysis , (d) The sound proximity of each replica comparing to the control flute, for each finger hole (numeric scale in semi-tones).

Internal 3D Printing of Intricate Structures

Participants: Ronan Gaugne, Valérie Gouranton and Bruno Arnaldi.

Additive technologies are increasingly used in Cultural Heritage process, for example in order to reproduce, complete, study or exhibit artefacts. 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. Medical images based digitization such as MRI or CT scan are also increasingly used in CH as they provide information on the internal structure of archaeological material. Different previous works illustrated the interest of combining 3D printing and CT scan in order to extract concealed artefacts from larger archaeological material. The method was based on 3D segmentation techniques within volume data obtained by CT scan 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. We propose 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 [25], [11]. This method was successfully applied to visualize the interior of a funeral urn and is currently applied on a set of tools agglomerated in a gangue of corrosion (see Figure 8).

This work was done in collaboration with Inrap and Image ET.

Figure 8. Front and bottom views of our 3D printed urn.

7.2. Physically-Based Simulation and Multisensory Feedback

7.2.1. Physically-based Simulation

Real-time tracking of deformable targets in 3D ultrasound sequences

Participants: Maud Marchal

Soft-tissue motion tracking is an active research area that consists in providing accurate evaluation about the location of anatomical structures. To do so, ultrasound imaging is often used since it is non-invasive, real-time and portable. Thus, several ultrasound tracking approaches have been developed in order to estimate soft tissue displacements that are caused by physiological motions and manipulations by medical tools. These methods have gained significant interest for image-guided therapies such as radio-frequency ablation or high-intensity focused ultrasound. In our work, we present a real-time approach that allows tracking deformable structures in 3D ultrasound sequences [8]. Our method consists in obtaining the target displacements by combining robust dense motion estimation and mechanical model simulation. We performed an evaluation of our method through simulated data, phantom data, and real-data. Results demonstrate that this novel approach has the advantage of providing correct motion estimation regarding different ultrasound shortcomings including speckle noise, large shadows and ultrasound gain variation. Furthermore, we show the good performance of our method with respect to state-of-the-art techniques by testing on the 3D databases provided by MICCAI CLUST'14 and CLUST'15 challenges.

This work was done in collaboration with LAGADIC team and b<>com.

7.2.2. 3D Haptic Interaction

DesktopGlove: a Multi-finger Force Feedback Interface Separating Degrees of Freedom Between Hands

Participants: Merwan Achibet and Maud Marchal

In virtual environments, interacting directly with our hands and fingers greatly contributes to immersion, especially when force feedback is provided for simulating the touch of virtual objects. Yet, common haptic interfaces are unfit for multi-finger manipulation and only costly and cumbersome grounded exoskeletons do provide all the efforts expected from object manipulation. To make multi-finger haptic interaction more accessible, we have proposed to combine two affordable haptic interfaces into a bimanual setup named DesktopGlove. With this approach, each hand is in charge of different components of object manipulation: one commands the global motion of a virtual hand while the other controls its fingers for grasping (see Figure 9). In addition, each hand is subjected to forces that relate to its own degrees of freedom so that users perceive a variety of haptic effects through both of them. Our results show that (1) users are able to integrate the separated degrees of freedom of DesktopGlove to efficiently control a virtual hand in a posing task, (2) DesktopGlove shows overall better performance than a traditional data glove and is preferred by users, and (3) users considered the separated haptic feedback realistic and accurate for manipulating objects in virtual environments [12].

This work was done in collaboration with MJOLNIR team.

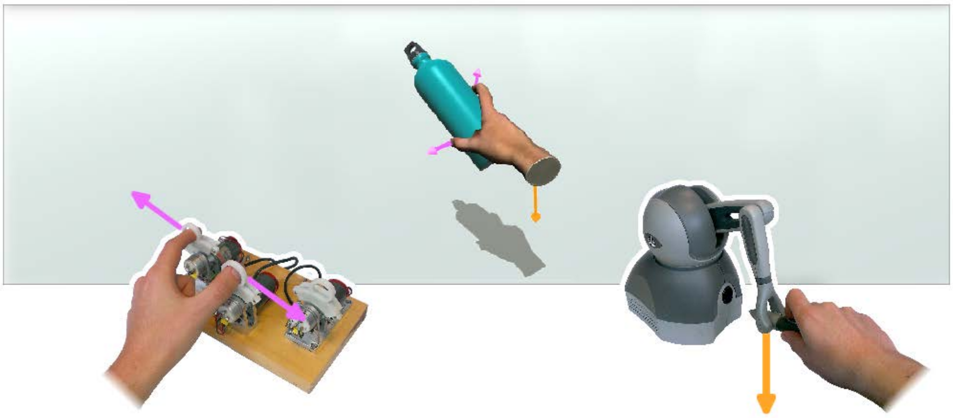


Figure 9. DesktopGlove separates the control of one virtual hand between both user's hands: a common haptic arm handles the global motion and a custom multi-finger interface controls the virtual fingers. The force feedback is split between both interfaces so that each hand is exposed to forces that relate to its own frame of reference.

ElasticArm: leveraging passive haptic feedback in virtual environments

Participants: Merwan Achibet, Adrien Girard, Anatole Lécuyer and Maud Marchal

Haptic feedback is known to improve 3D interaction in virtual environments but current haptic interfaces remain complex and tailored to desktop interaction. In [2], we describe an alternative approach called "Elastic-Arm" for incorporating haptic feedback in immersive virtual environments in a simple and cost-effective way. The Elastic-Arm is based on a body-mounted elastic armature that links the user's hand to his body and generates a progressive egocentric force when extending the arm. A variety of designs can be proposed with multiple links attached to various locations on the body in order to simulate different haptic properties and sensations such as different levels of stiffness, weight lifting, bimanual interaction, etc. Our passive haptic approach can be combined with various 3D interaction techniques and we illustrate the possibilities offered by the Elastic-Arm through several use cases based on well-known techniques such as the Bubble technique, redirected touching, and pseudo-haptics. A user study was conducted which showed the effectiveness of our pseudo-haptic technique as well as the general appreciation of the Elastic-Arm. We believe that the Elastic-Arm could be used in various VR applications which call for mobile haptic feedback or human-scale haptic sensations.

Vision-based adaptive assistance and haptic guidance for safe wheelchair corridor following

Participants: Maud Marchal

In case of motor impairments, steering a wheelchair can become a hazardous task. Joystick jerks induced by uncontrolled motions may lead to wall collisions when a user steers a wheelchair along a corridor. In [7] we introduce a low-cost assistive and guidance system for indoor corridor navigation in a wheelchair, which uses purely visual information, and which is capable of providing automatic trajectory correction and haptic guidance in order to avoid wall collisions. A visual servoing approach to autonomous corridor following serves as the backbone to this system. The algorithm employs natural image features which can be robustly extracted in real time. This algorithm is then fused with manual joystick input from the user so that progressive assistance and trajectory correction can be activated as soon as the user is in danger of collision. A force feedback in conjunction with the assistance is provided on the joystick in order to guide the user out of his dangerous trajectory. This ensures intuitive guidance and minimal interference from the trajectory correction system. In addition to being a low-cost approach, it can be seen that the proposed solution does not require an a-priori environment model. Experiments on a robotised wheelchair equipped with a monocular camera prove the capability of the system to adaptively guide and assist a user navigating in a corridor.

This work was done in collaboration with LAGADIC team.

7.2.3. Tactile Interaction at Fingertips

The fingertips are one of the most important and sensitive parts of our body. They are the first stimulated areas of the hand when we interact with our environment. Providing haptic feedback to the fingertips in virtual reality could, thus, drastically improve perception and interaction with virtual environments. Within this context, we proposed two contributions for tactile feedback and haptic interaction at the fingertips.

The Haptip

Participants: Adrien Girard, Yoren Gaffary, Anatole Lécuyer and Maud Marchal

In [5], we present a modular approach called HapTip to display such haptic sensations at the level of the fingertips. This approach relies on a wearable and compact haptic device able to simulate 2 Degree of Freedom (DoF) shear forces on the fingertip with a displacement range of 2 mm. Several modules can be added and used jointly in order to address multi-finger and/or bimanual scenarios in virtual environments. For that purpose, we introduce several haptic rendering techniques to cover different cases of 3D interaction, such as touching a rough virtual surface, or feeling the inertia or weight of a virtual object. In order to illustrate the possibilities offered by HapTip, we provide four use cases focused on touching or grasping virtual objects (see Figure 10). To validate the efficiency of our approach, we also conducted experiments to assess the tactile perception obtained with HapTip. Our results show that participants can successfully discriminate the directions of the 2 DoF stimulation of our haptic device. We found also that participants could well perceive different weights of virtual objects simulated using two HapTip devices. We believe that HapTip could be used in numerous applications in virtual reality for which 3D manipulation and tactile sensations are often crucial, such as in virtual prototyping or virtual training.

Figure 10. Illustrative use cases of our approach HapTip: the user can get in contact and tap a virtual bottle, touch a surface and feel its texture, and heft an object and feel its weight.

This work was done in collaboration with CEA List.

Studying one and two-finger perception of tactile directional cues

Participants: Yoren Gaffary, Anatole Lécuyer and Maud Marchal

In [20], we study the perception of tactile directional cues by one or two fingers, using either the index, middle, or ring finger, or any of their combination. Therefore, we use tactile devices able to stretch the skin of the fingertips in 2 DOF along four directions: horizontal, vertical, and the two diagonals. We measure the recognition rate in each direction, as well as the subjective preference, depending on the (couple of) finger(s) stimulated (see Figure 11). Our results show first that using the index and/or middle finger performs significantly better than using the ring finger on both qualitative and quantitative measures. The results when comparing one versus two-finger configurations are more contrasted. The recognition rate of the diagonals is higher when using one finger than two, whereas two fingers enable a better perception of the horizontal direction. These results pave the way to other studies on one versus two-finger perception, and raise methodological considerations for the design of multi-finger tactile devices.

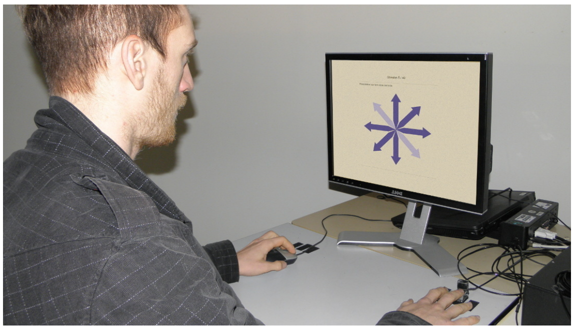


Figure 11. Experimental setup of our study on the perception of tactile directional cues by one or two fingers: the participant reports the direction of the stimulus he just perceived on the fingertip of his middle finger.

This work was done in collaboration with CEA List, IRMAR and Agrocampus Ouest.

7.3. Collaborative Virtual Environments

7.3.1. Acting in Collaborative Virtual Environments

VR Rehearsals for Acting with Visual Effects

Participants: Rozenn Bouville, Valérie Gouranton and Bruno Arnaldi,

We studied the use of Virtual Reality for movie actors rehearsals of VFX-enhanced scenes. The impediment behind VFX scenes is that actors must be filmed in front of monochromatic green or blue screens with hardly any cue to the digital scenery that is supposed to surround them. The problem is worsens when the scene includes interaction with digital partners. The actors must pretend they are sharing the set with imaginary creatures when they are, in fact, on their own on an empty set. To support actors in this complicated task, we introduced the use of VR for acting rehearsals not only to immerse actors in the digital scenery but to provide them with advanced features for rehearsing their play. Indeed, our approach combines a fully interactive environment with a dynamic scenario feature to allow actors to become familiar with the virtual elements while rehearsing dialogue and action at their own speed. The interactive and creative rehearsals enabled by the system can be either single-user or multiuser. Moreover, thanks to the wide range of supported platforms, VR rehearsals can take place either onset or offset. We conducted a preliminary study to assess whether VR training can replace classical training (see Figure 12). The results show that VR-trained actors deliver a performance just as good as ordinarily trained actors. Moreover, all the subjects in our experiment preferred VR training to classic training [17].

Synthesis and Simulation of Collaborative Surgical Process Models

Figure 12. The use of VR for acting rehearsal enables actors to rehearse being immersed in the virtual scenery before being shot on a green and empty set.

Participants: Guillaume Claude, Valérie Gouranton and Bruno Arnaldi

The use of Virtual Reality for surgical training has been mostly focused on technical surgical skills. We proposed a novel approach by focusing on the procedural aspects [4]. Our system relies on a specific workflow, which enbables to generate a model of the procedure based on real case surgery observations made in the operating room (see Figure 13). In addition, in the context of the project S3PM we then proposed an innovative workflow to integrate the generic model of the procedure (generated from the real-case surgery observation) as a scenario model in the VR training system (see Figure 14). We described how the generic procedure model could be generated, as well as its integration in the virtual environment [18].

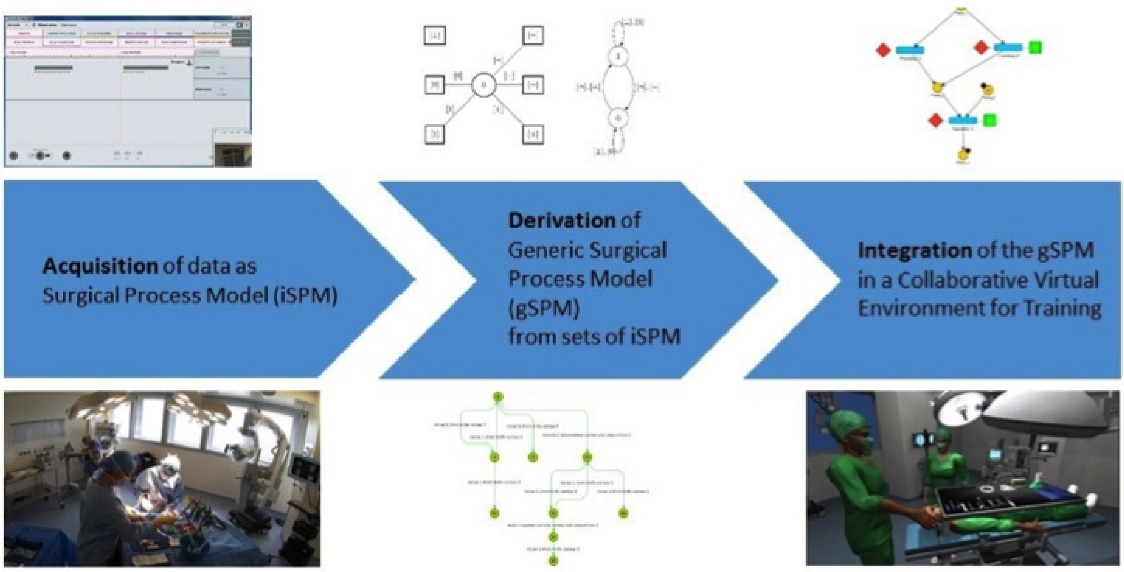


Figure 13. Collaborative Virtual Environments for Training in Surgical Procedures, based on observations during real surgeries. Observation data is integrated into a system providing a Generalised Surgical Process Model (gSPM) of the procedure. This Model is integrated as the scenario of the Virtual Environment.

This work was done in collaboration with HYCOMES team and LTSI Inserm Medicis.

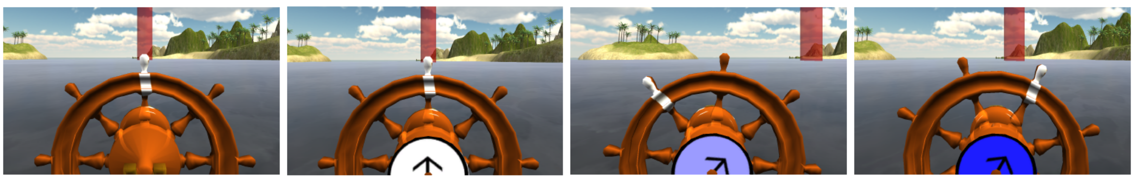
Figure 14. Virtual replica of a real operating room of Rennes hospital (CHU Rennes) in the Immersia CAVE-like setup (IRISA/Inria Rennes).

7.3.2. Awareness for Collaboration in Virtual Environments

Take-Over Control Paradigms in Collaborative Virtual Environments for Training

Participants: Gwendal Le Moulec, Ferran Argelaguet, Anatole Lécuyer and Valérie Gouranton

We studied the notion of Take-Over Control in Collaborative Virtual Environments for Training (CVET). The Take-Over Control represents the transfer (the take over) of the interaction control of an object between two or more users. This paradigm is particularly useful for training scenarios, in which the interaction control could be continuously exchanged between the trainee and the trainer, e.g. the latter guiding and correcting the trainee's actions. We proposed a formalization of the Take-Over Control followed by an illustration focusing in a use-case of collaborative maritime navigation. In the presented use-case, the trainee has to avoid an under-water obstacle with the help of a trainer who has additional information about the obstacle. The use-case allows to highlight the different elements a Take-Over Control situation should enforce, such as user's awareness. Different Take-Over Control techniques were provided and evaluated focusing on the transfer exchange mechanism and the visual feedback (see Figure 15). The results show that participants preferred the Take-Over Control technique which maximized the user awareness [24].



(a) *Initial state* *No feedback condition*

(b) *Initial state* $w_1 = 1$, $w_2 = 0$

(c) *In progress* $w_1 = 0.5$, $w_2 = 0.5$

(d) *Completed* $w_1 = 0$, $w_2 = 1$

Figure 15. Our illustrative use case inspired by maritime navigation for Take-Over Control during training in a collaborative virtual environment. The user was instructed to steer a boat towards a semi-transparent red column (target destination) by controlling the heading of the boat. A white handle indicated the rotation angle of the boat(a). The sequence (b,c,d) shows the evolution of the contribution of the trainer on the steering angle, from no control to full control.

Vishnu: Virtual Immersive Support for HelpiNg Users: An Interaction Paradigm for Collaborative Remote Guiding in Mixed Reality

Participants: Morgan Le Chénéchal, Valérie Gouranton and Bruno Arnaldi

Increasing networking performances as well as the emergence of Mixed Reality (MR) technologies make possible providing advanced interfaces to improve remote collaboration. We presented a novel interaction paradigm called Vishnu that aims to ease collaborative remote guiding. We focus on collaborative remote maintenance as an illustrative use case. It relies on an expert immersed in Virtual Reality (VR) in the remote workspace of a local agent helped through an Augmented Reality (AR) interface. The main idea of the Vishnu paradigm is to provide the local agent with two additional virtual arms controlled by the remote expert who can use them as interactive guidance tools. Many challenges come with this: collocation, inverse kinematics (IK), the perception of the remote collaborator and gestures coordination. Vishnu aims to enhance the maintenance procedure thanks to a remote expert who can show to the local agent the exact gestures and actions to perform (see Figure 16). Our pilot user study shows that it may decrease the cognitive load compared to a usual approach based on the mapping of 2D and de-localized informations, and it could be used by agents in order to perform specific procedures without needing to have an available local expert [22].

This work was done in collaboration with b<>com and Telecom Bretagne.

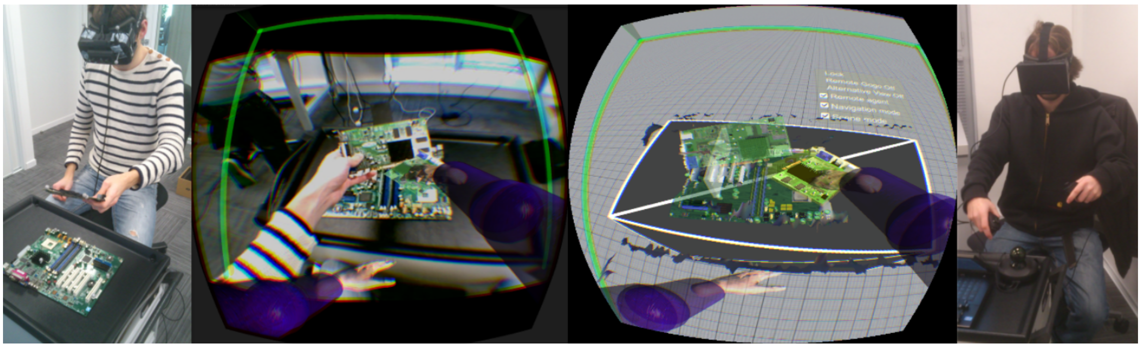


Figure 16. Illustration of the Vishnu approach: system and viewpoints of the agent (left) and the expert (right) in a motherboard assembly scenario.

When the Giant meets the Ant: An Asymmetric Approach for Collaborative and Concurrent Object Manipulation in a Multi-Scale Environment

Participants: Morgan Le Chénéchal, Jérémy Lacoche, Valérie Gouranton and Bruno Arnaldi

We proposed a novel approach to enable two or more users to manipulate an object collaboratively. Our goal is to benefits from the wide variety of todays VR devices. Our solution is based on an asymmetric collaboration pattern at different scales in which users benefit from suited points of views and interaction techniques according to their device setups. Each user application is adapted thanks to plasticity mechanisms. Our system provides an efficient way to co-manipulate an object within irregular and narrow courses, taking advantages of asymmetric roles in synchronous collaboration (see Figure 17). Moreover, it aims to provide a way to maximize the filling of the courses while the object moves on its path [23],[35].

This work was done in collaboration with b<>com and Telecom Bretagne.

7.4. Brain-Computer Interfaces

7.4.1. Contribution to a Reference Book on BCI

We have largely contributed to a reference book on BCI released in 2016 in French and English, and co-edited by Fabien Lotte, Maureen Clerc and Laurent Bougrain for ISTE (French version [36] [37]) and Wiley

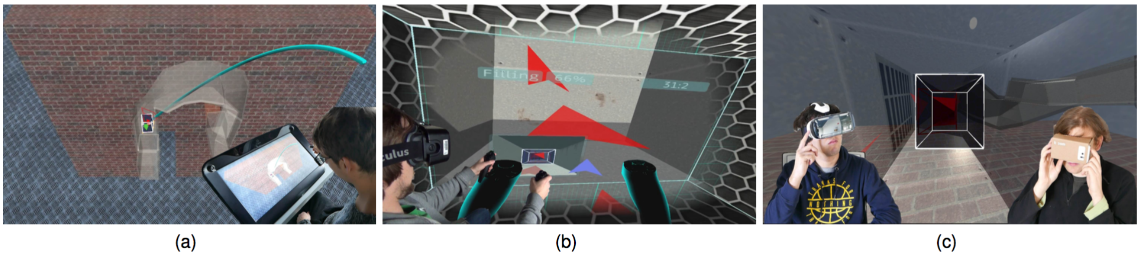


Figure 17. When the Giant meets the Ant: Collaborative manipulation of a virtual object (here, a cube) based on an asymmetric setting between two users who can be helped by two additional users. (a) The first participant has a global view of the scene and moves the object with a 3D bent ray. (b) The second user is placed inside the object and precisely rotates and scales it. (c) Two additional roles can be added. The first one helps to scale the object using a third person view of it. The other one is a spectator who switches between the other participants' viewpoints and helps them with oral communication.

(English version [39] [40]) publishers. This book provides keys for understanding and designing these multi-disciplinary interfaces, which require many fields of expertise such as neuroscience, statistics, informatics and psychology. This work corresponds to four different book chapters, all published in both French and English, which are presented hereafter.

Book chapter on BCI and videogames

Participants: Anatole Lécuyer

Videos games are often cited as a very promising field of applications for brain-computer interfaces. In a first chapter [30] [31], we described state of the art in the field of video games played "with the mind". In particular, we considered the results of the OpenViBE2 project: one of the most important research projects in this area. We presented a selection of prototypes developed during this OpenViBE2 project which is illustrative of the state of the art in this field and of the use of BCIs in video games, such as based on imagining a motion of the left and right hands to score goals, or in another example using the P300 cerebral potential to destroy spaceships in a remake of well-known Japanese game.

Book chapter on BCI softwares

Participants: Jussi Lindgren and Anatole Lécuyer

In a second chapter [28] [29], we described OpenViBE and other software platforms used to study the subject. The chapter gave an overview of such platforms. We described how the software components of the platforms reflect typical signal acquisition and signal processing stages used in BCI. Finally, we presented a high-level account of differences between major BCI platforms and gave a few pieces of advice and recommendation regarding BCI platform selection.

Book chapter on BCI and HCI

Participants: Andéol Evain, Ferran Argelaguet and Anatole Lécuyer

In a third chapter [34], we focused on the link between BCI and Human-Computer Interaction (HCI), and studied how HCI concepts can apply to BCIs. First, we presented an overview of the main concepts of HCI. We then studied the main characteristics of BCIs related to these concepts. This chapter also discussed the choice of cerebral patterns to use, depending on the interaction task and the use context. Finally, we presented the most promising new interaction paradigms for interaction with BCIs.

This work was done in collaboration with MJOLNIR team.

Book chapter on Neurofeedback

Participants: Lorraine Perronnet and Anatole Lécuyer

We proposed a fourth chapter called Brain training with Neurofeedback [33] [32]. We first defined the concept of neurofeedback (NF) and gave an overall view of the current status in this domain. Then we described the design of a NF training program and the typical course of a NF session, as well as the learning mechanisms underlying NF. We retraced the history of NF, explaining the origin of its questionable reputation and providing a foothold for understanding the diversity of existing approaches. We also discussed how the fields of NF and BCIs might potentially overlap in future with the development of "restorative" BCIs. Finally, we presented a few applications of NF and summarized the state of research of some of its major clinical applications.

This work was done in collaboration with VISAGES team.

7.4.2. BCI Methods and Techniques

Do the Stimuli of a BCI Have to be the Same as the Ones Used for Training it?

Participants: Andéol Evain, Ferran Argelaguet and Anatole Lécuyer

Does the stimulation used during the training on an SSVEP-based BCI have to be similar to that of the end use? We conducted an experiment in which we recorded six-channel EEG data from 12 subjects in various conditions of distance between targets , and of difference in color between targets [10]. Our analysis revealed that the stimulation configuration used for training which leads to the best classification accuracy is not always the one which is closest to the end use configuration. We found that the distance between targets during training is of little influence if the end use targets are close to each other, but that training at far distance can lead to a better accuracy for far distance end use. Additionally, an interaction effect is observed between training and testing color: while training with monochrome targets leads to good performance only when the test context involves monochrome targets as well, a classifier trained on colored targets can be efficient for both colored and monochrome targets. In a nutshell, in the context of SSVEP-based BCI, training using distant targets of different colors seems to lead to the best and more robust performance in all end use contexts.

This work was done in collaboration with MJOLNIR team.

A Novel Fusion Approach Combining Brain and Gaze Inputs for Target Selection

Participants: Andéol Evain, Ferran Argelaguet and Anatole Lécuyer

Gaze-based interfaces and Brain-Computer Interfaces (BCIs) allow for hands-free human-computer interaction. We investigated the combination of gaze and BCIs. We proposed a novel selection technique for 2D target acquisition based on input fusion [9]. This new approach combines the probabilistic models for each input, in order to better estimate the intent of the user. We evaluated its performance against the existing gaze and brain-computer interaction techniques. Twelve participants took part in our study, in which they had to search and select 2D targets with each of the evaluated techniques (see Figure 18). Our fusion-based hybrid interaction technique was found to be more reliable than the previous gaze and BCI hybrid interaction techniques for 10 participants over 12, while being 29% faster on average. However, similarly to what has been observed in hybrid gaze-and-speech interaction, gaze-only interaction technique still provides the best performance. Our results should encourage the use of input fusion, as opposed to sequential interaction, in order to design better hybrid interfaces.

This work was done in collaboration with MJOLNIR team.

7.4.3. BCI User Experience and Neurofeedback

Influence of Error Rate on Frustration of BCI Users

Participants: Andéol Evain, Ferran Argelaguet and Anatole Lécuyer

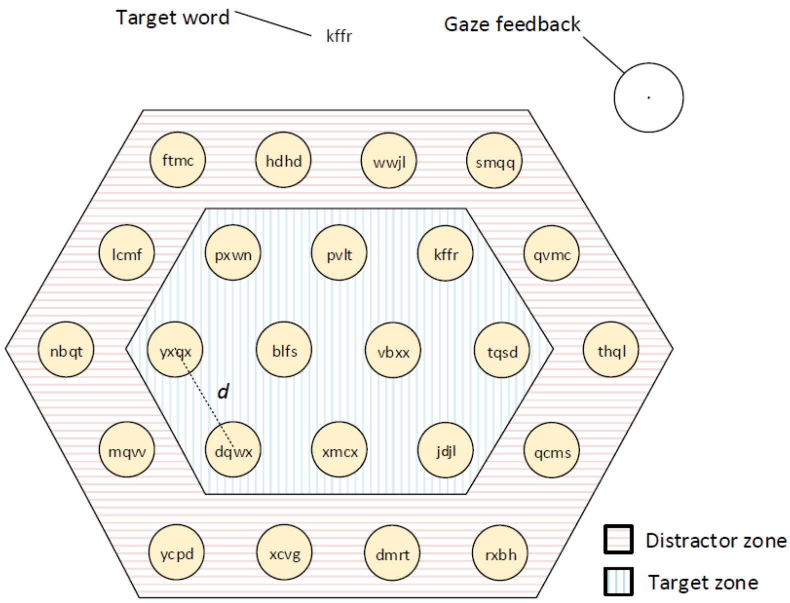


Figure 18. Experimental task combining gaze and brain inputs. The user has to look for the goal word displayed at the top of the screen, then, the user has to select the target with the exact same word. The detected gaze position is displayed under the form of a circle and a central point (visual feedback). For all trials the size of the targets remained constant, and only the length of the target word and the separation (d) between targets varied. The targets at the outer circle were distractors in which the target word was never placed.

Brain-Computer Interfaces (BCIs) are still much less reliable than other input devices. The error rates of BCIs range from 5% up to 60%. We assessed the subjective frustration, motivation, and fatigue of BCI users, when confronted to different levels of error rate [27]. We conducted a BCI experiment in which the error rate was artificially controlled (see Figure 19). Our results first show that a prolonged use of BCI significantly increases the perceived fatigue, and induces a drop in motivation. We also found that user frustration increases with the error rate of the system but this increase does not seem critical for small differences of error rate. Thus, for future BCIs, we advise to favor user comfort over accuracy when the potential gain of accuracy remains small.

This work was done in collaboration with MJOLNIR team.

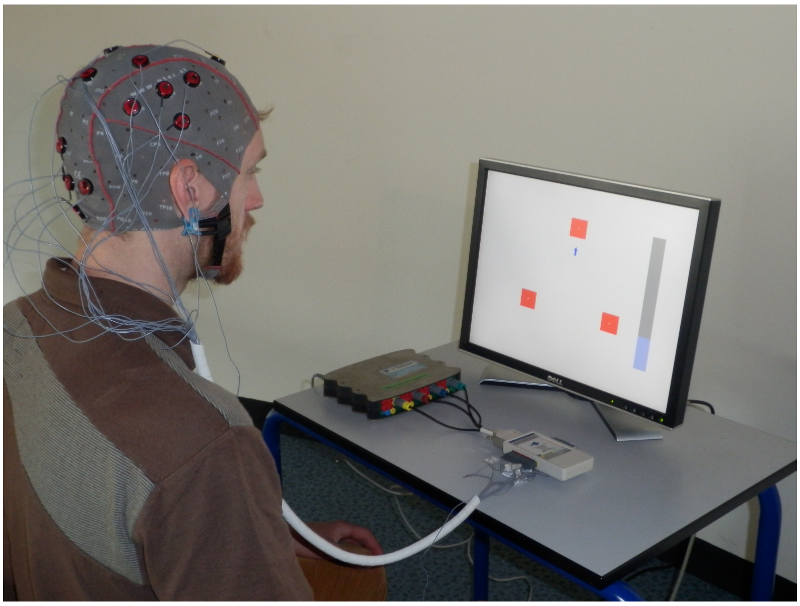


Figure 19. Experimental setup: a participant is using an SSVEP-based BCI which error rate is artificially controlled.

Design of an Experimental Platform for Hybrid EEG-fMRI Neurofeedback Studies

Participants: Marsel Mano, Lorraine Perronnet and Anatole Lécuyer

During a neurofeedback (NF) experiment one or more brain activity measuring technologies are used to estimate the changes of the acquired neural signals that reflect the changes of the subject's brain activity in real-time. There exist a variety of NF research applications that use only one type of neural signals (i.e. uni-modal) like EEG or fMRI, but there are very few NF researches that use two or more neural signals (i.e. multi-modal). We have developed a hybrid EEG-fMRI platform for bi-modal NF experiments, as part of the project Hemisfer. Our system is based on the integration and the synchronization of an MR-compatible EEG and fMRI acquisition subsystems. The EEG signals are acquired with a 64 channel MR-compatible solution from Brain Products and the MR imaging is performed on a 3T Verio Siemens scanner (VB17) with a 12-ch head coil. We have developed two real-time pipelines for EEG and fMRI that handle all the necessary signal processing, the Joint NF module that calculates and fuses the NF and a visualize module that displays the NF to the subject. The control and the synchronization of both subsystems with each-other and with the experimental protocol is handled by the NF Control. Our platform showed very good real-time performance with various pre-processing, filtering, and NF estimation and visualization methods. The entire fMRI process from acquisition to NF takes always less than 200ms, well below the TR of regular EPI sequences (2s). The same process for EEG, with NF update cycles varying 2-5Hz, is done in virtually real time (50Hz).

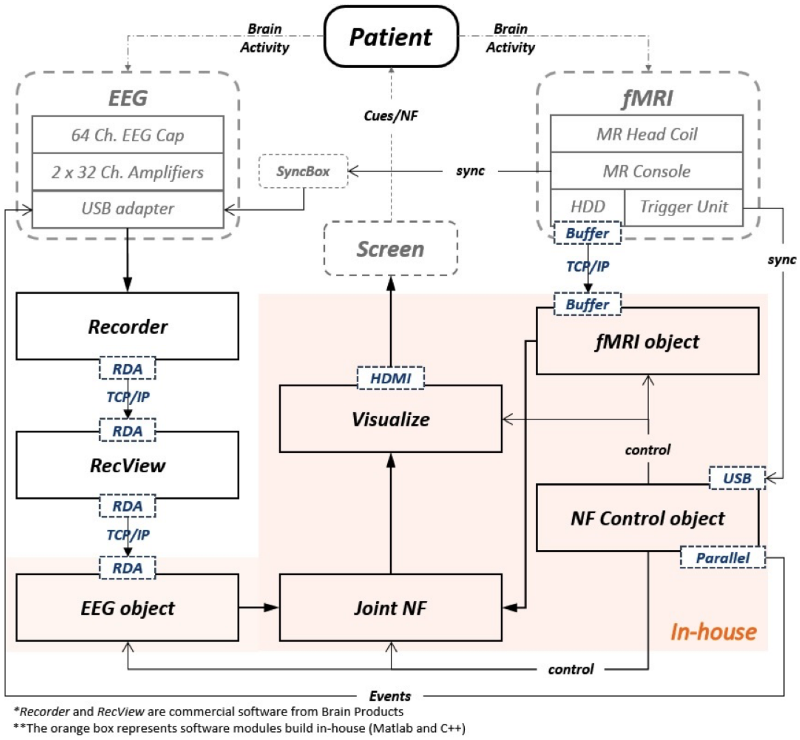


Figure 20. Architecture of our hybrid EEG-fMRI neurofeedback platform.

This work was done in collaboration with VISAGES team and presented as poster at OHBM 2016.

Unimodal versus Bimodal EEG-fMRI Neurofeedback

Participants: Lorraine Perronnet, Anatole Lécuyer and Marsel Mano

In the context of the HEMISFER project, we proposed a simultaneous EEG-fMRI experimental protocol in which 10 healthy participants performed a motor-imagery task in unimodal and bimodal neurofeedback conditions. With this protocol we were able to compare for the first time the effects of unimodal EEG-neurofeedback and fMRI-neurofeedback versus bimodal EEG-fMRI-neurofeedback by looking both at EEG and fMRI activations. We also introduced a new feedback metaphor for bimodal EEG-fMRI-neurofeedback that integrates both EEG and fMRI signal in a single bi-dimensional feedback (a ball moving in 2D). Such a feedback is intended to relieve the cognitive load of the subject by presenting the bimodal neurofeedback task as a single regulation task instead of two. Additionally, this integrated feedback metaphor gives flexibility on defining a bimodal neurofeedback target. Participants were able to regulate activity in their motor regions in all neurofeedback conditions. Moreover, motor activations as revealed by offline fMRI analysis were stronger during EEG-fMRI-neurofeedback than during EEG-neurofeedback. This result suggests that EEG-fMRI-neurofeedback could be more specific or more engaging than EEG-neurofeedback. Our results also suggest that during EEG-fMRI-neurofeedback, participants tended to regulate more the modality that was harder to control. Taken together our results shed light on the specific mechanisms of bimodal EEG-fMRI-neurofeedback and on its added-value as compared to unimodal EEG-neurofeedback and fMRI-neurofeedback.

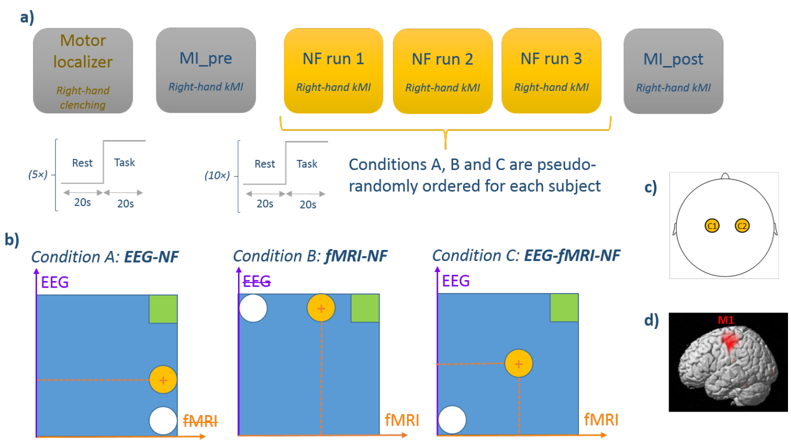


Figure 21. Experimental procedure for comparing unimodal versus bimodal EEG-fMRI neurofeedback.

This work was done in collaboration with VISAGES team and presented as poster at OHBM 2016. Experiments were conducted at NEURINFO platform from University of Rennes 1.

ILDA Project-Team

7. New Results

7.1. Wall Displays

Ultra-high-resolution wall displays feature a very high pixel density over a large physical surface, which makes them well-suited to the collaborative, exploratory visualization of large datasets (see Sections 6.3.1 and 6.3.2). We have continued working on the design, implementation and evaluation of interactive visualization techniques for such ultra-high-resolution wall-sized displays, focusing, in some of these projects, on the collaboration between users who perform different data manipulation and analysis tasks.

*Figure 4. **Left:** FITS-OW running on the WILDER platform, showing: multiple FITS images, (a) M31 on the left side, (b) three juxtaposed images that show observations of the Eagle nebula at different wavelengths, and (c) a much larger FITS image (86,499 × 13,474 pixels) used as a zoomable background over the entire wall; (d) the result-set of a SIMBAD query restricted to observations about galaxies; (e) basic measurements for galaxy M31; (e) a page of a research paper (PDF) discussing that particular galaxy; (f) the color map selector. **Right:** Results of a SIMBAD query superimposed on the corresponding FITS image, along with a sorted list of all items in the result-set. Selecting an element in this list updates the detailed info in the lower right window and highlights the source in the image. All windows can be freely repositioned on the wall.*

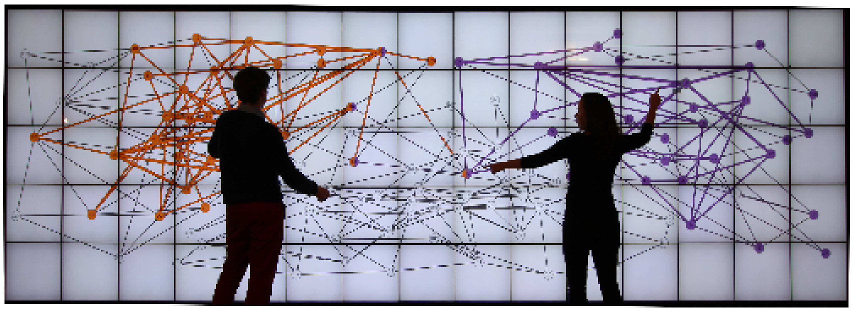
- We continued working on FITS-OW, an application designed for such wall displays, that enables astronomers to navigate in large collections of FITS images, query astronomical databases, and display detailed, complementary data and documents about multiple sources simultaneously. We published a paper about FITS-OW [7], in which we describe the system, reporting on the technical challenges we addressed in terms of distributed graphics rendering and data sharing over the computer clusters that drive wall displays. The article also describes how astronomers interact with their data using both the wall's touch-sensitive surface and handheld devices. This work was also featured as a short article in the SPIE Newsroom (see Section 10.3).
- Wall-sized displays support small groups of users working together on large amounts of data. Observational studies of such settings have shown that users adopt a range of collaboration styles, from loosely to closely coupled. Shared interaction techniques, in which multiple users perform a command collaboratively, have also been introduced to support co-located collaborative work. In [19], we operationalized five collaborative situations with increasing levels of coupling, and tested the effects of providing shared interaction support for a data manipulation task in each situation. The results show the benefits of shared interaction for close collaboration: it encourages

collaborative manipulation, it is more efficient and preferred by users, and it reduces physical navigation and fatigue. We also identify the time costs caused by disruption and communication in loose collaboration and analyze the trade-offs between parallelization and close collaboration. These findings inform the design of shared interaction techniques to support collaboration on wall-sized displays.

- We also studied how pairs explore graphs on a touch enabled wall-display [16], using two selection techniques adapted for collaboration: a basic localized selection, and a propagation selection technique that uses the idea of diffusion/transmission from an origin node. We assessed in a controlled experiment the impact of selection technique on a shortest path identification task. Pairs consistently divided space even though the task is not spatially divisible. The basic selection technique, that has a localized visual effect, led to parallel work that negatively impacted accuracy. The large visual footprint of the propagation technique led to close coordination, improving speed and accuracy for complex graphs only. We then observed the use of propagation on additional graph topology tasks, confirming pair strategies on spatial division and coordination.
- In [22], we focused on road traffic control center. Road traffic control centers are of vital importance to modern cities. Interviews with controllers in two such centers identified the need to incorporate the visualization of results from predictive traffic models with real traffic, to help operators choose among different interventions on the network. We explore this idea in a prototype that runs on a wall display, and supports direct touch and input from workstations and mobile devices. Apart from basic functionality to manage the current traffic such as changing traffic light duration or speed limits, the prototype incorporates traffic simulations for forecasting results of possible actions, highlighting their differences to current traffic. Based on needs identified in our interviews, we offered two techniques that visually combine simulated and real situations, taking advantage of the large display space: multiple independent views and DragMagic, a variation of magic lenses. A preliminary laboratory experiment suggests that both techniques are viable design options, even for monitoring several simulations and areas of interest, contrary to expectations from previous work. However DragMagics are easier to master.
- Immersion is the subjective impression of being deeply involved in a specific situation, and can be sensory or cognitive. In a position paper [23], we used a basic model of visual perception to study how ultra-high resolution wall displays can provide visual immersion. With their large size, depending on the position of viewers in front of them, wall displays can provide a surrounding and vivid environment. Users close to the wall can have their visual field filled by the wall and they are able to see clearly a large amount information with a fine resolution. However, when close to the wall, visual distortion due to large possible viewing angles, can affect the viewing of data. On the contrary, from far away, distortion is no longer an issue, but the viewers' visual field is not fully contained inside the wall, and the information details seen are less fine.

7.2. Gestures, Tangibles and Sound

- We designed a new way of implementing tangible interfaces with TouchTokens [4]. 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. We performed a formative user study to collect and analyze touch patterns with tokens of varying shape and size. The analysis of this pattern collection showed that individual users have a consistent grasp for each token, but that this grasp is user-dependent and that different grasp strategies can lead to confounding patterns. We thus designed a second set of tokens featuring notches that constrain users' grasp. Our recognition algorithm can classify the resulting patterns with a high level of accuracy (>95%) without any training, enabling application designers to associate rich touch input vocabularies with command triggers and parameter controls.



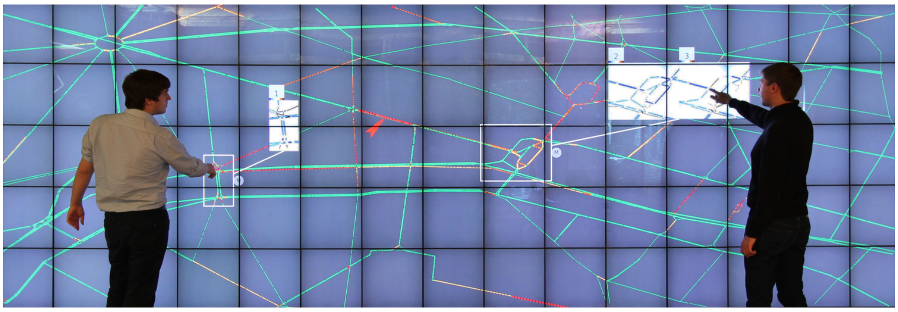
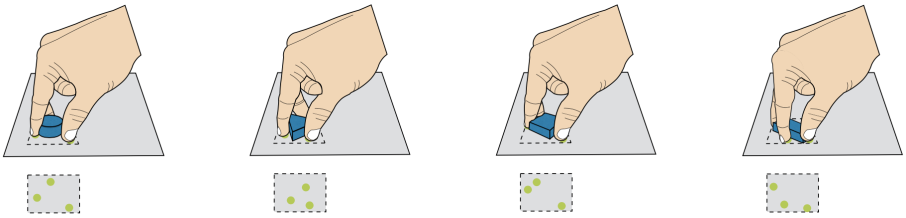


*Figure 5. **Top:** A pair using the propagation technique described in [16] to explore a graph. They discuss two communities, in orange and purple, selected using the propagation technique. The communities are linked by a specific node shown by the right user. The remaining 3 orange-purple nodes show how by propagating the purple community, it flows into the orange one through this node. **Bottom:** Visualization from [22] of traffic in a city with two "DragMagics" (white rectangles) showing one (left) and two (right) simulations associated with different possible interventions on the traffic. The simulation visualizations use difference color maps to highlight differences with the real traffic.*



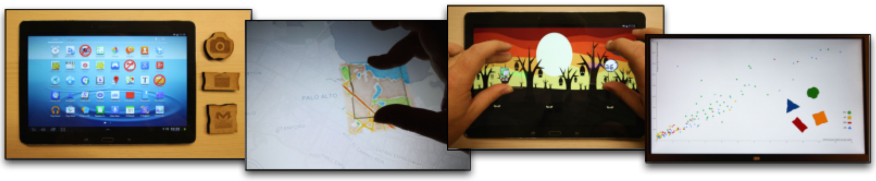


*Figure 6. **Top:** TouchTokens are passive tokens that guide users' fingers to specific spatial configurations, resulting in distinguishable touch patterns. **Bottom:** Proof-of-concept applications: access control, tangible magic lenses, character controllers in a game, data visualization.*

- In collaboration with IRCAM, we introduced SoundGuides [17], a user adaptable tool for auditory feedback on movement. The system is based on a interactive machine learning approach, where both gestures and sounds are first conjointly designed and conjointly learned by the system. The system can then automatically adapt the auditory feedback to any new user, taking into account the particular way each user performs a given gesture. SoundGuides is suitable for the design of continuous auditory feedback aimed at guiding users' movements and helping them to perform a specific movement consistently over time. Applications span from movement-based interaction techniques to auditory-guided rehabilitation. We first describe our system and report a study that demonstrates a "stabilizing effect" of our adaptive auditory feedback method.

7.3. Interacting with Linked Data

As part of the team's novel research theme on Semantics-Driven Data Manipulation 3.2 , and in collaboration with Aba-Sah Dadzie from the Open University, Emmanuel Pietriga coordinated a special issue of the Semantic Web Journal and wrote a follow-up [12] to the 2011 survey about Approaches to Visualizing Linked Data [42]. Linked Data promises to serve as a disruptor of traditional approaches to data management and use, promoting the push from the traditional Web of documents to a Web of data. The ability for data consumers to adopt a follow your nose approach, traversing links defined within a dataset or across independently-curated datasets, is an essential feature of this new Web of Data, enabling richer knowledge retrieval thanks to synthesis across multiple sources of, and views on, interrelated datasets. But for the Web of Data to be successful, we must design novel ways of interacting with the corresponding very large amounts of complex, interlinked, multi-dimensional data throughout its management cycle. The design of user interfaces for Linked Data, and more specifically interfaces that represent the data visually, play a central role in this respect. Contributions to this special issue on Linked Data visualization investigate different approaches to harnessing visualization as a tool for exploratory discovery and basic-to-advanced analysis. The papers in this volume illustrate the design and construction of intuitive means for end-users to obtain new insight and gather more knowledge, as they follow links defined across datasets over the Web of Data.

7.4. Visualization

- The attraction effect is a well-studied cognitive bias in decision making research, where one's choice between two alternatives is influenced by the presence of an irrelevant (dominated) third alternative. In collaboration with EPI Aviz, we examined in [13] whether this cognitive bias, so far only tested with three alternatives and simple presentation formats such as numerical tables, text and pictures, also appears in visualizations. In a series of crowdsource experiments, we observed this cognitive bias in visualizations (namely scatterplots), even in larger sets of alternatives, never considered before, where the number of alternatives is too large for numerical tables to be practical. We discussed implications for future research on how to further study and possibly alleviate the attraction effect.
- With colleagues from the University of Konstanz [14] we concluded previous work on data glyphs, i.e., visual marks that encode multiple dimensions to one or more visual variables. We provided a systematic review of experimental studies on data glyphs from the past 60 years, describing the types of glyphs and design variations tested, the tasks under which they were analyzed, and study results. Based on our meta analysis of all results, we further contributed a set of design implications and a discussion on open research directions.
- In [11], with colleagues from INRA, we provided an overview of a framework for Evolutionary Visual Exploration (EVE) that guides users in exploring large search spaces. EVE uses an interactive evolutionary algorithm to steer the exploration of multidimensional datasets towards two-dimensional projections that are interesting to the analyst. Our method smoothly combines automatically calculated metrics and user input in order to propose pertinent views to the user. While previously we showed that using EVE, domain experts were able to formulate interesting hypothesis and reach new insights when exploring freely, our new findings indicate that users, guided by the interactive evolutionary algorithm, are able to converge quickly to an interesting view of their data when a clear task is specified. Our work aims at building a bridge between the domains of visual analytics and interactive evolution.

IMAGINE Project-Team

6. New Results

6.1. User-centered Models for Shapes and Shape Assemblies

- **Scientist in charge**: Stefanie Hahmann.
- **Other permanent researchers**: Marie-Paule Cani, Jean-Claude Léon, Damien Rohmer.

Our goal, is to develop responsive shape models, i.e. 3D models that respond in the expected way under any user action, by maintaining specific application-dependent constraints (such as a volumetric objects keeping their volume when bent, or cloth-like surfaces remaining developable during deformation, etc). We are extending this approach to composite objects made of distributions and/or combination of sub-shapes of various dimensions.

6.1.1. Shape analysis

Figure 3. Left: Illustration of comparative study of 3D medial axis quality in [21]. Right: Hierarchies for similar shapes (dancers) in different poses to show that the proposed hierarchy is stable under articulation [22]. Coarser levels of the hierarchy are consistent even if finer levels are added in the presence of finer details. Also, note that the hierarchy is retained even with occlusion: The pink level of the left arm of the first dancer is occluded, but the blue level begins as it should.

Within the post-doc of Thomas Delame a comparative study between 3D skeletonization methods has been achieved. This work has been summarized as a Eurographics State of the Art [15]. Moreover, a comparative study of the quality between 3D medial axis was assessed and published in Vision, Modeling and Visualization [21].

We developed a multilevel analysis method of 2D shapes and used it to find similarities between the different parts of a shape [22]. Such an analysis is important for many applications such as shape comparison, editing, and compression. Our robust and stable method decomposes a shape into parts, determines a parts hierarchy, and measures similarity between parts based on a salience measure on the medial axis, the Weighted Extended Distance Function, providing a multi-resolution partition of the shape that is stable across scale and articulation. Comparison with our extensive user study on the MPEG-7 database, see below, demonstrates that our geometric results are consistent with user perception. This work has been accomplished within a collaboration between S. Hahmann, Kathryn Leonard (CSUCI), and Geraldine Morin and Axel Carlier (IRIT, ENSEEIHT). K. Leonard was visiting the Imagine team during several month in 2016 as an invited professor funded by the ERC Expressive grant.

We also conducted a large user-study and made the results available throughout an open access data base: The 2D Shape Structure database [9] is a public, user-generated dataset of 2D shape decompositions into a hierarchy of shape parts with geometric relationships retained. It is the outcome of a large-scale user study obtained by crowdsourcing, involving over 1200 shapes in 70 shape classes, and 2861 participants. A total of 41953 annotations has been collected with at least 24 annotations per shape. For each shape, user decompositions into main shape, one or more levels of parts, and a level of details are available. This database reinforces a philosophy that understanding shape structure as a whole, rather than in the separated categories of parts decomposition, parts hierarchy, and analysis of relationships between parts, is crucial for full shape understanding. We provide initial statistical explorations of the data to determine representative ("mean") shape annotations and to determine the number of modes in the annotations. The primary goal of this work is to make this rich and complex database openly available (through the website http://2dshapesstructure.github.io), providing the shape community with a ground truth of human perception of holistic shape structure. This paper has received the « Reproducibility Award » (http://www.reproducibilitystamp.com).

6.1.2. Surface design

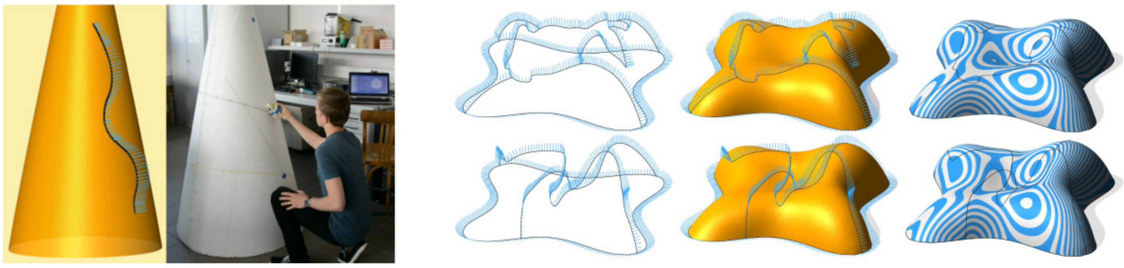


Figure 4. Left: Illustration of results of [14].

Recent surface acquisition technologies based on micro-sensors produce three-space tangential curve data which can be transformed into a network of space curves with surface normals. In the thesis of Tibor Stanko, which is funded by the CEA-LETI, we dispose such a mobile device equipped with several micro-sensors. The goal of the thesis is to develop surface acquisitions methods using this equipped mobile device. As a first step, we address the theoretical and algorithmic problem of surfacing an arbitrary closed 3D curve network with given surface normals. Thanks to the normal vector input, the patch finding problem can be solved unambiguously and an initial piecewise smooth triangle mesh is computed. The input normals are propagated throughout the mesh. Together with the initial mesh, the propagated normals are used to compute mean curvature vectors. We then compute the final mesh as the solution of a new variational optimization method based on the mean curvature vectors. The intuition behind this original approach is to guide the standard Laplacian-based variational methods by the curvature information extracted from the input normals. The normal input increases shape fidelity and allows to achieve globally smooth and visually pleasing shapes. This work has been presented at Eurographics 2016 as a short paper [25] and GTMG 2016 [26] and has been published as a journal paper [14].

Morse-Smale complexes have been proposed to visualize topological features of scalar fields defined on manifold domains. Herein, three main problems have been addressed in the past: (a) efficient computation of the initial combinatorial structure connecting the critical points; (b) simplification of these combinatorial structures; (c) reconstruction of a scalar field in accordance to the simplified Morse-Smale complex. The PhD thesis of Leo Allemand-Giorgis faces the third problem by proposing a novel approach for computing a scalar field coherent with a given simplified MS complex that privileges the use of piecewise polynomial functions [31]. Based on techniques borrowed from shape preserving design in Computer Aided Geometric Design, our method constructs the surface cell by cell using piecewise polynomial curves and surfaces.

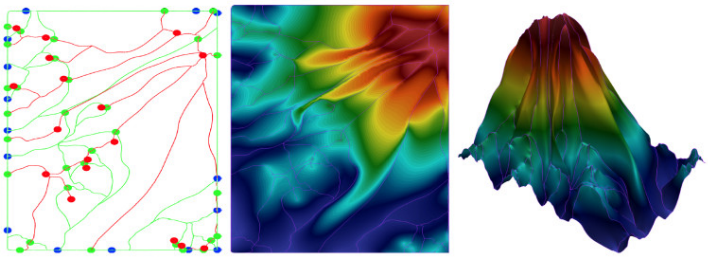


Figure 5. The terrain data set of Mt Rainier: Surface reconstruction (c) with contour lines (b) from the MS complex with 69 critical points(a) [31].

6.2. Motion & Sound Synthesis

- **Scientist in charge**: François Faure.
- **Other permanent researchers**: Marie-Paule Cani, Damien Rohmer, Rémi Ronfard.

Animating objects in real-time is mandatory to enable user interaction during motion design. Physically-based models, an excellent paradigm for generating motions that a human user would expect, tend to lack efficiency for complex shapes due to their use of low-level geometry (such as fine meshes). Our goal is therefore two-folds: first, develop efficient physically-based models and collisions processing methods for arbitrary passive objects, by decoupling deformations from the possibly complex, geometric representation; second, study the combination of animation models with geometric responsive shapes, enabling the animation of complex constrained shapes in real-time. The last goal is to start developing coarse to fine animation models for virtual creatures, towards easier authoring of character animation for our work on narrative design.

6.2.1. Physically-based models

We proposed a survey on the exhisting adaptative physically based models in Computer Graphics in collaboration with IST Austria, University of Minnesota, and NANO-D Inria team. Models where classified according to the strategy they use for adaptation, from time-stepping and freezing techniques to geometric adaptivity in the form of structured grids, meshes, and particles. Applications range from fluids, through deformable bodies, to articulated solids. The survey has been published as a Eurographics state of the art [13].

In collaboration with the *Reproduction et Développement des Plantes* Lab (ENS Lyon), we proposed a realistic three-dimensional mechanical model of the indentation of a flower bud using the SOFA library, in order to provide a framework for the analysis of force-displacement curves obtained experimentally [12].

6.2.2. Simulating paper material with sound

We developed within the PhD from Camille Schreck a dedicated approach to model a real time deforming virtual sheet of paper. First we developed a geometrical model interleaving physically based elastic deformation with a dedicated geometrical correction and remeshing. The key idea consists in modeling the surface using a set of generalized cones able to model developable ruled surfaces instead of the more traditional set of triangles. This surface can handle length preservation with respect to the 2D pattern, and permanent non smooth crumpling appearance. This geometrical model published in ACM Transactions on Graphics in Dec. 2015 [5] has been presented at ACM SIGGRAPH this summer and is currently under investigation to be part of Inria Showroom. This model has then been extended to real time sound synthesis of crumpled paper within the collaboration with Doug James (Stanford University). This method was the first to handle real-time shape dependent sound synthesis. During the interactive deformation, sudden curvature changes and friction are detected. These sound generating events are then associated to a geometrical region where the sound resonates and defined efficiently using previous geometrical model. Finally, the sound is synthesized using a

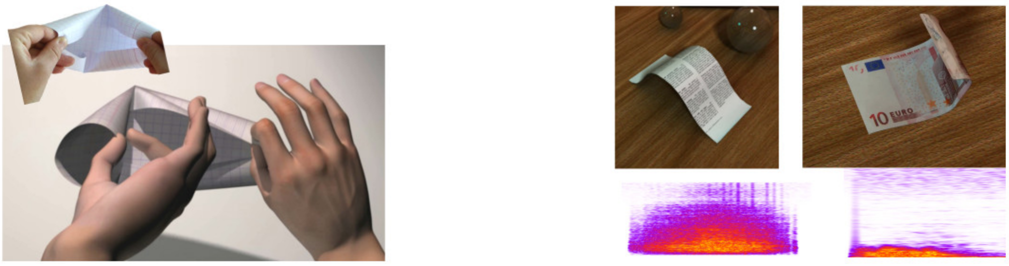


Figure 6. Left: Geometrical deformation using our geometrical model from [5]. Right: Various paper deformation and type leading to different synthesized sounds [24].

pre-recorded sound data base of crumple and friction events sorted with respect to the resonator region size. This work has been published at Symposium on Computer Animation [24] and received the best paper award.

6.2.3. Human motion

Figure 7. Live tracking and visualization of a plausible anatomical skeleton following the pose the subject [17].

Armelle Bauer defended her PhD in November, co-advised with TIMC (Jocelyne Troccaz as principal advisor), on Augmented Reality for the interactive visualization of human anatomy. This is one of the main achievements of the Living Book of Anatomy project, funded by Labex Persyval. This work was partly published at the Motion in Games conference (MIG 2016) [17]. It served as a basis for the follow-up ANR project Anatomy2020 involving Anatoscope, TIMC and LIG laboratories, and Univ Lyon 2.

6.3. Knowledge-based Models for Narrative Design

- **Scientist in charge**: Rémi Ronfard.
- **Other permanent researchers**: Marie-Paule Cani, Frédéric Devernay, François Faure, Jean-Claude Léon, Olivier Palombi.

Our long term goal is to develop high-level models helping users to express and convey their own narrative content (from fiction stories to more practical educational or demonstrative scenarios). Before being able to specify the narration, a first step is to define models able to express some a priori knowledge on the background scene and on the object(s) or character(s) of interest. Our first goal is to develop 3D ontologies able to express such knowledge. The second goal is to define a representation for narration, to be used in future storyboarding frameworks and virtual direction tools. Our last goal is to develop high-level models for virtual cinematography such as rule-based cameras able to automatically follow the ongoing action and semi-automatic editing tools enabling to easily convey the narration via a movie.

6.3.1. Virtual cameras

Filming live action requires a coincidence of many factors: actors of course, but also lighting, sound capture, set design, and finally the camera (position, frame, and motion). Some of these, such as sound and lighting, can be more or less reworked in post-production, but the camera parameters are usually considered to be fixed at shooting time. We developed two kinds of image-based rendering technique, which allows to change in post-production either the camera frame (in terms of pan, tilt, and zoom), or the camera position.

To be able to change the camera frame after shooting, we developed techniques to construct a video panorama from a set of cameras placed roughly at the same position. Video panorama exhibits a specific problem, which is not present in photo panorama: because the projection centers of the cameras can not physically be at the same location, there is residual parallax between the video sequences, which produce artifacts when the videos are stitched together. Sandra Nabil has worked during her PhD on producing video panoramas without visible artifacts, which can be used to freely pick the camera frame in terms of pan, tilt and zoom during the post-production phase.

Modifying the camera position itself is an even greater challenge, since it either requires a perfect 3D reconstruction of the scene or a dense sampling of the 4D space of optical rays at each time (called the 4D lightfield). During the PhD of Gregoire Nieto, we developed image-based rendering (IBR) techniques which are designed to work in cases where the 3D reconstruction cannot be obtained with a high precision, and the number of cameras used to capture the scene is low, resulting in a sparse sampling of the 4D lightfield.

6.3.2. Virtual actors

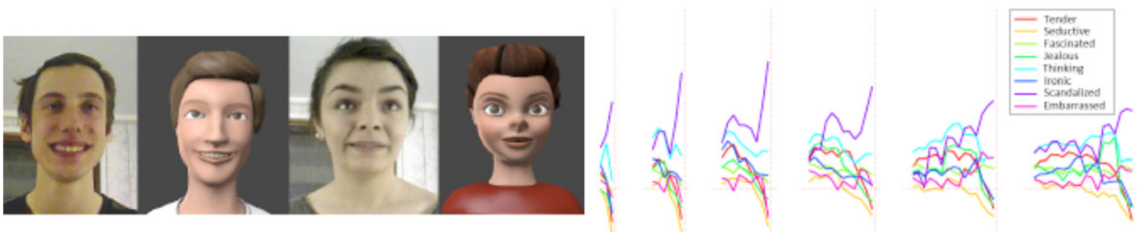


Figure 8. Left: Examples of video and animation frames for a dramatic attitude (seductive) played by two semi-professional actors. Right: Prosodic contours for 8 dramatic attitudes, showing evidence that "scandalized" and "thinking" strongly stand out from other attitudes.

Following up on Adela Barbelescu's PhD thesis, we tested the capability of audiovisual parameters (voice frequency, rhythm, head motion and facial expressions) to discriminate among different dramatic attitudes in both real actors (video) and virtual actors (3D animation). Using Linear Discriminant Analysis classifiers, we showed that sentence-level features present a higher discriminating rate among the attitudes and are less dependent on the speaker than frame and sylable features. We also performed perceptual evaluation tests, showing that voice frequency is correlated to the perceptual results for all attitudes, while other features, such as head motion, contribute differently, depending both on the attitude and the speaker. Those new results were presented at the Interspeech conference [16].

6.4. Creating and Interacting with Virtual Prototypes

- **Scientist in charge**: Jean-Claude Léon.

- **Other permanent researchers**: Marie-Paule Cani, Frédéric Devernay, Olivier Palombi, Damien Rohmer, Rémi Ronfard.

The challenge is to develop more effective ways to put the user in the loop during content authoring. We generally rely on sketching techniques for quickly drafting new content, and on sculpting methods (in the sense of gesture-driven, continuous distortion) for further 3D content refinement and editing. The objective is to extend these expressive modeling techniques to general content, from complex shapes and assemblies to animated content. As a complement, we are exploring the use of various 2D or 3D input devices to ease interactive 3D content creation.

6.4.1. Sculpting Virtual Worlds

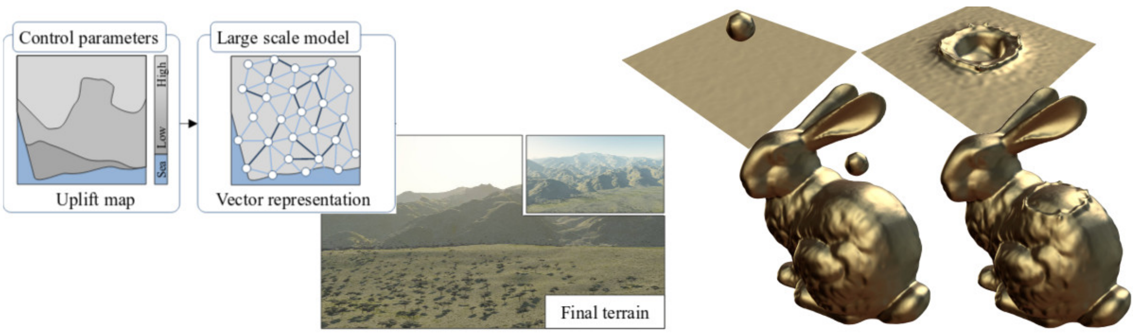


Figure 9. Left: Generating a large-scale terrain following fluvial erosion principle from a coarse set of control parameters [10]. Right: Copy of an animated drop of fluid and its effect on the underlying surface into another animation [23].

Extending expressive modeling paradigms to full virtual worlds, with complex terrains, streams and oceans, and vegetation is a challenging goal. To achieve this, we need to combine procedural methods that accurately simulate physical, geological and biological phenomena shaping the world, which high level user control. This year, our work in the area was three-folds:

Firstly, in collaboration with Jean Braun, professor in geo-morphology and other colleagues, we designed the fist efficient simulation method able to take into account large-scale fluvial erosion to shape mountains. This method was published at Eurographics [10]. We also designed an interactive sculpting system with multi-touch finger interaction, able to shape mountain ranges based on tectonic forces. This method, combined in real-time with our erosion simulation process, was submitted for a journal publication.

Secondly, we extended the "Worldbrush" system proposed in 2015 (Emilien et al, Siggraph 2015) in order to consistently populate virtual worlds with learned statistical distributions of trees and plants. The main contributor to this project was James Gain, our visiting professor. After clustering the input terrain into a number of characteristic environmental conditions, we computed sand-box (small-scale) simulations of ecosystems (plant growth) for each of these conditions, and then used learned statistical models (an extension of worldbrush) to populate the full terrain with consistent sets of species. This work was submitted for publication.

Third, we extended interactive sculpting paradigms to the sculpting of liquid simulation results, such as editing waves on a virtual ocean [23]. Liquid simulations are both compute intensive and very hard to control, since they are typically edited by re-launching the simulations with slightly different initial conditions until the user is satisfied. In contrast, our method enables users to directly edit liquid animation results (coming in the form of animated meshes) in order to directly output new animations. More precisely, the method offer semi-automatic clustering methods enabling users to select features such as droplets and waves, edit them in space and time and them paste them back into the current liquid animation or to another one.

6.4.2. Sketch based design

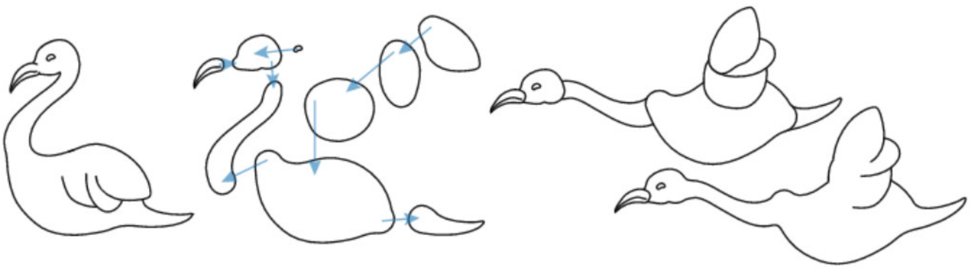


Figure 10. Progressive extraction of sub-parts of an input sketch in depth with automatic contour completion [37].

Using 2D sketches is one of the easiest way for creating 3D contents. While prior knowledge on the object being sketch can be used to retrieve the missing information, and thus consistently inferring 3D, interpreting more general sketches and generating 3D shapes from them in indeed a challenging long-term goals. This year, our work in the area was two-folds:

Firstly, we participated to a course on Sketch-based Modeling, presented at both Eurographics 2016 and Siggraph Asia 2016 [18]. The parts we worked on was sketch-based modeling from prior knowledge, with the examples of our works on animals, garments (developable surfaces) and trees.

Secondly, we advanced towards the interpretation of general sketches representing smooth, organic shapes. The key features of our methods are a new approach for aesthetic contour completion, and an interactive algorithm for progressively interpreting internal silhouettes (suggestive contours) in order to progressively extract sub-parts of the shape from the drawing. These parts are ordered in depth. Our first results were presented as a poster at the Siggraph 2016 conference [37], and then extended an submitted for publication. We are now extending them towards the inference of 3D, organic shapes from a 2D sketch.

MANAO Project-Team

7. New Results

7.1. Analysis and Simulation

7.1.1. Principles of Light Field Imaging

Light field imaging offers powerful new capabilities through sophisticated digital processing techniques that are tightly merged with unconventional optical designs. This combination of imaging technology and computation necessitates a fundamentally different view of the optical properties of imaging systems and poses new challenges for the traditional signal and image processing domains. We aimed to provide a comprehensive review [14] of the considerations involved and the difficulties encountered in working with light field data during 25 years of research.

7.1.2. Physically-Based Reflectance Model Combining Reflection and Diffraction

Reflectance properties express how objects in a virtual scene interact with light; they control the appearance of the object: whether it looks shiny or not, whether it has a metallic or plastic appearance. Having a good reflectance model is essential for the production of photo-realistic pictures. Measured reflectance functions provide high realism at the expense of memory cost. Parametric models are compact, but finding the right parameters to approximate measured reflectance can be difficult. Most parametric models use a model of the surface micro-geometry to predict the reflectance at the macroscopic level. We have shown [26] that this micro-geometry causes two different physical phenomena: reflection and diffraction. Their relative importance is connected to the surface roughness. Taking both phenomena into account, we develop a new reflectance model that is compact, based on physical properties and provides a good approximation of measured reflectance.

7.1.3. Multi-Scale and Structured SV-BRDF Model for Scratched Materials

We developed a Spatially-Varying BRDF model tailored to the multi-scale rendering of scratched materials such as metals, plastics or finished woods. Our approach takes advantage of the regular structure of scratch distributions to achieve high performance without compromising visual quality (fig. 8). The BRDF for a single scratch is simulated using an optimized 2D ray-tracer and compactly stored in a three-component 2D texture. In contrast to existing models, our approach takes into account all inter-reflections inside a scratch, including Fresnel effects. At render time, the SV-BRDF for the scratch distribution under a pixel or ray footprint is obtained by linear combination of individual scratch BRDFs. Our model can be evaluated using both importance and light sampling, in direct and global illumination settings. Our approach provides users with controls over the profile, micro-BRDF, density and orientation of scratches. All these material properties are updated at interactive rates. This work has been published at ACM Siggraph 2016 [16] and one our result has been selected as the cover of the ACM Siggraph 2016 proceedings. It is part of the PhD Thesis "Control of anisotropic materials appearance" [11] defended this year.

7.1.4. Cues for Perception of Appearance

Thanks for the FP7 ITN PRISM, we have participated to several user studies to understand the perception of an object appearance. First, for fluids and other deformable materials, we find [25] that observers show a high degree of constancy in matching the viscosity across the different variations. However, volume differences between test and match stimulus, especially with static stimuli, caused large effects of over- and under-estimation of viscosity. We also find that a number of cues related to curvatures, periodic movements of the liquids, and the way they spread out predict aspects of the observer's performance, but that humans achieve better constancy than the cues predict.

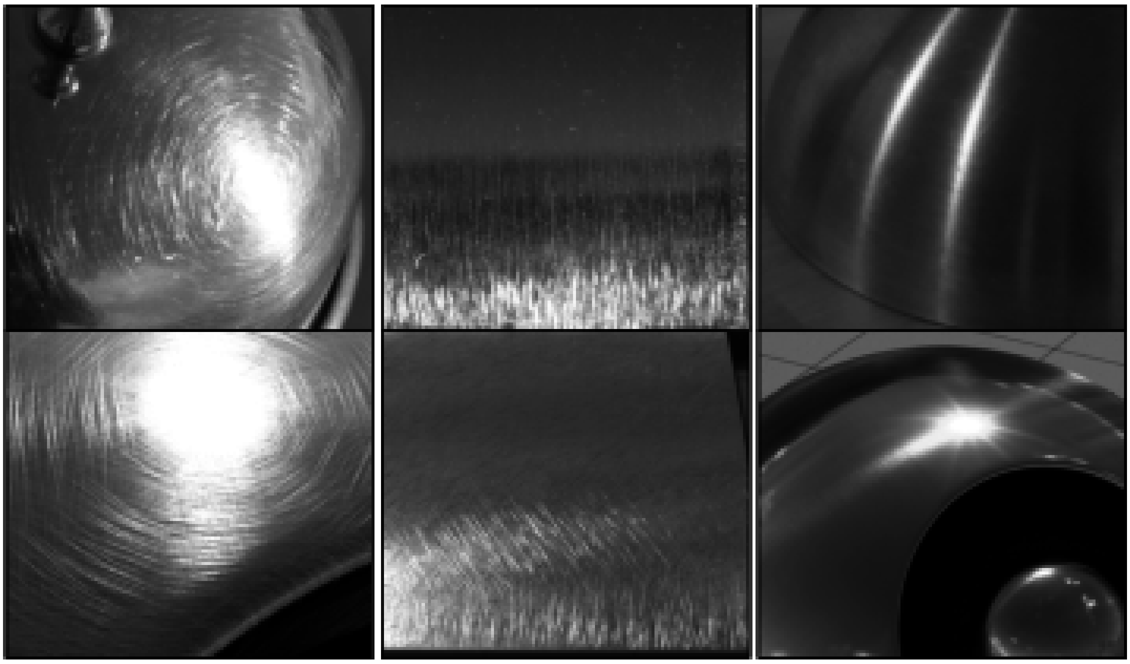


Figure 8. Our scratch BRDF (top) can reproduce several effects similar to real photographs (bottom).

We have also investigated gloss haze [22]. The results reveals that haziness is a distinct visual dimension orthogonal to the commonly studied glossiness and blurriness. Coatedness appears to be nearly synonymous with haziness, as this is one of the main physical causes of haze in real world materials. Polish seems to be a combination of glossiness and haziness, as materials go from dull to hazy to highly glossy during the physical polishing process. The inferred tactile quality of friction is apparently uncorrelated with haziness. Our results demonstrate that haze is indeed a distinct perceptual dimension of gloss, which is systematically related to the kurtosis of the specular lobe.

7.2. From Acquisition to Display

7.2.1. Spatial Augmented Reality

Spatial augmented reality allows to improve or modify the perception of the reality with virtual information displayed directly in the real world, using video-projection. Many fields such as tourism, entertainment, education, medicine, industry or cultural heritage may benefit from it. Recent computer science techniques allow to measure, analyse and visualise the geometry of the surface of real objects, as for instance archeological artefacts. We have proposed a SAR interaction and visualisation technique (part of the PhD thesis "Interaction techniques, personalized experience and surface reconstruction for spatial augmented reality" [12] defended this year) that combines the advantages of the study of both real and 3D archeological artefacts. Thus, we superimpose on the object an expressive rendering based on curvatures with SAR, allowing for example to show details of engravings. Next, we simulate the use of a flashlight with the help of a 6-degree-of-freedom controller. The user can then specify the area on the object to be augmented and adjust the various necessary parameters of the expressive rendering. One of the main caracteristics of SAR is to enable multiple users to simultaneously participate to the same experience. However, depending on the target application, this can be seen as a drawback.

We have also proposed a new display device [27] that allows to create experiences in SAR that are both multiuser and personalised by taking into account the user point of view. In order to do so, the projection display, set in front of the object to augment, is made from a material that is both retro-reflective and semi-transparent. We suggest two different uses of this new device, as well as two scenarios of application.

7.2.2. Isotropic BRDF Measurements

Image-based BRDF measurements on spherical material samples present a great opportunity to shorten significantly the acquisition time with respect to more traditional, non-multiplexed measurement methods for isotropic BRDFs. However, it has never been analyzed deeply, what measurement accuracy can be achieved in such a setup; what are the main contributing uncertainty factors and how do they relate to calibration procedures. We have developed [20] a new set of isotropic BRDF measurements with their radiometric and geometric uncertainties acquired within such an imaging setup. We have analyzed the most prominent optical phenomena that affect measurement accuracy and pave the way for more thorough uncertainty analysis in forthcoming image-based BRDF measurements. Our newly acquired data with their quantified uncertainties will be helpful for comparing the quality and accuracy of the different experimental setups and for designing other such image-based BRDF measurement devices.

7.3. Rendering, Visualization and Illustration

7.3.1. Cache-friendly Sampling

Monte-Carlo integration techniques for global illumination are popular on GPUs thanks to their massive parallel architecture, but efficient implementation remains challenging. The use of randomly de-correlated low-discrepancy sequences in the path-tracing algorithm allows faster visual convergence. However, the parallel tracing of incoherent rays often results in poor memory cache utilization, reducing the ray bandwidth efficiency. Interleaved sampling [65] partially solves this problem, by using a small set of distributions split in coherent ray-tracing passes, but the solution is prone to structured noise. On the other hand, ray-reordering methods [83] group stochastic rays into coherent ray packets but their implementation add an additional sorting cost on the GPU [74], [50]. We have introduced [19] a micro-jittering technique for faster multi-dimensional Monte-Carlo integration in ray-based rendering engines. Our method, improves ray coherency between GPU threads using a slightly altered low-discrepancy sequence rather than using ray-reordering methods. Compatible with any low-discrepancy sequence and independent of the importance sampling strategy, our method achieves comparable visual quality with classic de-correlation methods, like Cranley-Patterson rotation [66], while reducing rendering times in all scenarios.

7.3.2. Multi-Resolution Meshes for Feature-Aware Hardware Tessellation

Hardware tessellation is de facto the preferred mechanism to adaptively control mesh resolution with maximal performances. However, owing to its fixed and uniform pattern, leveraging tessellation for feature-aware LOD rendering remains a challenging problem. In [15], we relax this fundamental constraint by introducing a new spatial and temporal blending mechanism of tessellation levels, which is built on top of a novel hierarchical representation of multi-resolution meshes. This mechanism allows to finely control topological changes so that vertices can be removed or added at the most appropriate location to preserve geometric features in a continuous and artifact-free manner (cf. Figure 9). We then show how to extend edge-collapse based decimation methods to build feature-aware multi-resolution meshes that match the tessellation patterns. Our approach is fully compatible with current hardware tessellators and only adds a small overhead on memory consumption and tessellation cost. This work as been published at Eurographics 2016 [15].

7.3.3. Shape Depiction for Transparent Objects

Shading techniques are useful to deliver a better understanding of object shapes. When transparent objects are involved, depicting the shape characteristics of each surface is even more relevant. We have developed [21] a method for rendering transparent scenes or objects using classical tools for shape depiction in real time. Our method provides an efficient way to compute screen space curvature on transparent objects by using a novel

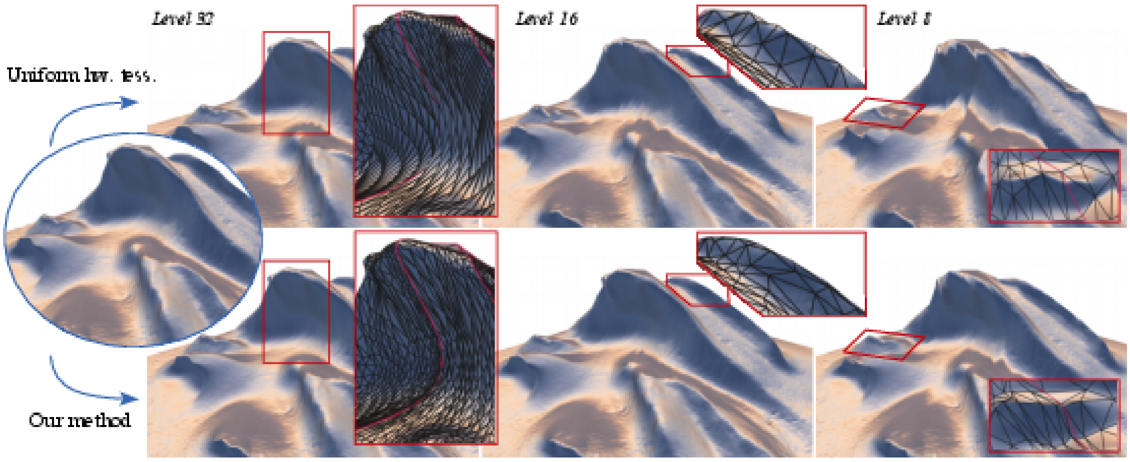


*Figure 9. Uniform hardware tessellation **(top)** fails at representing accurately sharp features and areas of high curvature, such as the top and deep part of the drifts, which produces tessellation artifacts. Our method **(bottom)** better preserves those regions by adapting the triangle size and aligning their edges with those features.*

screen space representation of a scene derived from Order Independent Transparency techniques. Moreover, we propose a customizable stylization that modulates the transparency per fragment, according to its curvature and its depth, which can be adapted for various kinds of applications.

7.4. Editing and Modeling

7.4.1. Flow-guided Warping for Image-based Shape Manipulation

Manipulating object shape in images usually require a-priori on their 3D geometry, and either user interactions or huge databases of 3D objects. In collaboration with the Maverick team (Inria Rhone Alpes), we have developped a method that manipulates perceived object shape from a single input color image without the need of addional 3D information, user input or 3D data. The key idea is to give the illusion of shape sharpening or rounding by exaggerating orientation patterns in the image that are strongly correlated to surface curvature (fig. 10). We build on a growing literature in both human and computer vision showing the importance of orientation patterns in the communication of shape, which we complement with mathematical relationships and a statistical image analysis revealing that structure tensors are indeed strongly correlated to surface shape features. We then rely on these correlations to introduce a flow-guided image warping algorithm, which in effect exaggerates orientation patterns involved in shape perception. We evaluate our technique by 1) comparing it to ground truth shape deformations, and 2) performing two perceptual experiments to assess its effects. Our algorithm produces convincing shape manipulation results on synthetic images and photographs, for various materials and lighting environments. This work has been published at ACM Siggraph 2016 [17].

7.4.2. Local Shape Editing at the Compositing Stage

Modern compositing software permit to linearly recombine different 3D rendered outputs (e.g., diffuse and reflection shading) in post-process, providing for simple but interactive appearance manipulations. Renderers also routinely provide auxiliary buffers (e.g., normals, positions) that may be used to add local light sources or depth-of-field effects at the compositing stage. These methods are attractive both in product design and movie production, as they allow designers and technical directors to test different ideas without having to re-render an entire 3D scene. In this work, we extended this approach to the editing of local shape: users modify the rendered normal buffer, and our system automatically modifies diffuse and reflection buffers to

(a) Input image - ©Expertissim (b) Shape sharpening (c) Shape rounding

Figure 10. Our warping technique takes as input (a) a single image (Jules Bennes, after Barye: "walking lion") and modifies its perceived surface shape, either making it sharper in (b) or rounder in (c).

provide a plausible result. Our method is based on the reconstruction of a pair of diffuse and reflection prefiltered environment maps for each distinct object/material appearing in the image. We seamlessly combine the reconstructed buffers in a recompositing pipeline that works in real-time on the GPU using arbitrarily modified normals. This work as been published at the Eurographics Symposium on Rendering [24].

7.4.3. Topology-Aware Neighborhoods for Point-Based Simulation and Reconstruction

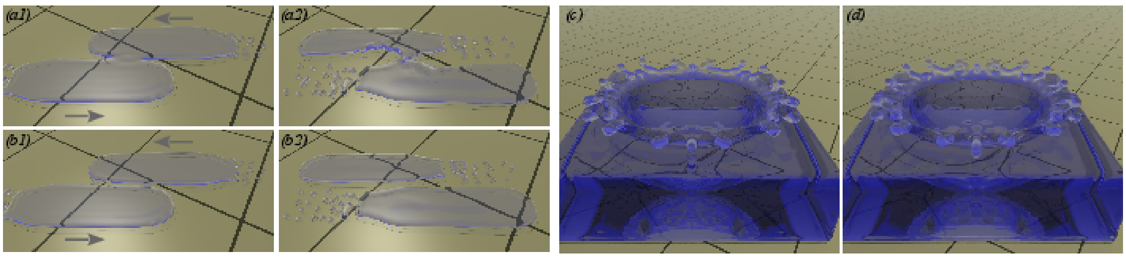


Figure 11. Two SPH fluid simulations using a standard Euclidean particle neighborhood (a,c), and our new topological neighborhood (b,d). On the left, two fluid components are crossing while moving in opposite directions. Our new neighborhood performs accurate merging computations and avoids both unwanted fusion in the reconstruction and incorrect fluid interaction in the simulation. On the right, our accurate neighborhoods lead to different shape of the splash, and enable the reconstruction of the fluid with an adequate topology while avoiding bulging at distance.

Particle based simulations are widely used in computer graphics. In this field, several recent results have improved the simulation itself or improved the tension of the final fluid surface. In current particle based implementations, the particle neighborhood is computed by considering the Euclidean distance between fluid particles only. Thus particles from different fluid components interact, which generates both local incorrect behavior in the simulation and blending artifacts in the reconstructed fluid surface. In collaboration with IRIT, we developed a better neighborhood computation for both the physical simulation and surface reconstruction steps (fig. 11). Our approach tracks and stores the local fluid topology around each particle using a graph structure. In this graph, only particles within the same local fluid component are neighbors and other disconnected fluid particles are inserted only if they come into contact. The graph connectivity also takes into account the asymmetric behavior of particles when they merge and split, and the fluid surface is reconstructed accordingly, thus avoiding their blending at distance before a merge. In the simulation, this neighborhood information is exploited for better controlling the fluid density and the force interactions at the vicinity of its boundaries. For instance, it prevents the introduction of collision events when two distinct fluid components are crossing without contact, and it avoids fluid interactions through thin waterproof walls. This leads to an overall more consistent fluid simulation and reconstruction. This work as been published at the Eurographics/ ACM SIGGRAPH Symposium on Computer Animation [18].

MAVERICK Project-Team

7. New Results

7.1. Computer-aided image manipulation

7.1.1. Automatic lighting design from photographic rules

Participants: Jérémy Wambecke, Romain Vergne, Georges-Pierre Bonneau, Joëlle Thollot.

Figure 2. Our lighting setup produces realistic images for any kind of opaque surfaces, where shapes of objects are always properly conveyed.

Lighting design is crucial in 3D scenes modeling for its ability to provide cues to understand the objects shape. However a lot of time, skills, trials and errors are required to obtain a desired result. Existing automatic lighting methods for conveying the shape of 3D objects are based either on costly optimizations or on non-realistic shading effects. Also they do not take the material information into account. In this work, we propose a new method that automatically suggests a lighting setup to reveal the shape of a 3D model, taking into account its material and its geometric properties (see Figure 2). Our method is independent from the rendering algorithm. It is based on lighting rules extracted from photography books, applied through a fast and simple geometric analysis. We illustrate our algorithm on objects having different shapes and materials, and we show by both visual and metric evaluation that it is comparable to optimization methods in terms of lighting setups quality. Thanks to its genericity our algorithm could be integrated in any rendering pipeline to suggest appropriate lighting. It has been published in WICED'2016 [8].

7.1.2. Automatic Texture Guided Color Transfer and Colorization

Participants: Benoit Arbelot, Romain Vergne, Thomas Hurtut, Joëlle Thollot.

This work targets two related color manipulation problems: *Color transfer* for modifying an image colors and *colorization* for adding colors to a greyscale image. Automatic methods for these two applications propose to modify the input image using a reference that contains the desired colors. Previous approaches usually do not target both applications and suffer from two main limitations: possible misleading associations between input and reference regions and poor spatial coherence around image structures. In this work, we propose a unified framework that uses the textural content of the images to guide the color transfer and colorization (see Figure 3). Our method introduces an edge-aware texture descriptor based on region covariance, allowing for local color transformations. We show that our approach is able to produce results comparable or better than state-of-the-art methods in both applications. It has been published in Expressive'2016 [4] and an extended version has been submitted to C&G.

7.1.3. Flow-Guided Warping for Image-Based Shape Manipulation

Participants: Romain Vergne, Pascal Barla, Georges-Pierre Bonneau, Roland W. Fleming.

Color transfer Colorization

Figure 3. Our framework allows for automatic local color transfer (left) and colorization (right) based on textural properties.

(a) Input image - ©Expertissim (b) Shape sharpening (c) Shape rounding

Figure 4. Our warping technique takes as input (a) a single image (Jules Bennes, after Barye: "walking lion") and modifies its perceived surface shape, either making it sharper in (b) or rounder in (c).

We present an interactive method that manipulates perceived object shape from a single input color image thanks to a warping technique implemented on the GPU. The key idea is to give the illusion of shape sharpening or rounding by exaggerating orientation patterns in the image that are strongly correlated to surface curvature. We build on a growing literature in both human and computer vision showing the importance of orientation patterns in the communication of shape, which we complement with mathematical relationships and a statistical image analysis revealing that structure tensors are indeed strongly correlated to surface shape features. We then rely on these correlations to introduce a flow-guided image warping algorithm, which in effect exaggerates orientation patterns involved in shape perception. We evaluate our technique by 1) comparing it to ground truth shape deformations, and 2) performing two perceptual experiments to assess its effects. Our algorithm produces convincing shape manipulation results on synthetic images and photographs, for various materials and lighting environments (see Figure 4). This work has been published in ACM TOG 2016 [3].

7.1.4. Local Shape Editing at the Compositing Stage

Participants: Carlos Jorge Zubiaga Peña, Gael Guennebaud, Romain Vergne, Pascal Barla.

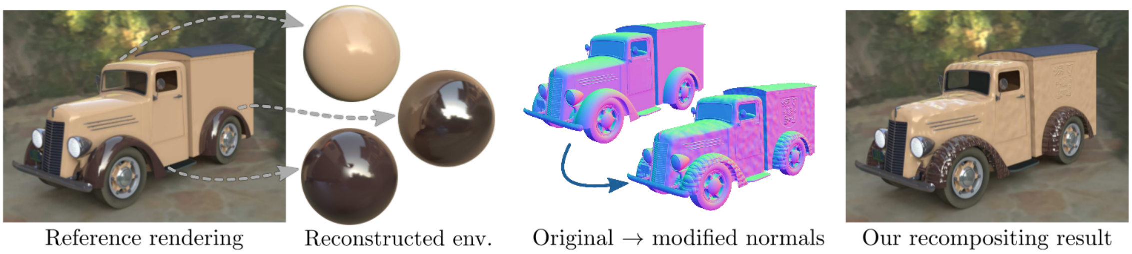


Figure 5. Our method permits to modify surface shape by making use of the shading and auxiliary buffers output by modern renderers. We first reconstruct shading environments for each object/material combination of the Truck scene, relying on normal and shading buffers. When normals are then modified by the compositing artist, the color image is recomposited in real-time, enabling interactive exploration. Our method reproduces inter-reflections between objects, as seen when comparing the reconstructed environments for rear and front mudguards.

Modern compositing software permit to linearly recombine different 3D rendered outputs (e.g., diffuse and reflection shading) in post-process, providing for simple but interactive appearance manipulations. Renderers also routinely provide auxiliary buffers (e.g., normals, positions) that may be used to add local light sources or depth-of-field effects at the compositing stage. These methods are attractive both in product design and movie production, as they allow designers and technical directors to test different ideas without having to re-render an entire 3D scene. We extend this approach to the editing of local shape: users modify the rendered normal buffer, and our system automatically modifies diffuse and reflection buffers to provide a plausible result (see Figure 5). Our method is based on the reconstruction of a pair of diffuse and reflection prefiltered environment maps for each distinct object/material appearing in the image. We seamlessly combine the reconstructed buffers in a recompositing pipeline that works in real-time on the GPU using arbitrarily modified normals. This work has been published in EGSR (EI & I) 2016 [13].

7.1.5. Map Style Formalization: Rendering Techniques Extension for Cartography

Participants: Hugo Loi, Benoit Arbelot, Romain Vergne, Joëlle Thollot.

Cartographic design requires controllable methods and tools to produce maps that are adapted to users' needs and preferences. The formalized rules and constraints for cartographic representation come mainly from the conceptual framework of graphic semiology. Most current Geographical Information Systems (GIS) rely on

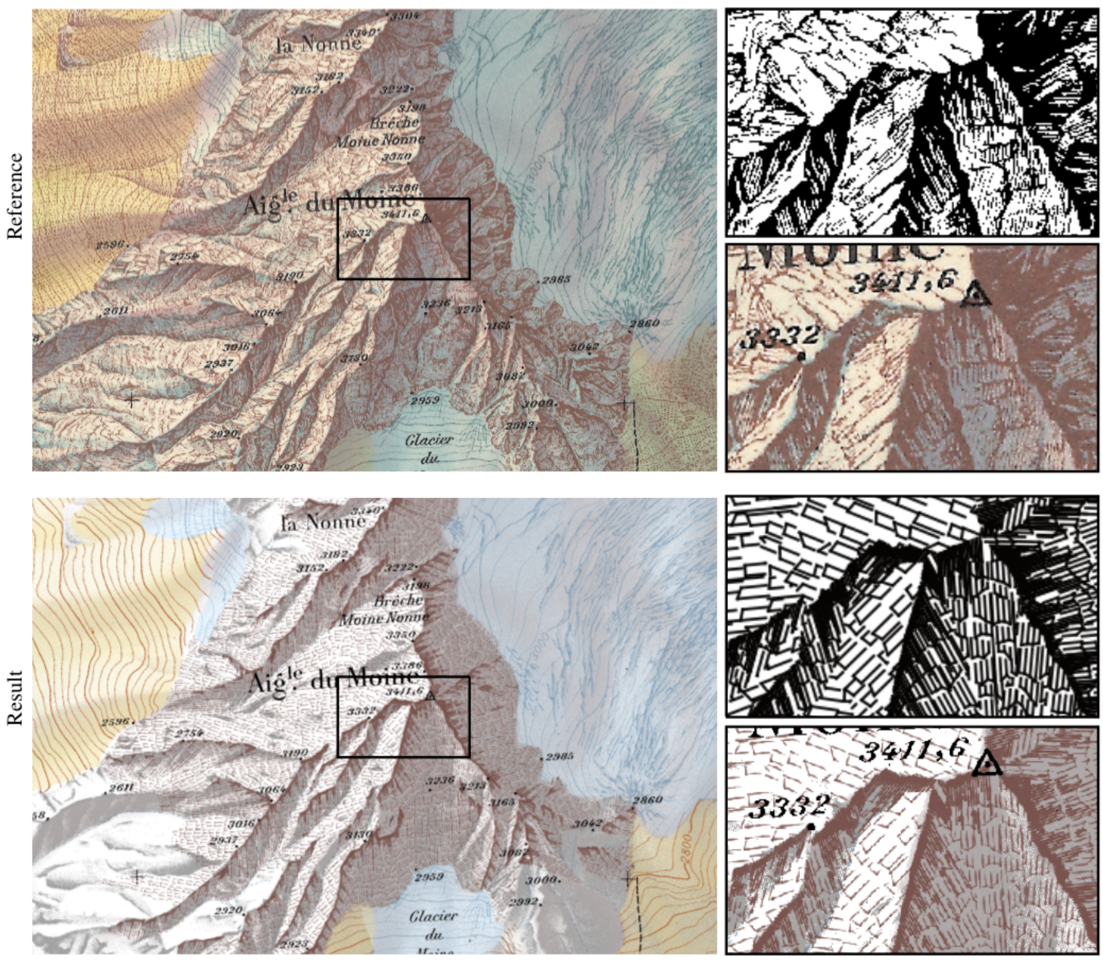


Figure 6. Reference and Resulting Mountain map “Aiguille du Moine”, 1:10k scale: extracts of reference (first line) and resulting rocky areas (second line): on the right, zooms on, first the hatching primitives, second the stylized same ones. For a fair comparison, we provide resulting map at a resolution similar to the reference map.

the Styled Layer Descriptor and Semiology Encoding (SLD/SE) specifications which provide an XML schema describing the styling rules to be applied on geographic data to draw a map. Although this formalism is relevant for most usages in cartography, it fails to describe complex cartographic and artistic styles. In order to overcome these limitations, we propose an extension of the existing SLD/SE specifications to manage extended map stylizations, by the means of controllable expressive methods. Inspired by artistic and cartographic sources (Cassini maps, mountain maps, artistic movements, etc.), we propose to integrate into our system three main expressive methods: linear stylization, patch-based region filling and vector texture generation. We demonstrate how our pipeline allows to personalize map rendering with expressive methods in several examples. This work is the result of the MAPSTYLE ANR and has been published at Expressive 20016 [5].

7.2. Illumination Simulation and Materials

7.2.1. A Physically-Based Reflectance Model Combining Reflection and Diffraction

Participant: Nicolas Holzschuch.

Reflectance properties express how objects in a virtual scene interact with light; they control the appearance of the object: whether it looks shiny or not, whether it has a metallic or plastic appearance. Having a good reflectance model is essential for the production of photo-realistic pictures. Measured reflectance functions provide high realism at the expense of memory cost. Parametric models are compact, but finding the right parameters to approximate measured reflectance can be difficult. Most parametric models use a model of the surface micro-geometry to predict the reflectance at the macroscopic level. We have shown that this micro-geometry causes two different physical phenomena: reflection and diffraction. Their relative importance is connected to the surface roughness. Taking both phenomena into account, we developped a new reflectance model that is compact, based on physical properties and provides a good approximation of measured reflectance (See Figure 7).

7.2.2. A Robust and Flexible Real-Time Sparkle Effect

Participant: Beibei Wang.

We present a fast and practical procedural sparkle effect for snow and other sparkly surfaces which we integrated into a recent video game. Following from previous work, we generate the sparkle glints by intersecting a jittered 3D grid of sparkle seed points with the rendered surface. By their very nature, the sparkle effect consists of high frequencies which must be dealt with carefully to ensure an anti-aliased and noise free result (See Figure 8). We identify a number of sources of aliasing and provide effective techniques to construct a signal that has an appropriate frequency content ready for sampling at pixels at both foreground and background ranges of the scene. This enables artists to push down the sparkle size to the order of 1 pixel and achieve a solid result free from noisy flickering or other aliasing problems, with only a few intuitive tweakable inputs to manage [9].

7.2.3. Capturing Spatially Varying Anisotropic Reflectance Parameters using Fourier Analysis

Participants: Nicolas Holzschuch, Alban Fichet.

Reflectance parameters condition the appearance of objects in photorealistic rendering. Practical acquisition of reflectance parameters is still a difficult problem. Even more so for spatially varying or anisotropic materials, which increase the number of samples required. We present an algorithm for acquisition of spatially varying anisotropic materials, sampling only a small number of directions. Our algorithm uses Fourier analysis to extract the material parameters from a sub-sampled signal. We are able to extract diffuse and specular reflectance, direction of anisotropy, surface normal and reflectance parameters from as little as 20 sample directions (See Figure 9). Our system makes no assumption about the stationarity or regularity of the materials, and can recover anisotropic effects at the pixel level. This work has been published at Graphics Interface 2016 [6].

7.2.4. Estimating Local Beckmann Roughness for Complex BSDFs

Participant: Nicolas Holzschuch.

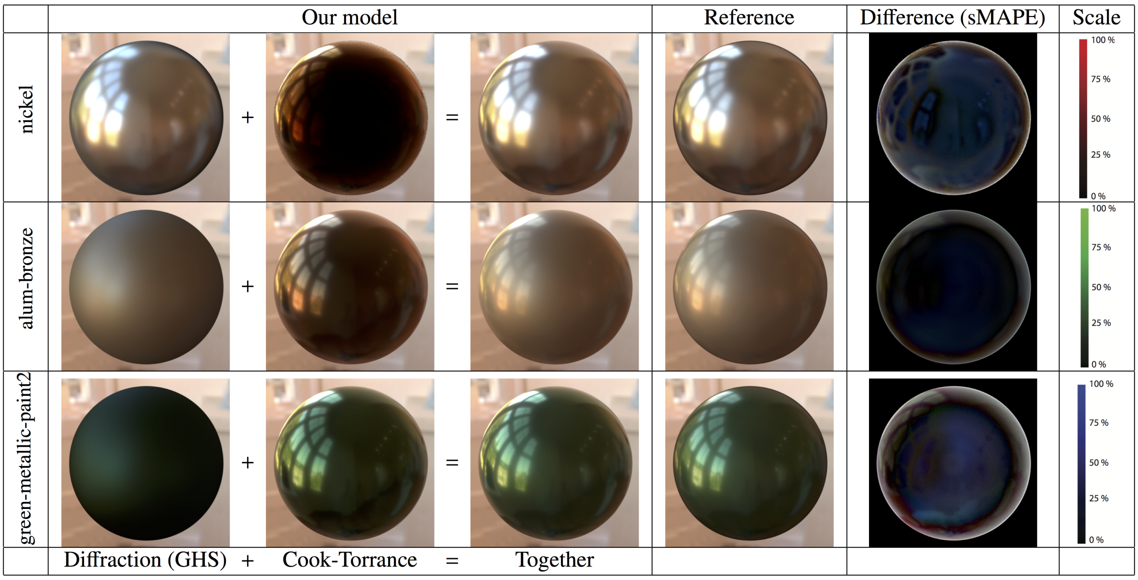


Figure 7. Surface micro-geometry contributes to its visible aspect (material reflectance). Two physical phenomena are acting together: reflection on micro-facets and diffraction. Our reflectance model combines them, with the proper energy repartition between them. The importance of diffraction depends on the roughness of the material. Even when it it relatively small, as for `green-metallic-paint2`*, it has a significant impact of the aspect of the material. Our model explains even a very difficult material like* `alum-bronze` *(middle row) as a single material.*

Figure 8. Two scenes rendered with our sparkle effect

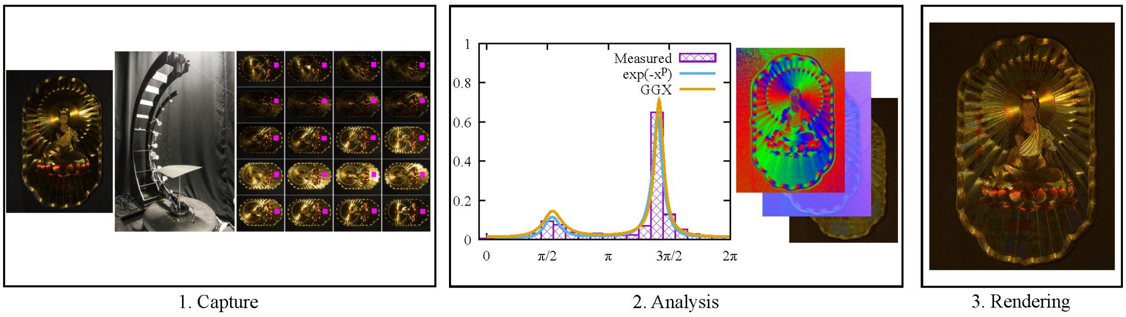


Figure 9. Our acquisition pipeline: first, we place a material sample on our acquisition platform, and acquire photographs with varying incoming light direction. In a second step, we extract anisotropic direction, shading normal, albedo and reflectance parameters from these photographs and store them in texture maps. We later use these texture maps to render new views of the material.

Many light transport related techniques require an analysis of the blur width of light scattering at a path vertex, for instance a Beckmann roughness. Such use cases are for instance analysis of expected variance (and potential biased countermeasures in production rendering), radiance caching or directionally dependent virtual point light sources, or determination of step sizes in the path space Metropolis light transport framework: recent advanced mutation strategies for Metropolis Light Transport, such as Manifold Exploration and Half Vector Space Light Transport employ local curvature of the BSDFs (such as an average Beckmann roughness) at all interactions along the path in order to determine an optimal mutation step size. A single average Beckmann roughness, however, can be a bad fit for complex measured materials and, moreover, such curvature is completely undefined for layered materials as it depends on the active scattering layer. We propose a robust estimation of local curvature for BSDFs of any complexity by using local Beckmann approximations, taking into account additional factors such as both incident and outgoing direction (See Figure 10). This work has been published as a Siggraph 2016 Talk [18].

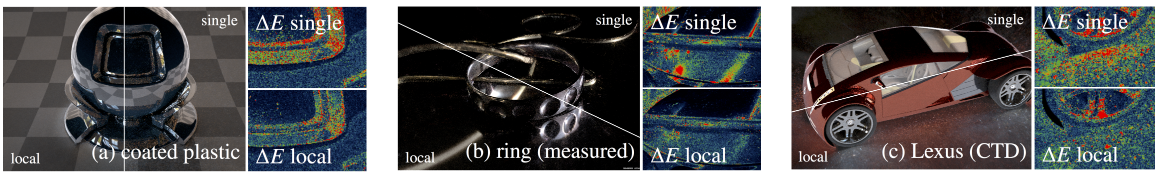


Figure 10. Indirect lighting (exposure in b and c increased for printouts) on three test scenes rendered with different materials: (a) multilayer coated plastic material, (b) measured materials on a ring, (c) CTD material on a car. The insets show difference to reference in CIE'76 ΔE. Top: single Gaussian, bottom: our local Gaussian approximation. We can render both analytic (a, c) and measured materials (b) more robustly because the local Gaussian approximation facilitates more even exploration of path space.

7.2.5. MIC based PBGI

Participant: Beibei Wang.

Point-Based Global Illumination (PBGI) is a popular rendering method in special effects and motion picture productions. The tree-cut computation is in gen eral the most time consuming part of this algorithm, but it can be formulated for efficient parallel execution, in particular regarding wide-SIMD hardware. In this context, we propose several vectorization schemes, namely single, packet and hybrid, to maximize the utilization of modern CPU architectures. Whil e for the single scheme, 16 nodes from the hierarchy are processed for a single receiver in parallel, the packet scheme handles one node for 16 receivers. These two schemes work well for scenes having smooth geometry and diffuse material. When the scene contains high frequency bumps maps and glossy reflection s, we use a hybrid vectorization method. We conduct experiments on an Intel Many Integrated Core architecture and report preliminary results on several sce nes, showing that up to a 3x speedup can be achieved when compared with non-vectorized execution [19].

7.2.6. *Point-Based Light Transport for Participating Media with Refractive Boundaries*

Participants: Beibei Wang, Jean-Dominique Gascuel, Nicolas Holzschuch.

Illumination effects in translucent materials are a combination of several physical phenomena: absorption and scattering inside the material, refraction at its surface. Because refraction can focus light deep inside the material, where it will be scattered, practical illumination simulation inside translucent materials is difficult. In this paper, we present an a Point-Based Global Illumination method for light transport on translucent materials with refractive boundaries. We start by placing volume light samples inside the translucent material and organising them into a spatial hierarchy. At rendering, we gather light from these samples for each camera ray. We compute separately the samples contributions to single, double and multiple scattering, and add them (See Figure 11). Our approach provides high-quality results, comparable to the state of the art, with significant speed-ups (from 9× to 60× depending on scene c omplexity) and a much smaller memory footprint [10], [12].

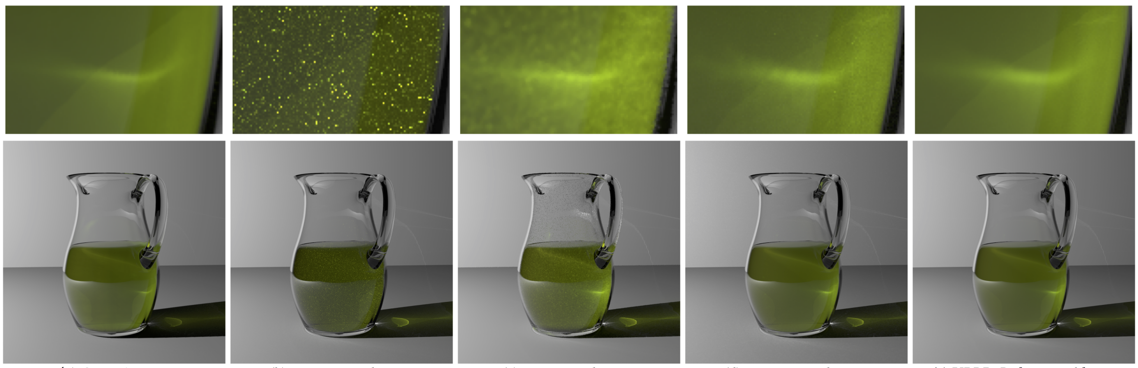


Figure 11. Our algorithm (a), compared with Bi-Directional Path Tracing (BDPT) (b), Photon Mapping with Beam-Radiance Estimate (BRE) (c) and Unified Points, Beams and Paths (UPBP) (d) (e). Our algorithm is up to 60 times faster than UPBP, with similar quality. Material: olive oil, $\alpha = 0.0042, 0.4535, 0.0995$*;* $\ell = 9.7087, 11.6279, 2.7397$*. For this material with low albedo* α *and large mean-free-path* ℓ*, low-order scattering effects dominate.*

7.3. Complex Scenes

In order to render both efficiently and accurately ultra-detailed large scenes, this approach consists in developing representations and algorithms able to account compactly for the quantitative visual appearance of a regions of space projecting on screen at the size of a pixel.

7.3.1. Appearance pre-filtering

Participants: Guillaume Loubet, Fabrice Neyret.

We address the problem of constructing appearance-preserving level of details (LoDs) of complex 3D models such as trees and propose a hybrid method that combines the strength of mesh and volume representations. Our main idea is to separate macroscopic (i.e. larger than the target spatial resolution) and microscopic (sub-resolution) surfaces at each scale and to treat them differently, because meshes are very efficient at representing macroscopic surfaces while sub-resolution geometry benefit from volumetric approximations. We introduce a new algorithm based on mesh analysis that detects the macroscopic surfaces of a 3D model at a given resolution. We simplify these surfaces with edge collapses and provide a method for pre-filtering their BRDFs parameters. To approximate microscopic details, we use a heterogeneous microflake participating medium and provide a new artifact-free voxelization algorithm that preserves local occlusion. Thanks to our macroscopic surface analysis, our algorithm is fully automatic and can generate seamless LoDs at arbitrarily coarse resolutions for a wide range of 3D models. We validated our method on highly complex geometry and show that appearance is consistent across scales while memory usage and loading times are drasticall y reduced (see Figure 12). This work has been submitted to EG2017.

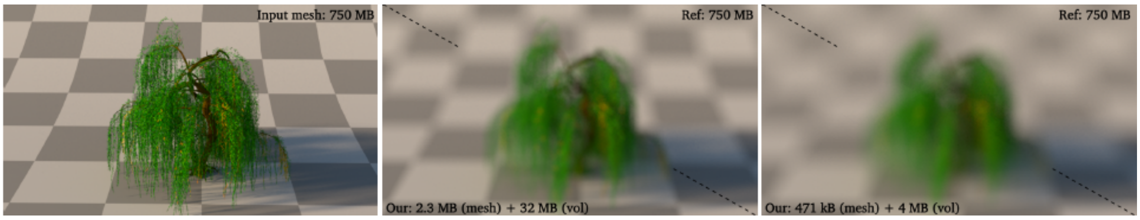


Figure 12. A weeping willow 3D model pre-filtered with our method. Our LoDs use meshes for representing macroscopic surfaces and a volumetric representation to approximate sub-resolution geometry. This approach allows for accurate preservation of the appearance of complex geometry acro ss scales while memory usage is drastic reduced. These images have been rendered with 256spp and a thin lense camera model in Mitsuba

7.4. Texture Synthesis

7.4.1. Understanding and controlling contrast oscillations in stochastic texture algorithms using Spectrum of Variance

Participants: Fabrice Neyret, Eric Heitz.

We identify and analyze a major issue pertaining to all power-spectrum based texture synthesis algorithms from Fourier synthesis to procedural noise algori thms like Perlin or Gabor noise, namely, the oscillation of contrast (see Figure 13). One of our key contributions is to introduce a simple yet powerf ul descriptor of signals, the Spectrum of Variance (not to be confused with the PSD), which, to our surprise, has never been leveraged before. In this new framework, several issues get easy to understand measure and control, with new handles, as we illustrate. We finally show that fixing oscillation of contra st opens many doors to a more controllable authoring of stochastic texturing. We explore some of the new reachable possibilities such as constrained noise content and bridges towards very different families of look such as cellular patterns, points-like distributions or reaction-diffusion [17].

7.5. Visualization and Geometric Design

7.5.1. Surfacing Curve Networks with Normal Control

Participant: Georges-Pierre Bonneau.

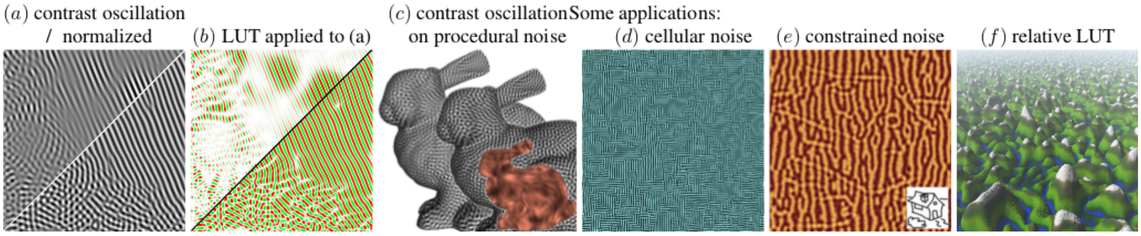


Figure 13. Power-spectrum based texturing algorithms (e.g., Gabor, Fourier synthesis) suffer from unexpected low frequency contrast variations (a,b,c top) even when the spectrum has no low frequency (the contrast field is display in red in (c)). This prevents precise authoring with non- linear transform, like color LUT (b top). Our renormalization method allows to control the stationarity (a,b,c bottom). It also opens many doors for noise authoring such as the generation of reaction-diffusion-like strips and spots (b bottom), cellular-like patterns (d), content constraints (e), or the parame trization of height maps relative to local extrema (f).

Members of Maverick involved: Georges-Pierre Bonneau

This is a joint work with team-project IMAGINE (Tibor Stanko and Stefanie Hahmann) at Inria-Grenoble and CEA-Leti (Nathalie Saguin). Recent surface acquisition technologies based on microsensors produce three-space tangential curve data which can be transformed into a network of space curves with surface normals. This work addresses the problem of surfacing an arbitrary closed 3D curve network with given surface normals. Thanks to the normal vector input, the patch finding problem can be solved unambiguously and an initial piecewise smooth triangle mesh is computed. The input normals are propagated throughout the mesh. Together with the initial mesh, the propagated normals are used to compute mean curvature vectors. We compute the final mesh as the solution of a new variational optimization method based on the mean curvature vectors. The intuition behind this original approach is to guide the standard Laplacian-based variational methods by the curvature information extracted from the input normals. The normal input increases shape fidelity and allows to achieve globally smooth and visually pleasing shapes [2], [7]. This is a joint work with team-project IMAGINE (Tibor Stanko and Stefanie Hahmann) at Inria-Grenoble and CEA-Leti (Nathalie Saguin).

7.5.2. Piecewise polynomial Reconstruction of Scalar Fields from Simplified Morse-Smale Complexes

Participants: Léo Allemand-Giorgis, Georges-Pierre Bonneau.

Morse-Smale (MS) complexes have been proposed to visualize topological features of scalar fields defined on manifold domains. Herein, three main problems have been addressed in the past: (a) efficient computation of the initial combinatorial structure connecting the critical points; (b) simplification of these combinatorial structures; (c) reconstruction of a scalar field in accordance to the simplified Morse-Smale complex. The present work faces the third problem by proposing a novel approach for computing a scalar field coherent with a given simplified MS complex that privileges the use of piecewise polynomial functions. Based on techniques borrowed from shape preserving design in Computer Aided Geometric Design, our method constructs the surface cell by cell using piecewise polynomial curves and surfaces. The benefit and limitations of using polynomials for reconstruction surfaces from topological data are studied in this work [14].

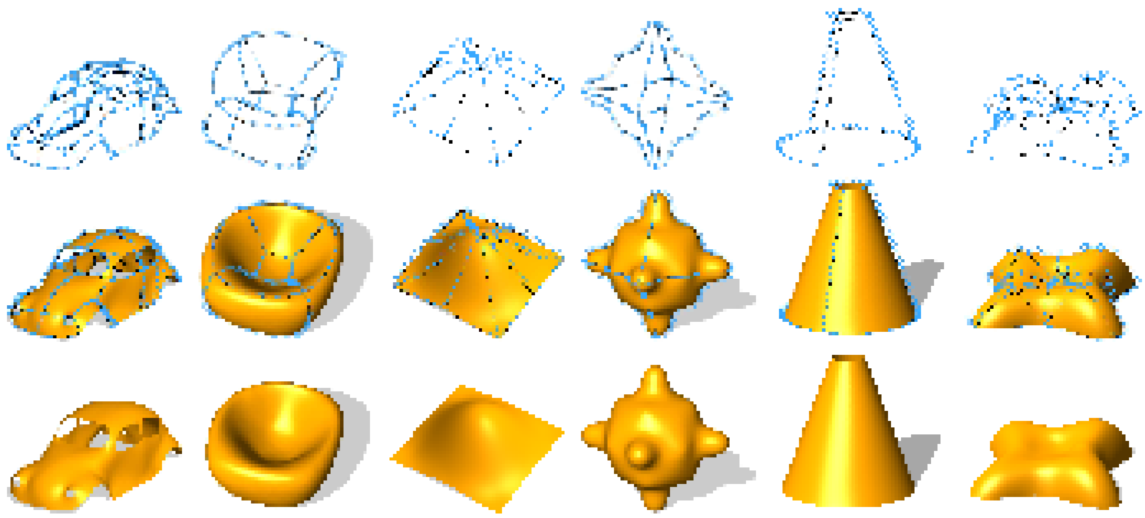


Figure 14. In [2] and [7] we address the problem of surfacing an arbitrary closed 3D curve network with given surface normals (top row). Our interpolating surfaces are visualized with (middle row) and without (bottom row) input curves.

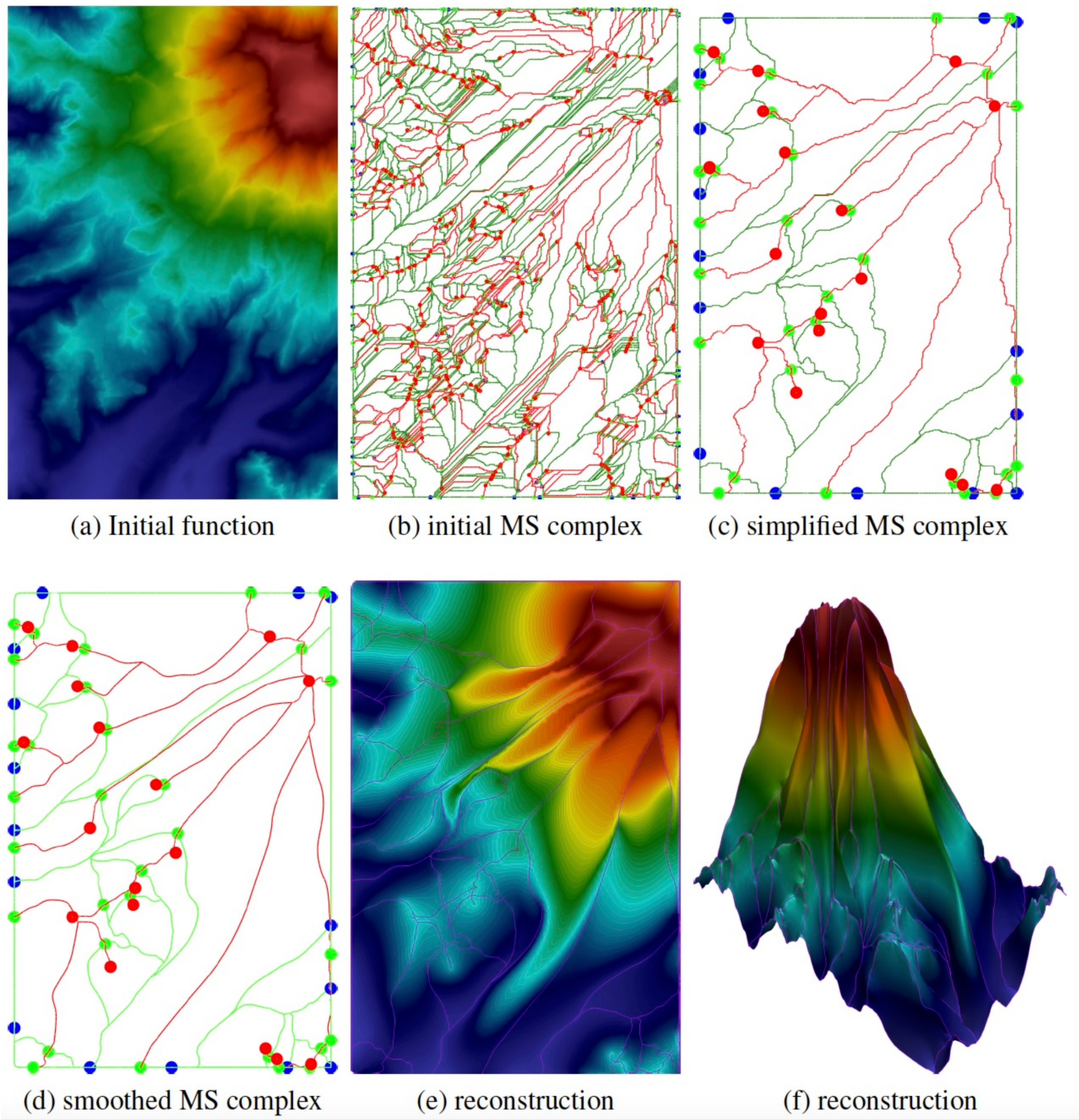


Figure 15. The terrain data set of Mt Rainier (a) has 1931 critical points (b). The simplified Morse-Smale complex with 69 critical points is reconstructed using our methods. The final function approximates the original one, with a topology that is simplified in a controlled-manner.

MIMETIC Project-Team

7. New Results

7.1. Outline

In 2016, MimeTIC pursued its efforts in improving virtual human simulation by initiating new projects in this domain, such as the Inria PRE with CAIRN team, and recruiting Antonio Mucherino in Inria half-delegation. Our main goal is to provide more natural human motion in real-time applications, which is a transversal requirement in many of MimeTIC's research domains.

- In Biomechanics, being able to rapidly simulate plausible human motion enables to explore new approaches to provide real-time feedback to users in many application domains, such a rehabilitation, sports training, ergonomics and industrial training.
- In computer graphics, simulating natural motion either relies on heavy mechanical simulation and optimal control or adapting motion capture data. We wish to push dynamic simulation a step forward to propose new biomechanically-based simulation, such as actuating the virtual human with muscles instead of rotating servos. We also wish to simplify the process of retargeting motion capture data, which is a process still difficult to automatize. In both cases, we also promote the idea of understanding how human perception behaves when facing inaccurate simulation, in order to provide accurate simulations only when necessary.
- In virtual reality, real-time motion capture and simulation are essential when using head mounted display devices as users cannot perceive their own body during immersive experiences. Hence, simulating natural avatar motion and reacting efficiently to the user's actions are key points to ensure good Presence and Embodiment. MimeTIC is collaborating with other teams in VR, such as Hybrid, to address this complex pluridisciplinary question.
- In digital storytelling, interactive autonomous virtual characters lever the potentiality of proposing complex stories on social and human themes. More stories are now created with the goal of proposing several interactive storylines, which massively enhances the possibilities of interactive entertainment, computer games and digital applications. Projects in MimeTIC explore for instance how to provide a seamless control of the balance between the autonomy of characters and the unfolding of the story through the narrative discourse.

Hence, the organization of the results is reflecting these main challenges in motion analysis, virtual human simulation, interaction in VR, and digital storytelling.

7.2. Motion Analysis

In motion analysis, we continued designing new approaches to measure human performance in specific applications, such as clinical gait assessment, ergonomics and sports. We also developed an original approach to concurrently analyze and simulate human motion, by addressing the problem of redundancy in musculoskeletal models.

7.2.1. Clinical gait assessment based on Kinect data

Participant: Franck Multon.

In clinical gait analysis, we proposed a method to overcome the main limitations imposed by the low accuracy of the Kinect measurements in real medical exams. Indeed, inaccuracies in the 3D depth images lead to badly reconstructed poses and inaccurate gait event detection. In the latter case, confusion between the foot and the ground leads to inaccuracies in the foot-strike and toe-off event detection, which are essential information to get in a clinical exam. To tackle this problem we assumed that heel strike events could be indirectly estimated by searching for the extreme values of the distance between the knee joints along the walking longitudinal axis. As Kinect sensor may not accurately locate the knee joint, we used anthropometrical data to select a body point located at a constant height where the knee should be in the reference posture. Compared to previous works using a Kinect, heel strike events and gait cycles are more accurately estimated, which could improve global clinical gait analysis frameworks with such a sensor. Once these events are correctly detected, it is possible to define indexes that enable the clinician to have a rapid state of the quality of the gait. We therefore proposed a new method to assess gait asymmetry based on depth images, to decrease the impact of errors in the Kinect joint tracking system. It is based on the longitudinal spatial difference between lower-limb movements during the gait cycle. The movement of artificially impaired gaits was recorded using both a Kinect placed in front of the subject and a motion capture system. The proposed longitudinal index distinguished asymmetrical gait, while other symmetry indices based on spatiotemporal gait parameters failed using such Kinect skeleton measurements. This gait asymmetry index measured with a Kinect is low cost, easy to use and is a promising development for clinical gait analysis.

This method has been challenged with other classical approaches to assess gait asymmetry using either cheap Kinect data or Vicon data. We demonstrate the superiority of the approach when using Kinect data for which traditional approaches failed to accurately detect gait asymmetry. It has been validated on healthy subjects who were forced to walk with a 5cm sole placed below each foot alternatively [2].

This work has been done in collaboration with the MsKLab from Imperial College London, to design new gait asymmetry indexes that could be used in daily clinical analysis.

7.2.2. New automatic methods to assess motion in industrial contexts based on Kinect

Participants: Franck Multon, Pierre Plantard.

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 use of low-cost motion capture systems (i.e., a Microsoft Kinect) to measure the 3D pose of a subject in natural environments, such as on a workstation, with many occlusions and inappropriate sensor placements. Predicting the potential accuracy of the measurement for such complex 3D poses and sensor placements is challenging with classical experimental setups. After having evaluated the actual accuracy of the pose reconstruction method delivered by the Kinect, we have identified that occlusions were a very important problem to solve in order to obtain reliable ergonomic assessments in real cluttered environments. To this end, we extended previous correction methods proposed by Hubert Shum (Northumbria University) which consist in identifying the reliable and unreliable parts of the Kinect skeleton data, and to replace unreliable ones by prior knowledge recorded in a database. In collaboration with Hubert Shum, we extended this approach to deal with long occlusions that occur in real manufacturing conditions. To this end we proposed a new data structure named Filtered Pose Graph to speed-up the process, and select example poses that improve the quality of the correction, especially ensuring continuity. We have demonstrated a significant increase of the quality of the correction, especially when large tracking errors occur with the Kinect system [16].

This method has been applied to a complete ergonomic process outputting RULA scores based on the reconstructed and corrected poses. We also demonstrated that it delivers new ergonomic information compared to traditional approaches based on isolated pictures: it provides time spent above a given RULA score which is a valuable information to support decision in ergonomics [15]. We also challenged this method with a reference motion capture system in laboratory conditions. In order to evaluate the actual use in ergonomics, we also compared the ergonomic scores obtained with this automatic method to two experts' scores in real factories. The results show very good agreements between automatic and manual assessments, and have been published in Applied Ergonomics journal [25].

This work was partially funded by the Faurecia company through a Cifre convention.

7.2.3. Evaluation and analysis of sports gestures: application to tennis serve

Participants: Richard Kulpa, Marion Morel, Benoit Bideau, Pierre Touzard.

Following the previous studies we made on tennis serve, we were able to evaluate the link between performance and risk of injuries. To go further, we made new experiments to quantify the influence of fatigue on the performance of tennis serve, that is to say the kinematic, kinetic and performance changes that occur in the serve throughout a prolonged tennis match play [12], [13]. To this end, we recorded serves of several advanced tennis players with a motion capture system before, at mid-match, and after a 3-hour tennis match. Before and after each match, we also recorded electromyographic data of 8 upper limb muscles obtained during isometric maximal voluntary contraction. These experiments showed a decrease in mean power frequency values for several upper limb muscles that is an indicator of local muscular fatigue. Decreases in serve ball speed, ball impact height, maximal angular velocities and an increase in rating of perceived exertion were also observed between beginning and end of match. However, no change in timing of maximal angular velocities was observed. The consistency in timing of maximal angular velocities suggests that advanced tennis players are able to maintain the temporal pattern of their serve technique, in spite of the muscular fatigue development [12]. Moreover, we showed that passive shoulder internal rotation and total range of motion are significantly decreased during a 3-hour tennis match that is identified as an injury risk factor among tennis players [13].

Overall, 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 novices and experts (who perform the reference motions). Only based on a database of experts' 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 joint. We applied our method on tennis serves and karate katas [22].

7.2.4. Interactions between walkers

Participants: Anne-Hélène Olivier, Armel Crétual, Julien Bruneau, Richard Kulpa, Sean Lynch, Laurentius Meerhoff, Julien Pettré.

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. This year, we developed new experiments in our immersive platform. In the context of Sean Lynch's PhD on the visual perception of human motion during interactions in locomotor tasks, we designed a study to investigate whether local limb motion is required to successfully avoid a single dynamic obstacle or if global motion alone provides sufficient information (Figures 4 .a and 4 .b). Sixteen healthy subjects were immersed in a virtual environment that required navigating towards a target, whilst an obstacle crossed its path. Within the virtual environment, four occluding walls prevented the subject observing the complete environment at the initiation of movement, ensuring steady state was reached prior to obstacle interaction. The velocity and heading of the obstacle were programmed to result in a range of future crossing distance (varying from 0.1 to 1.2m) in front and behind the subject. The velocity and heading of the obstacle were fixed, and the subject used a joystick to control its orientation to avoid collision. Five obstacle appearances were presented in a randomized order; a full body (control condition), trunk- or legs- only (i.e., local motion only), and a cylinder or sphere representing the center of gravity (COG) (i.e., global motion only). No significant difference for obstacle appearance was found on number of collisions. However, in both

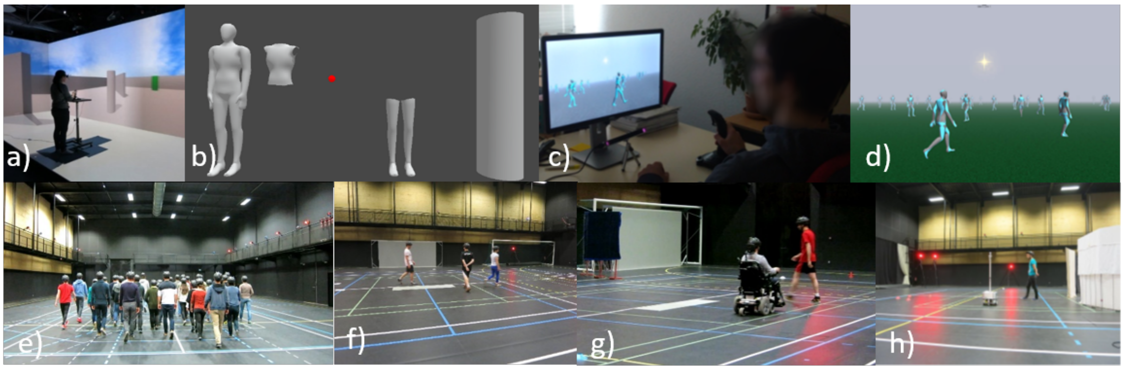


Figure 4. Experiments performed to investigate interactions between walkers.

global motion only conditions, subjects adopted alternative collision avoidance strategies compared to the full body control condition. Distance regulation and collision avoidance within daily activities may be principally regulated by global rather than local motion. Underlying mechanisms may differ accordingly to shape and size, however there is no impediment for successful completion of collision avoidance.

Second, we provide lot of efforts to investigate the complex case of multiple interactions while in previous studies we mainly focused on pairwise interactions. We developed a new experiment using an eye tracker to provide insight about the selection process of the interactions (Figures 4 .c and 4 .d). We proposed to study the human gaze during a navigation task in a crowded virtual environment. The characteristics of each virtual agent was known and controlled. Then, by recording the gaze activity, we are able to highlight the characteristics of each agent the participant was looking at. Results first showed a strong link between the fixated agents and the trajectory adaptations of the participants which means that participants looked at agents they are interacting with, which is an important result to validate the use of the eye tracker in such a situation. Concerning the characteristics of the fixated agents, results showed that human gaze, during navigation, is attracted by dangerous individuals: they were the ones presenting the higher risk of future collision with the participants. Future work is needed to evaluate the influence of other factors such as walking speed or the nature of agents trajectories. This year, we also performed an important experimental campaign including 80 participants to investigate collective behavior (Figure 4 .e). When people walk together in the street, they have to coordinate their own motion with the ones of their neighbors. From these local interactions, group motion emerges. The objective of this study was to understand how a collective behavior can emerge from these local interactions between individuals. Especially, the study aimed at identifying what is the neighborhood of a walker in a group from a perceptual point of view (who influences your motion). This work was performed in collaboration with William Warren (Brown University, Providence) and Cécile Appert-Rolland (CNRS, Orsay). Data analysis is still in process but from these results we hope to develop new knowledge on pedestrian behavior. These new results will help us to design new or improve existing crowd simulators based on local interactions. These simulators have important economic and societal roles. For example, they allow to validate the design of public places/building, which aims at hosting dense levels of public in perfectly safe conditions. The study of multiple interactions was also strengthened with the arrival of Laurentius Meerhoff as a post-doctoral student with a regional SAD funding in May 2016. Experiments involving 3 walkers were conducted (Figure 4 .f). We investigated how collision is avoided in small groups of people and whether people can successfully interact with the whole environment, or whether under some circumstance agents had to resort to sequential treatment. We proposed a method to detect whether the treatment was sequential or simultaneous and we showed the initial relative position between walkers strongly affects how interaction is engaged with.

Third, we started working on the interaction between a walker and a person on a motorized wheelchair (Figure 4 .g). This work was performed in collaboration with the Inria Lagadic team. The main objective was to design a control law that allows the wheelchair to automatically navigate in a crowded place without any collision. This is important for people who have difficulties to drive their wheelchair because of cognitive impairments. However, before reaching this objective, some steps are required to understand how walkers and persons on a wheelchair interact together. To this end, we developed a study where we recorded the trajectory of walkers and a person on a wheelchair in a collision avoidance and reaching scenario. Results will help to model such a control law for natural interactions.

Finally, we continue working on the interaction between a walker and a moving robot. This work was performed in collaboration with Philippe Souères and Christian Vassallo (LAAS, Toulouse). 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. Our work is a first step in this direction. Our goal is to describe how, during locomotion, humans avoid collision with a moving robot. We just published in Gait and Posture our results on collision avoidance between participants and a non-reactive robot (we wanted to avoid the effect of a complex loop by a robot reacting to participants' motion). Our objective was to determine whether the main characteristics of such interaction preserve the ones previously observed: accurate estimation of collision risk, anticipated and efficient adaptations. We observed that collision avoidance between a human and a robot has similarities with human-human interactions (estimation of collision risk, anticipation) but also leads to major differences [18]. Humans preferentially give way to the robot, even if this choice is not optimal with regard to motion adaptation to avoid the collision. In this new study, we considered the situation where the robot was reactive to the walker's motion (Figure 4 .h). First of all, it results that humans have a good understanding of the robot behavior and their reaction are smoother and faster with respect to the case with a non-collaborative robot. Second, humans adapt similarly to human-human study and the crossing order is respected in almost all cases. These results have strong similarities with the ones observed with two humans crossing each other.

7.2.5. Biomechanics for motion analysis-synthesis

Participants: Charles Pontonnier, Georges Dumont, Ana Lucia Cruz Ruiz, Antoine Muller, Diane Haering.

In the context of Ana Lucia Cruz Ruiz's PhD, whose goal is to define and evaluate muscle-based controllers for avatar animation, we developed an original control approach to reduce the redundancy of the musculoskeletal system for motion synthesis, based on the muscle synergy theory. For this purpose we ran an experimental campaign of overhead throwing motions. We recorded the muscle activity of 10 muscles of the arm and the motion of the subjects. Thanks to a synergy extraction algorithm, we extracted a reduced set of activation signals corresponding to the so called muscle synergies and used them as an input in a forward dynamics pipeline. Thanks to a two stage optimization method, we adapted the model's muscle parameters and the synergy signals to be as close as possible of the recorded motion. The results are compelling and ask for further developments [5]. We also proposed a classification about muscle-based controllers for animation that has been published in Computer Graphics Forum [6]. Ana Lucia defended her thesis on December 2nd, 2016.

We are also developing an analysis pipeline thanks to the work of Antoine Muller. This pipeline aims at using a modular and multiscale description of the human body to let users be able to analyse human motion. For now, the pipeline is able to assemble different biomechanical models in a convenient descriptive graph, to calibrate these models thanks to experimental data and to compute inverse dynamics to get joint torques from experimental motion capture data. We also investigated the capacity of motion-based methods to calibrate body segment inertial parameters in characterizing the part of the residuals due to the kinematical error into the dynamical one [23].

We also begin to work on the determination of maximal torque envelopes of the elbow thanks to the work of Diane Haering, Inria Post-doctoral fellow at MimeTIC. These results have a great potential of application i) to quantify the articular load during work tasks and ii) to help calibrating muscle parameters for musculoskeletal simulations. Preliminary results have been presented to an international biomechanics conference [21].

Finally, in collaboration with the CERAH (Centre d'étude et d'appareillage des handicapés, institut des invalides, Créteil, France), we proposed an identification-based method for knee prosthesis characteristics. The method is based on a forward dynamics framework enabling a matching between experimental data and model behavior [26].

7.3. Virtual Human Simulation

In addition to this last contribution on biomechanically-inspired character simulation, at the crossroad between motion analysis and simulation, we also explored two main directions for virtual human simulation in 2016. Firstly, with the arrival of Antonio Mucherino in the team, we pushed the idea of extending the idea of interaction meshes (introduced in 2010 by Taku Komura in Edinburgh) to model the constraints intrinsically associated with the motion. This approach requires developing new distance geometry algorithms in order to take time and rigid body constraints into account. Secondly, we continued to push the idea of using perceptual studies to efficiently adapt simulation in order to save computation time for less important details.

Julien Pettré moved to the Lagadic Inria team in March 2016. However we continue collaborating with him on crowd simulation problems, e.g., developing models related to interactions between pedestrians and designing perceptual studies to improve the realism of simulations.

7.3.1. Recent advances in discretizable distance geometry

Participants: Antonio Mucherino, Ludovic Hoyet, Franck Multon.

Since September 2016, Antonio Mucherino has a half-time Inria detachment in the MimeTIC team, in order to collaborate on exploring distance geometry-based problems in representing and editing human motion. In collaboration with various French and international partners, he has been working on the different facets of the discretization of the distance geometry. In 2016, he has mainly focused on the two following points. Firstly, since the discretization assumptions require the existence of a vertex ordering on the graph G which is used for representing a problem instance, he presented a new algorithm for the automatic detection of vertex orders that are also able to optimize a given set of objectives [7]. With the aim of making its exploration more efficient, the idea is to reduce in size the search space obtained with the discretization, while keeping in its interior the entire solution set. Secondly, he has started to investigate the possibility to extend the distance geometry (and its discretization) to a wider range of applications, by studying the overlaps between two different geometrical applications, arising in two different domains [3].

More related to the integration with the work in MimeTIC, we are currently exploring applying distance geometry approaches to other applications of interest for virtual human simulations, such as human motion editing and retargeting, and crowd simulations.

7.3.2. Perception of Secondary Motions in Crowd Scenarios

Participants: Ludovic Hoyet, Anne-Hélène Olivier, Richard Kulpa, Julien Pettré.

Creating plausible virtual character animations is of importance in topics researched in MimeTIC, especially for interactive applications where balancing realism and computational load is a requisite. Recently, we investigated how to improve realism of virtual crowd animations by exploring the effects of introducing secondary shoulder motions at the animation level. Typically, a crowd engine pipeline animates numerous moving characters according to a two-step process. First, a crowd simulator generates the characters' global 2D displacement trajectories in the environment, then an animation engine transforms these global trajectories into full body motions. This two-step decomposition is interesting for computational reasons, as crowd simulators raise quadratic complexity issues by nature. For the sake of simplicity, simulation models are often limited to 2D moving circles with 3 degrees of freedom (DoF), i.e., two translations and a rotation. The complete set of internal trajectories (30 to 60 DoF per character) is then considered at the animation step only, where characters are processed independently. This two-step process avoids combining the complexity of crowd simulators with the dimensionality of character kinematic models. However, it also leads to the notion of interactions between characters to be considered only at the simulation level, and to be lost at the animation level. Body animations are therefore not influenced by the presence of neighbours, only global trajectories are.

Final animations therefore often lead to residual collisions and/or characters walking as if they were alone, showing no sign to the influence of others.

In this work, we investigated the value of adding secondary motions on the perceived visual quality of crowd animations (i.e., perceived residual collisions and animation naturalness). We focused on adding shoulder motions to characters passing at close distances, and explored this question through two perceptual experiments. To understand the effects of shoulder motions on walking interactions, we first focused on understanding how these secondary motions affect how viewers perceive local interactions between two characters. We found that shoulder motions have strong positive effects on the visual quality of two-character animations, where such animations are perceived to be significantly more natural, and residual collisions become significantly less perceptible. Then we evaluated the benefits of displaying shoulder motions in the situation of crowded scenes, where shoulder motions are diluted into much more visually complex animations, and demonstrated positive effects on the animation naturalness. This increase of visual quality is obtained at a very low computational overhead, which demonstrates the relevance of the direction explored by our work. Our general conclusion is that adding secondary motions in character interactions has a significant impact on the visual quality of crowd animations, with a very light impact on the computational cost of the whole animation pipeline. Our results advance crowd animation techniques by enhancing the simulation of complex interactions between crowd characters with simple secondary motion triggering techniques.

These results were accepted and presented in SIGGRAPH 2016, the premier and most selective computer graphics scientific event, and published in ACM Transaction on Graphics [11].

7.4. Human Motions in VR

To carry-out natural and efficient interactions with a digital world, it is firstly necessary to recognize and evaluate the action of the user. We consequently initiated a collaboration with the Intuidoc IRISA team to adapt methods previously used in 2D gesture recognition to 3D motion. With the increasing use of head mounted display devices (especially cheap devices recently spread in the large public), the problem of avatar simulation and embodiment has become an important challenge. In this context, we initiated collaborative works with Hybrid to better understand embodiment and consequently imagine the future generation of avatars. Concurrently, we continued to explore the use of such technology in various application domains where human performance is a key point, such as ergonomics.

7.4.1. Motion recognition and classification

Participants: Franck Multon, Richard Kulpa, 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 [20]. 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.

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. Avatar Embodiment in Virtual Reality

Participant: Ludovic Hoyet.

With the massive development of virtual reality products investigated by major industrial companies (Google, Facebook, HTC, Sony, etc), there is a new need for understanding what makes users immersed in virtual environments, especially regarding their relation to their virtual representation (i.e., avatar). Amongst others, an important factor is for users to feel incarnated in their avatar, which is called *virtual embodiment*. As more and more technological limitations are now being unlocked, understanding such factors become important to lever new immersive applications, e.g., in education, ergonomics or entertainment.

In collaboration with the EPI Hybrid (Ferran Argelaguet and Anatole Lécuyer), we explore the capacity of avatars to convey such a sense of “virtual embodiment”, i.e., the extent to which we accept an avatar to be our representation in the virtual environment. The question of embodiment originates from the famous Rubber Hand Illusion experiment of Botvinick and Cohen (1998). This experiment demonstrated that when participants are presented with a fake rubber hand positioned beside their real hidden hand, and that both hands are synchronously stroked by an experimenter, after some time participants consider their real hand to be positioned at the location of the fake rubber hand. Today, understanding how similar phenomena happen in virtual environments is crucial to provide a maximum immersion for users. For instance, previous work demonstrated that racial biases can be reduced when users are incarnated in virtual characters of a different race, or explored body weight perception by altering the morphology of the avatar. The innovative aspect of our contributions is that we explore this embodiment effect in terms of interactions of the user with the virtual environment.

So far, we explored how people appropriate avatars by evaluating how they accept different representations of their virtual hand in virtual environments. Using various representations ranging from simplistic to highly realistic when interacting in virtual environments [19], we demonstrated that the sense of ownership (i.e., the impression that the virtual hand is actually our own hand) is increased when displaying highly realistic hand representations, but that the sense of agency (i.e., the impression to be able to control the actions of the virtual hand) is stronger for less realistic representations. With the potential of VR to alter and control avatars in different ways, e.g., the user representation, we also explored how structural differences of the hand representation can influence embodiment through controlling a six-digit virtual hand [10]. We found that participants responded positively to the possibility of controlling the virtual hand despite the structural difference, and accepted it as their own to some extent. Overall, results from such experiments further our understanding of the capacity of avatars to elicit a sense of embodiment in the users, and help to design more immersive VR experiences.

7.4.3. VR and Ergonomics

Participants: Charles Pontonnier, Georges Dumont, Pierre Plantard, Franck Multon.

The use of virtual reality tools for ergonomics applications is a very important challenge in order to generalize the use of such devices for the design of workstations.

We proposed a framework for collaborative ergonomic design in virtual environments. The framework consists in defining design modes and metaphors that help the users (engineers, ergonomists, end-users) to find a good trade-off between their own design constraints that can be contradictory at some point. We evaluated the framework and concluded that the active user has to be carefully chosen with regard to the design specifications, since the active user is favouring systematically its own constraints. This work has been published in the Journal on Multimodal User Interfaces [14].

7.5. Digital Storytelling

A transversal research of MimeTIC is digital storytelling as it enables to analyse, capture, model and simulate scenarios involving several humans (real and/or virtual). In this context, it is important to propose annotation tools and languages being able to capture such scenarios and stylistic informations before being able to simulate new ones. Moreover, when living an immersive experience in VR the user may want to have a summarize of his experience, which goes beyond simply replaying the recorded motions. Narration techniques can be positively used to highlight key events and actions, with nonlinear storytelling and intelligent camera placement to convey the desired emotion. The research in this field in MimeTIC contributes to the creation of complex stories on social and human themes. Such approaches are more and more required to create interactive storylines, which massively enhances the possibilities of interactive entertainment, training, computer games and digital applications.

7.5.1. Trip Synopsis: virtual camera control applied to route visualisation

Participant: Marc Christie.

Computerized route planning tools are widely used today by travelers all around the globe, while 3D terrain and urban models are becoming increasingly elaborate and abundant. This makes it feasible to generate a virtual 3D flyby along a planned route. Such a flyby may be useful, either as a preview of the trip, or as an after-the-fact visual summary. However, a naively generated preview is likely to contain many boring portions, while skipping too quickly over areas worthy of attention. We have therefore proposed a general interest-driven framework that automatically computes a flyby along a planned route [9]. This flyby relies on an interest function to derive how close and how slow the camera should focus on the interesting areas, while skipping interest-less regions by using elevated smoothed camera motions. To address the problem, we devised a specific iterative solving process that incrementally approaches the optimal camera trajectory by adjusting position and speed.

7.5.2. Flashbacks in narratives

Participants: Marc Christie, Hui-Yin Wu.

The flashback is a well-known storytelling device used to invoke surprise, suspense, or fill in missing details in a story. Film literature provides a deeper and more complex grounding of flashbacks by explaining their role to stimulate the viewer's memory in order to guide and change viewer comprehension. Yet, in adapting flashback mechanisms to AI storytelling systems, existing approaches have not fully modelled the roles of a flashback event on the viewer's comprehension and memory. To expand the scope of AI generated stories, we propose a formal definition of flashbacks based on the identification of four different impacts on the viewer's beliefs. We then establish a cognitive model that can predict how viewers would perceive a flashback event. We finally design a user evaluation to demonstrate that our model correctly predicts the effects of different flashbacks. This opens great opportunities for creating compelling and temporally complex interactive narratives grounded on cognitive models [29].

7.5.3. Embedded Cinematography Patterns for film Analysis

Participants: Marc Christie, Hui-Yin Wu.

Cinematography carries messages on the plot, emotion, or more general feeling of the film. Yet cinematographic devices are often overlooked in existing approaches to film analysis. To solve this limitation, we present Embedded Constrained Patterns (ECPs), a dedicated query language to search annotated film clips for sequences that fulfill complex stylistic constraints [28]. ECPs are groups of framing and sequencing constraints defined using vocabulary in film textbooks. Using a set algorithm, all occurrences of the ECPs can be found in annotated film sequences. We use a film clip from the Lord of the Rings to demonstrate a range of ECPs that can be detected, and analyse them in relation to story and emotions in the film.

MINT Project-Team

6. New Results

6.1. ControllAR: Appropriation of visual Feedback on Control Surfaces

Florent Berthaut, Alex Jones

Despite the development of touchscreens, many expert systems for working with digital multimedia content, such as in music composition and performance, video editing or visual performance, still rely on control surfaces. This can be due to the accuracy and appropriateness of their sensors, the haptic feedback that they offer, and most importantly the way they can be adapted to the specific subset of gestures and tasks that users need to perform. On the other hand, visual feedback on controllers remains limited and/or fixed, preventing similar personalizing. In this paper, we propose ControllAR, a novel system that facilitates the appropriation of rich visual feedback on control surfaces through remixing of graphical user interfaces and augmented reality display. We then use our system to study current and potential appropriation of visual feedback in the case of digital musical instruments and derive guidelines for designers and developers.

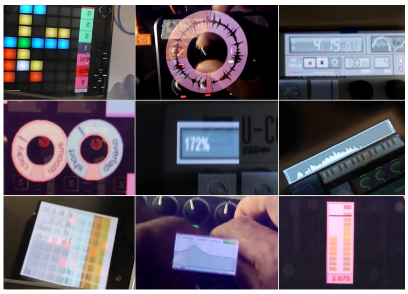


Figure 1. ControllAR: (left) ControllAR is used to augment a control surface with the remixed graphical user interface of music software, (right) Visual feedback designed by electronic musicians during our study belong to three categories: mappings feedback, processes feedback and content feedback.

6.2. Talaria: Continuous Drag & Drop on a Wall Display

Hanaë Rateau, Yosra Rekik, Laurent Grisoni, Joaquim Jorge

We present an interaction technique combining tactile actions and Midair pointing to access out-of-reach content on large displays without the need to walk across the display. Users can start through a Touch gesture on the display surface and finish Midair by pointing to push content away or inversely to retrieve a content. The technique takes advantage of wellknown semantics of pointing in human-to-human interaction. These, coupled with the semantics of proximal relations and deictic proxemics make the proposed technique very powerful as it leverages on well-understood human-human interaction modalities. Experimental results show this technique to outperform direct tactile interaction on dragging tasks. From our experience we derive four guidelines for interaction with large-scale displays.

6.3. Multi fngers interaction on a surface haptic display

Sofiane Ghenna, Christophe Giraud-Audine, Michel Amberg, Frédéric Giraud, Betty Lemaire-Semail

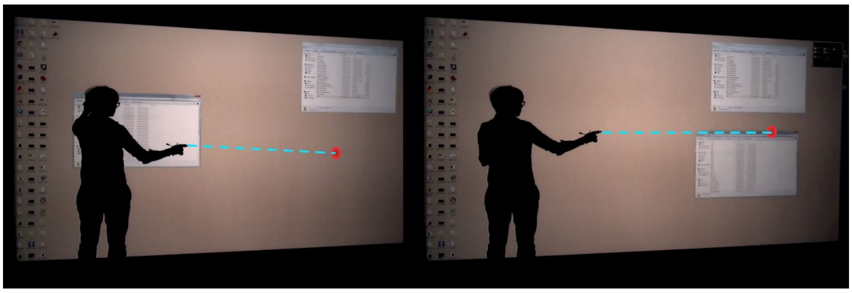
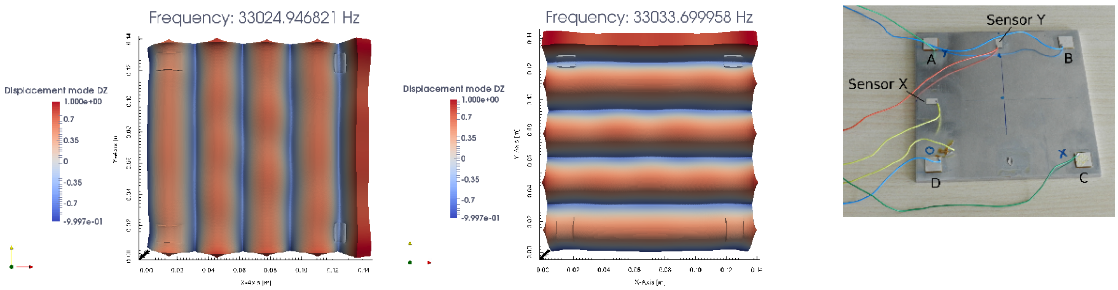


Figure 2. Talaria

In this study, we develop and implement a method for superimposing two vibration modes in order to produce different tactile stimuli on two fingers located in different positions. The tactile stimulation is based on the squeeze film effect which decreases the friction between a fingertip and a vibrating plate.

Experimental test have been conducted on a 1D tactile device. They show that it is possible to continuously control the friction on two fingers moving independently. Then, we developed the design of a 2D device based on the same principle, which gives rise to the design of a two fingers tactile display. Evaluations were conducted using a modal analysis with experimental validation.



(a) (b) (c)

Figure 3. Vibration modes and mode shapes using FEM with the position of the actuators (in white in a and b), and the prototype (c).

6.4. Finding the Minimum Perceivable Size of a Tactile Element on an Ultrasonic Based Haptic Tablet

Farzan Kalantari, Laurent Grisoni, Frédéric Giraud, Yosra Rekik

Tactile devices with ultrasonic vibrations (based on squeeze film effect) using piezoelectric actuators are one of the existing haptic feedback technologies. In this study we have performed two psychophysical experiments on an ultrasonic haptic tablet, in order to find the minimum size of a tactile element on which all the users are able to perfectly identify different types of textures. Our results show that the spatial resolution of the tactile element on haptic touchscreen actually varies, depending on the number and types of tactile feedback information. A first experiment exhibits three different tactile textures, chosen as being easily recognized by users. We use these textures in a second experiment, and evaluate minimal spatial area on which the chosen set of textures can be recognized. Among other, we find the minimal size depends on the texture nature.

6.5. BOEUF: A Unified Framework for Modeling and Designing Digital Orchestras

Florent Berthaut, Luke Dahl, Patricia Plénacoste

Orchestras of Digital Musical Instruments (DMIs) enable new musical collaboration possibilities, extending those of acoustic and electric orchestras. However the creation and development of these orchestras remain constrained. In fact, each new musical collaboration system or orchestra piece relies on a fixed number of musicians, a fixed set of instruments (often only one), and a fixed subset of possible modes of collaboration. In this paper, we describe a unified framework that enables the design of Digital Orchestras with potentially different DMIs and an expand-able set of collaboration modes. It relies on research done on analysis and classification of traditional and digital orchestras, on research in Collaborative Virtual Environments, and on interviews of musicians and composers. The BOEUF framework consists of a classification of modes of collaboration and a set of components for modelling digital orchestras. Integrating this framework into DMIs will enable advanced musical collaboration modes to be used in any digital orchestra, including spontaneous jam sessions.

Current work on this project consists in the implementation of BOEUF in the PureData programming language and in the study of its impact on musical collaboration during short improvised jam sessions.

Mjolnir Team

7. New Results

7.1. Introduction

The following sections summarize our main results of the year. For a complete list, see the list of publications at the end of this report.

7.2. Understanding and modeling users

Participants: Géry Casiez, Christian Frisson, Alix Goguey, Stéphane Huot, Sylvain Malacria, Mathieu Nancel, Thibault Raffaillac, Nicolas Roussel.

7.2.1. Touch interaction with finger identification: which finger(s) for what?

The development of robust methods to identify which finger is causing each touch point, called "finger identification," will open up a new input space where interaction designers can associate system actions to different fingers [11]. However, relatively little is known about the performance of specific fingers as single touch points or when used together in a “chord”. We presented empirical results for accuracy, throughput, and subjective preference gathered in five experiments with 48 participants exploring all 10 fingers and 7 two-finger chords. Based on these results, we developed design guidelines for reasonable target sizes for specific fingers and two-finger chords, and a relative ranking of the suitability of fingers and two-finger chords for common multi-touch tasks. Our work contributes new knowledge regarding specific finger and chord performance and can inform the design of future interaction techniques and interfaces utilizing finger identification [28].

7.2.2. Training and use of brain-computer interfaces

Brain-Computer Interfaces (BCIs) are much less reliable than other input devices, with error rates ranging from 5% up to 60%. To assess the subjective frustration, motivation, and fatigue of users when confronted to different levels of error rate, we conducted an BCI experiment in which it was artificially controlled. Our results show that a prolonged use of BCI significantly increases the perceived fatigue, and induces a drop in motivation [38]. We also found that user frustration increases with the error rate of the system but this increase does not seem critical for small differences of error rate. For future BCIs, we thus advise to favor user comfort over accuracy when the potential gain of accuracy remains small.

We have also investigated if the stimulation used for training an SSVEP-based BCI have to be similar to the one used *in fine* for interaction. We recorded 6-channels EEG data from 12 subjects in various conditions of distance between targets, and of difference in color between targets. Our analysis revealed that the stimulation configuration used for training which leads to the best classification accuracy is not always the one which is closest to the end use configuration [15]. We found that the distance between targets during training is of little influence if the end use targets are close to each other, but that training at far distance can lead to a better accuracy for far distance end use. Additionally, an interaction effect is observed between training and testing color: while training with monochrome targets leads to good performance only when the test context involves monochrome targets as well, a classifier trained on colored targets can be efficient both for colored and monochrome targets. In a nutshell, in the context of SSVEP-based BCI, training using distant targets of different colors seems to lead to the best and more robust performance in all end use contexts.

7.2.3. Evaluation metrics for touch latency compensation

Touch systems have a delay between user input and corresponding visual feedback, called input "latency" (or "lag"). Visual latency is more noticeable during continuous input actions like dragging, so methods to display feedback based on the most likely path for the next few input points have been described in research papers and patents. Designing these "next-point prediction" methods is challenging, and there have been no standard metrics to compare different approaches. We introduced metrics to quantify the probability of 7 spatial error "side-effects" caused by next-point prediction methods [35]. Types of side-effects were derived using a thematic analysis of comments gathered in a 12 participants study covering drawing, dragging, and panning tasks using 5 state-of-the-art next-point predictors. Using experiment logs of actual and predicted input points, we developed quantitative metrics that correlate positively with the frequency of perceived side-effects. These metrics enable practitioners to compare next-point predictors using only input logs.

7.2.4. Application use in the real world

Interface designers, HCI researchers or usability experts often need to collect information regarding usage of interactive systems and applications in order to interpret quantitative and behavioral aspects from users – such as our study on the use of trackpads described before – or to provide user interface guidelines. Unfortunately, most existing applications are closed to such probing methods: source code or scripting support are not always available to collect and analyze users' behaviors in real world scenarios.

InspectorWidget [26] is an open-source cross-platform application we designed to track and analyze users' behaviors in interactive software. The key benefits of this application are: 1) it works with closed applications that do not provide source code nor scripting capabilities; 2) it covers the whole pipeline of software analysis from logging input events to visual statistics through browsing and programmable annotation; 3) it allows post-recording logging; and 4) it does not require programming skills. To achieve this, InspectorWidget combines low-level event logging (e.g. mouse and keyboard events) and high-level screen capturing and interpretation features (e.g. interface widgets detection) through computer vision techniques.

7.2.5. Trackpad use in the real world

Trackpads (or *touchpads*) allow to control an on-screen cursor with finger movements on their surface. Recent models also support force sensing and multi-touch interactions, which make it possible to scroll a document by moving two fingers or to switch between virtual desktops with four fingers, for example. But despite their widespread use, little is known about how users interact with them, and which gestures they are most familiar with. To better understand this, we conducted a three-steps field study with Apple Macbook's multi-touch trackpads.

The first step of our study consisted in collecting low-level interaction data such as contact points with the trackpad and the multi-touch gestures performed while interacting. We developed a dedicated interaction logging application that we deployed on the workstation of 11 users for a duration of 14 days, and collected a total of over 82 millions contact points and almost 220 000 gestures. We then investigated finger chords (i.e., fingers used) and hand usage when interacting with a trackpad. In that purpose, we designed a dedicated mirror stand that can be easily positioned in front of the laptop's embedded web camera to divert its capturing field (Figure 1 , left). This mirror stand is combined with a background application taking photos when a multi-finger gesture is performed. We deployed this setup on the computer of 9 users for a duration of 14 days. Finally, we deployed a system preference collection application to gather the trackpad system preferences (such as transfer function and gestures associated) of 80 users. Our main findings are that touch contacts on the trackpad are performed on a limited sub-surface and are relatively slow (Figure 1 , right); that the consistency of user finger chords varies depending on the frequency of a gesture and the number of fingers involved; and that users tend to rely on the default system preferences of the trackpad [34].

7.3. Interactive visualization and animations

Participants: Amira Chalbi-Neffati, Fanny Chevalier, Nicolas Roussel.

Figure 1. Left: Mirror positioned in front of a built-in camera to divert its field of view and analyze finger-chords usages; Right: frequency distribution of all touch events of our participants, overlaid on a trackpad.

7.3.1. Social network analysis

The egocentric analysis of dynamic networks focuses on discovering the temporal patterns of a subnetwork around a specific central actor, i.e. an ego-network. These types of analyses are useful in many application domains, such as social science and business intelligence, providing insights about how the central actor interacts with the outside world. *EgoLines* is an interactive visualization we designed to support the egocentric analysis of dynamic networks . Using a "subway map" metaphor, a user can trace an individual actor over the evolution of the ego-network (Figure 2). The design of EgoLines is grounded in a set of key analytical questions pertinent to egocentric analysis, derived from interviews with three domain experts and general network analysis tasks. The results of controlled experiments with end-users and domain experts show its effectiveness in egocentric analysis tasks. Egolines can be tested at http://fannychevalier.net/egolines.html

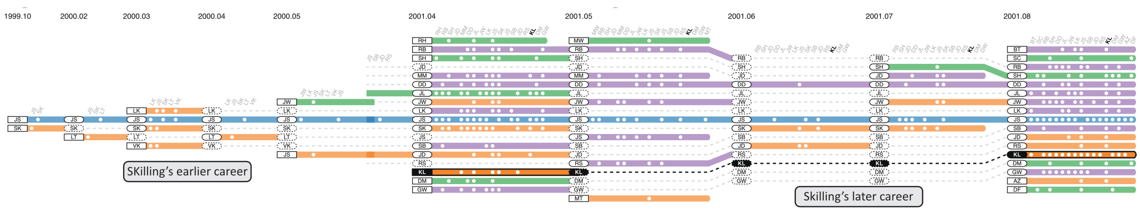


Figure 2. Egolines used to explore the dynamic network of email communications among employees at the Enron company.

7.3.2. Cross-sectional cohort phenotype

Cross-sectional phenotype studies are used by genetics researchers to better understand how phenotypes vary across patients with genetic diseases, both within and between cohorts. Analyses within cohorts identify patterns between phenotypes and patients, e.g. co-occurrence, and isolate special cases, e.g. potential outliers). Comparing the variation of phenotypes between two cohorts can help distinguish how different factors affect disease manifestation, e.g. causal genes, age of onset.). *PhenoStacks* is a novel visual analytics tool we designed to support the exploration of phenotype variation within and between cross-sectional patient cohorts . By leveraging the semantic hierarchy of the Human Phenotype Ontology, phenotypes are presented in context, can be grouped and clustered, and are summarized via overviews to support the exploration of phenotype distributions (Figure 3). The HPO is rarely used for visualization and was only recently first employed in PhenoBlocks [49]. In this prior work, we used the HPO to visualize phenotypes in clinical diagnosis settings, supporting the pairwise comparison of patient phenotypes using explicit encoding. In this new work, we turn our focus to genetics researchers conducting cross-sectional cohort studies, where the distribution of phenotypes is compared across many patients. The design of PhenoStacks was motivated by formative interviews with genetics researchers. The results of a deployment evaluation with four expert genetics researchers suggest that PhenoStacks can help identify phenotype patterns, investigate data quality issues, and inform data collection design. PhenoStacks is available from http://phenostacks.org/

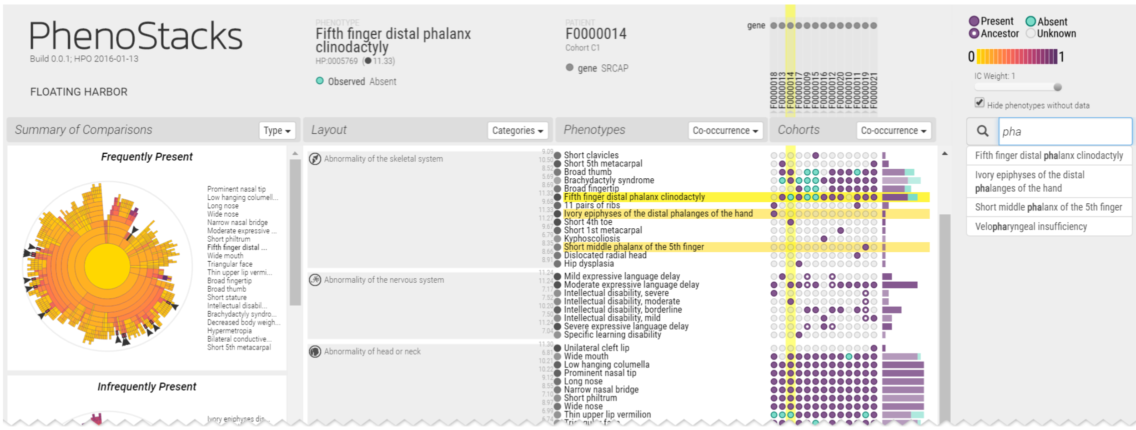


Figure 3. Exploration of phenotypic variation in cross-sectional cohorts of patients with a rare genetic disease using PhenoStacks.

7.3.3. *Human routine behavior*

Human routines are blueprints of behavior, which allow people to accomplish purposeful repetitive tasks at many levels, ranging from the structure of their day to how they drive through an intersection. People express their routines through actions that they perform in the particular situations that triggered those actions. An ability to model routines and understand the situations in which they are likely to occur could allow technology to help people improve their bad habits, inexpert behavior, and other suboptimal routines. However, existing routine models do not capture the causal relationships between situations and actions that describe routines. Byproducts of an existing activity prediction algorithm can be used to model those causal relationships in routines [23]. We applied this algorithm on two example datasets, and showed that the modeled routines are meaningful — that they are predictive of people's actions and that the modeled causal relationships provide insights about the routines that match findings from previous research. Our approach offers a generalizable solution to model and reason about routines. We show that the extracted routine patterns are at least as predictive of behaviors in the two behavior logs as the baseline we establish with existing algorithms.

To make the routine behavior models created using our approach accessible to participants and allow them to investigate the extracted routine patterns, we developed a simple visualization tool. To maintain a level of familiarity, we base our visual encoding of routine behavior elements on a traditional visual representation of an MDP as a graph (Figure 4). Our MDP graph contains nodes representing states (as circles) and actions (as squares), directed edges from state nodes to action nodes (indicating possible actions people can perform in those states), and directed edges from actions to states (indicating state transitions for any given state and action combination).

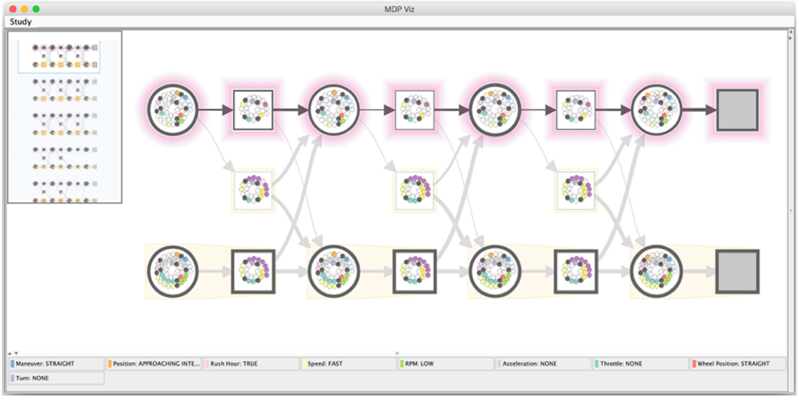


Figure 4. Our visual analytics tool showing the main routine and one likely variation of non-aggressive drivers extracted using our approach.

7.3.4. Meta-analysis of data based on user-authored annotations

User-authored annotations of data can support analysts in the activity of hypothesis generation and sense-making, where it is not only critical to document key observations, but also to communicate insights between analysts. *Annotation Graphs* are dynamic graph visualizations that enable meta-analysis of data based on user-authored annotations . The annotation graph topology encodes annotation semantics, which describe the content of and relations between data selections, comments, and tags. We present a mixed-initiative approach to graph layout that integrates an analyst's manual manipulations with an automatic method based on similarity inferred from the annotation semantics. Annotation graphs are implemented within a system, C8, that supports authoring annotations during exploratory analysis of a dataset (Figure 5). In this work, we develop and evaluate the system through an iterative user-centered design process with three experts, situated in the domain of analyzing HCI experiment data. The results suggest that annotation graphs are effective as a method of visually extending user-authored annotations to data meta-analysis for discovery and organization of ideas.

7.3.5. Fundamentals of animated transitions

Animations are increasingly used in interactive systems in order to enhance the usability and aesthetics of user interfaces. While animations are proven to be useful in many cases, we still find defective ones causing many problems, such as distracting users from their main task or making data exploration slower. The fact that such animations still exist proves that animations are not yet very well understood as a cognitive aid, and that we have not yet definitely decided what makes a well designed one. Our work on this topic aims at better understanding the different aspects of animations for user interfaces and exploring new methods and guidelines for designing them.

From bouncing icons that catch attention, to transitions helping with orientation, to tutorials, animations can serve numerous purposes. In , we revisit Baecker and Small's pioneering work *Animation at the Interface*, 25 years later. We review academic publications and commercial systems, and interviewed 20 professionals of various backgrounds. Our insights led to an expanded set of roles played by animation in interfaces today for keeping in context, teaching, improving user experience, data encoding and visual discourse. We illustrate each role with examples from practice and research, discussed evaluation methods and point to opportunities for future research. This expanded description of roles aims at inspiring the HCI research community to find novel uses of animation, guide them towards evaluation and spark further research.

We have also studied different aspects of animations for visual analysis tasks. We have worked on the design of a new model for animated transitions, explored certain aspects of visual grouping for these transitions, and

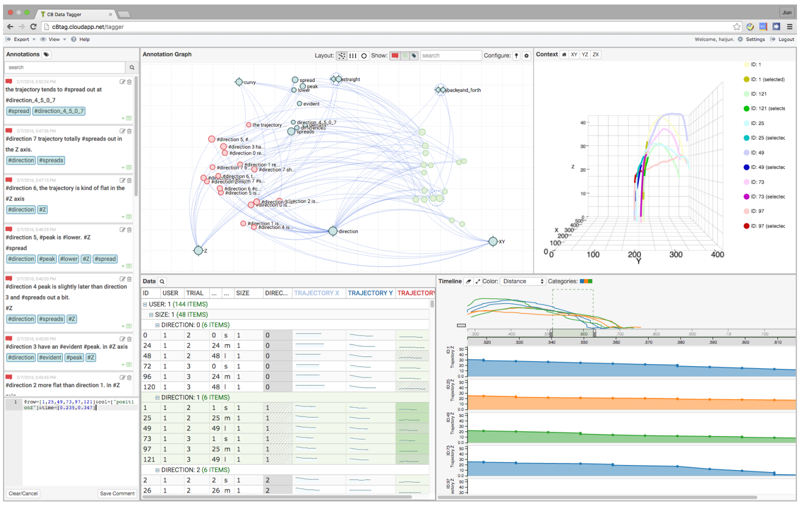


Figure 5. C8 used to analyze the results of an HCI user study that records participants pointing at a target on a tabletop display with different experimental conditions.

studied the impact of their temporal structure on data interpretation. These works, while still in progress, have been presented at the IHM doctoral consortium [39].

7.4. Interaction techniques

Participants: Géry Casiez, Fanny Chevalier, Stéphane Huot, Sylvain Malacria, Justin Mathew, Thomas Pietrzak, Nicolas Roussel.

7.4.1. Interaction in 3D environments

In virtual environments, interacting directly with our hands and fingers greatly contributes to the sense of immersion, especially when force feedback is provided for simulating the touch of virtual objects. Yet, common haptic interfaces are unfit for multi-finger manipulation and only costly and cumbersome grounded exoskeletons do provide all the efforts expected from object manipulation. To make multi-finger haptic interaction more accessible, we propose to combine two affordable haptic interfaces into a bimanual setup named DesktopGlove [18]. With this approach, each hand is in charge of different components of object manipulation: one commands the global motion of a virtual hand while the other controls its fingers for grasping. In addition, each hand is subjected to forces that relate to its own degrees of freedom so that users perceive a variety of haptic effects through both of them. Our results show that (1) users are able to integrate the separated degrees of freedom of DesktopGlove to efficiently control a virtual hand in a posing task, (2) DesktopGlove shows overall better performance than a traditional data glove and is preferred by users, and (3) users considered the separated haptic feedback realistic and accurate for manipulating objects in virtual environments.

We also investigated how head movements can serve to change the viewpoint in 3D applications, especially when the viewpoint needs to be changed quickly and temporarily to disambiguate the view. We studied how to use yaw and roll head movements to perform orbital camera control, i.e., to rotate the camera around a specific point in the scene [33]. We reported on four user studies. Study 1 evaluated the useful resolution of head movements and study 2 informed about visual and physical comfort. Study 3 compared two interaction techniques, designed by taking into account the results of the two previous studies. Results show that head roll is more efficient than head yaw for orbital camera control when interacting with a screen. Finally, Study

4 compared head roll with a standard technique relying on the mouse and the keyboard. Moreover, users were allowed to use both techniques at their convenience in a second stage. Results show that users prefer and are faster (14.5%) with the head control technique.

7.4.2. Storyboard sketching for stereo 3D films and Virtual Reality stories

The resurgence of stereoscopic and Virtual Reality (VR) media has motivated filmmakers to evolve new stereo- and VR-cinematic vocabularies, as many principles for stereo 3D film and VR story are unique. Concepts like plane separation, parallax position, and depth budgets in stereo, and presence, active experience, blocking and stitching in VR are missing from early planning due to the 2D nature of existing storyboards. Motivated to foresee difficulties exclusive to stereoscopy and VR, but also to exploit the unique possibilities of these medium, the 3D and VR cinematography communities encourages filmmakers to start thinking in stereo/VR as early as possible. Yet, there are very few early stage tools to support the ideation and discussion of a stereoscopic film or a VR story. Traditional solutions for early visual development and design, in current practices, are either strictly 2D or require 3D modeling skills, producing content that is consumed passively by the creative team.

To fill the gap in the filmmakers' toolkit, we proposed *Storeoboard* [31], a system for stereo-cinematic conceptualization, via storyboard sketching directly in stereo (Figure 6); and a novel multi-device system supporting the planning of virtual reality stories. Our tools are the first of their kind, allowing filmmakers to explore, experiment and conceptualize ideas in stereo or VR early in the film pipeline, develop new stereo- and VR-cinematic constructs and foresee potential difficulties. Our solutions are the design outcome of interviews and field work with directors, stereographers, storyboard artists and VR professionals. Our core contributions are thus: 1) a principled approach to the design and development of the first stereoscopic storyboard system that allows the director and artists to explore both the stereoscopic space and concepts in real-time, addressing key HCI challenges tied to sketching in stereoscopy; and 2) a principled survey of the state of the art in cinematic VR planning to design the first multi-device system that supports a storyboard workflow for VR film. We evaluated our tools with focus group and individual user studies with storyboard artists and industry professionals. In [31], we also report on feedback from the director of a live action, feature film on which Storeoboard was deployed. Results suggest that our approaches provide the speed and functionality needed for early stage planning, and the artifacts to properly discuss steroscopic and VR films.

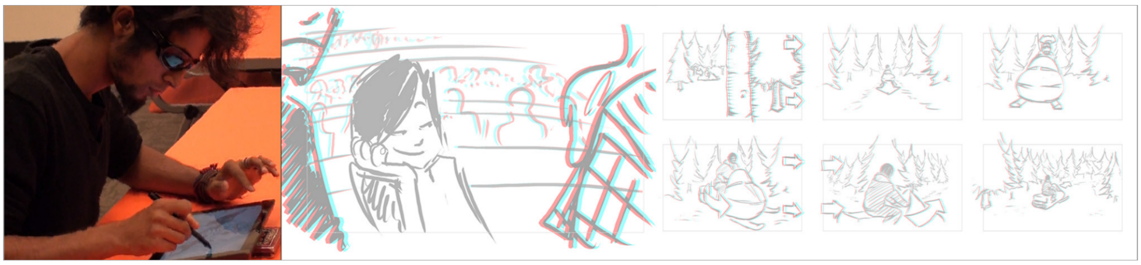


Figure 6. Storeoboard augments sketch-based storyboards with stereoscopic 3D planes for a fluid and flexible authoring of stereoscopic storyboards.

7.4.3. Tactile displays and vibrotactile feedback

Tactile displays have predominantly been used for information transfer using patterns or as assistive feedback for interactions. With recent advances in hardware for conveying increasingly rich tactile information that mirrors visual information, and the increasing viability of wearables that remain in constant contact with the skin, there is a compelling argument for exploring tactile interactions as rich as visual displays. As Direct

Manipulation underlies much of the advances in visual interactions, we introduced *Direct Manipulation-enabled Tactile display* [29]. We defined the concepts of a tactile screen, tactile pixel, tactile pointer, and tactile target which enable tactile pointing, selection and drag & drop. We built a proof of concept tactile display and studied its precision limits. We further developed a performance model for DMTs based on a tactile target acquisition study, and studied user performance in a real-world DMT menu application. The results show that users are able to use the application with relative ease and speed.

We have also explored vibrotactile feedback with wearable devices such as smartwatches and activity trackers, which are becoming prevalent. These devices provide continuous information about health and fitness, and offer personalized progress monitoring, often through multimodal feedback with embedded visual, audio, and vibrotactile displays. Vibrations are particularly useful when providing discreet feedback, without users having to look at a display or anyone else noticing, thus preserving the flow of the primary activity. Yet, current use of vibrations is limited to basic patterns, since representing more complex information with a single actuator is challenging. Moreover, it is unclear how much the user's current physical activity may interfere with their understanding of the vibrations. We addressed both issues through the design and evaluation of ActiVibe, a set of vibrotactile icons designed to represent progress through the values 1 to 10 [24]. We demonstrate a recognition rate of over 96% in a laboratory setting using a commercial smartwatch. ActiVibe was also evaluated in situ with 22 participants for a 28-day period. We show that the recognition rate is 88.7% in the wild and give a list of factors that affect the recognition, as well as provide design guidelines for communicating progress via vibrations.

7.4.4. Force-based autoscroll

Autoscroll, also known as edge-scrolling, is a common interaction technique in graphical interfaces that allows users to scroll a viewport while in dragging mode: once in dragging mode, the user moves the pointer near the viewport's edge to trigger an "automatic" scrolling. In spite of their wide use, existing autoscroll methods suffer from several limitations [45]. First, most autoscroll methods over-rely on the size of the control area, that is, the larger it is, the faster scrolling rate can be. Therefore, the level of control depends on the available distance between the viewport and the edge of the display, which can be limited. This is for example the case with small displays or when the view is maximized. Second, depending on the task, the users' intention can be ambiguous (e.g. dragging and dropping a file is ambiguous as the user's target may be located within the initial viewport or in a different one on the same display). To reduce this ambiguity, the size of the control area is drastically smaller for drag-and-drop operations which consequently also affects scrolling rate control as the user has a limited input area to control the scrolling speed.

We explored how force-sensing input, which is now available on commercial devices such as the Apple Magic Trackpad 2 or iPhone 6S, can be used to overcome the limitations of autoscroll. Indeed, force-sensing is an interesting candidate because: 1) users are often already applying a (relatively soft) force on the input device when using autoscroll and 2) varying force on the input device does not require to move the pointer, thus making it possible to offer control to the user while using a small and consistent control area regardless of the task and the device. We designed and proposed ForceEdge, a novel interaction technique mapping the force applied on a trackpad to the autoscrolling rate [19]. We implemented a software interface that can be used to design different transfer functions that map the force to autoscrolling rate and test these mappings for text selection and drag-and-drop tasks. Our pilot studies showed encouraging results and future work will focus on conducting more robust evaluations, as well as testing ForceEdge on mobile devices.

7.4.5. Combined Brain and gaze inputs for target selection

Gaze-based interfaces and Brain-Computer Interfaces (BCIs) allow for hands-free human–computer interaction. We investigated the combination of gaze and BCIs and proposed a novel selection technique for 2D target acquisition based on input fusion. This new approach combines the probabilistic models for each input, in order to better estimate the intent of the user. We evaluated its performance against the existing gaze and brain–computer interaction techniques. Twelve participants took part in our study, in which they had to search and select 2D targets with each of the evaluated techniques. Our fusion-based hybrid interaction technique was found to be more reliable than the previous gaze and BCI hybrid interaction techniques for 10 participants over 12, while being 29% faster on average. However, similarly to what has been observed in hybrid gaze-and-speech interaction, gaze-only interaction technique still provides the best performance. Our results should encourage the use of input fusion, as opposed to sequential interaction, in order to design better hybrid interfaces [14].

7.4.6. Actuated desktop devices

Desktop workstation remains the most common setup for office work tasks such as text editing, CAD, data analysis or programming. While several studies investigated how users interact with their devices (e.g. pressing keyboard keys, moving the cursor, etc.), it is not clear how they arrange their devices on the desk and whether we can leverage existing users' behaviors.

We designed the LivingDesktop [22], an augmented desktop with devices capable of moving autonomously. The LivingDesktop can control the position and orientation of the mouse, keyboard and monitors, offering different degrees of control for both the system (autonomous, semi-autonomous) and the user (manual, semi-manual) as well as different perceptive qualities (visual, haptic) thanks to a large range of device motions. We implemented a proof-of-concept of the LivingDesktop combining rail, robotic base and magnetism to control the position and orientation of the devices. This new setup presents several interesting features: (1) it improves ergonomics by continuously adjusting the position of its devices to help users adopting ergonomic postures and avoiding static postures for extended periods; (2) it facilitates collaborative works between local (e.g. located in the office) and remote co-workers; (3) it leverages context by reacting to the position of the user in the office, the presence of physical objects (e.g. tablets, food) or users' current activity in order to maintain a high level of comfort; (4) it reinforces physicality within the desktop workstation to increase immersion.

We conducted a scenario evaluation of the LivingDesktop. Our results showed the perceived usefulness of collaborative and ergonomics applications, as well as how it inspired our participants to elaborate novel application scenario, including social communication or accessibility.

7.4.7. Latency compensation

Human-computer interactions are greatly affected by the latency between the human input and the system visual response and the compensation of this latency is an important problem for the HCI community. We have developed a simple forecasting algorithm for latency compensation in indirect interaction using a mouse, based on numerical differentiation. Several differentiators were compared, including a novel algebraic version, and an optimized procedure was developed for tuning the parameters of the algorithm. The efficiency was demonstrated on real data, measured with a 1ms sampling time. These results are developed in [37] and patent has been filed on a subsequent technique for latency compensation [42].

POTIOC Project-Team

7. New Results

7.1. HOBIT

Participants: David Furio, Benoit Coulais, Martin Hachet

Practical work in optics learning allows supporting the construction of knowledge, in particular when the concept to be learned remains diffuse. To overcome the limitations of the current experimental setups, we have designed a hybrid system that combines physical interaction and numerical simulation. This system relies on 3D-printed replicas of optical elements, which are augmented with pedagogical information (see Figure 3). In a first step, we have focused on the well-known Michelson interferometer experiment, widely studied in under graduate programs of Science. A 3-months user study with 101 students and 6 teachers showed that, beyond the practical aspects offered by this system, such an approach enhances the technical and scientific learning compared to a standard Michelson interferometer experiment. A second version of HOBIT is currently being developed. This new version will let us simulate and augment multiple experiments related with optics, like polarization or Young's interferometer.

A paper presenting HOBIT has been (conditionaly) accepted at ACM CHI 2017.

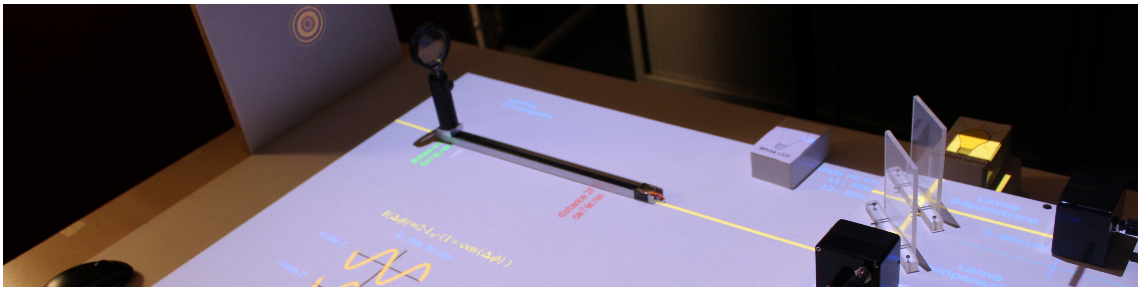


Figure 3. HOBIT: Hybrid Optical Bench for Innovative Teaching

7.2. Inner Garden

Participants: Joan Sol Roo, Renaud Gervais, Jeremy Frey, Martin Hachet

Digital technology has completely integrated our daily lives; we use it for entertainment, productivity and our social lives. However, the potential of leveraging technology to improve its users' overall happiness and life satisfaction is still largely untapped. Mindfulness, the act of paying a deliberate and non-judgmental attention to the present moment, has been shown to have a positive impact on a person's subjective well-being. With this in mind we created Inner Garden, an ambient mixed reality installation, inspired by a zen garden, comprised of an augmented sandbox along with a virtual reality modality to support mindful experiences (Figure 4 . By shaping the sand, the user creates a living miniature world that is projected back onto the sand. Moreover, using a VR headset, she can take a moment to herself by actually going inside her own garden to meditate. The natural elements of the garden are connected to real-time physiological measurements, such as breathing, helping staying focused on the body. We evaluated the system through a first user study and consulted meditation teachers, who envisioned the use of the garden in their teaching, especially for novice practitioners. The reception of the system seems to indicate that technology can, when designed carefully, both engage the users and foster well-being.

Figure 4. Inner Garden, an ambient mixed reality installation to support mindful experiences

A paper presenting Inner Garden has been (conditionaly) accepted at ACM CHI 2017.

7.3. Hybridation of Spatial Augmented Reality and Virtual Reality

Participants: Joan Sol Roo and Martin Hachet

Spatial Augmented Reality (SAR) allows a user, or a group of users, to benefit from digital augmentations embedded directly into the physical world. This enables co-located information and unobstructed interaction. On the other hand, SAR suffers from limitations that are inherently linked to its physical dependency, which is not the case for see-through or immersive displays. In this work, we explore how to facilitate the transition from SAR to VR, and vice versa, integrating both into a unified experience (Figure 5). We developed a set of interaction techniques and obtained first feedback from informal interviews.

A technote presenting this work has been (conditionaly) accepted at IEEE 3DUI 2017.

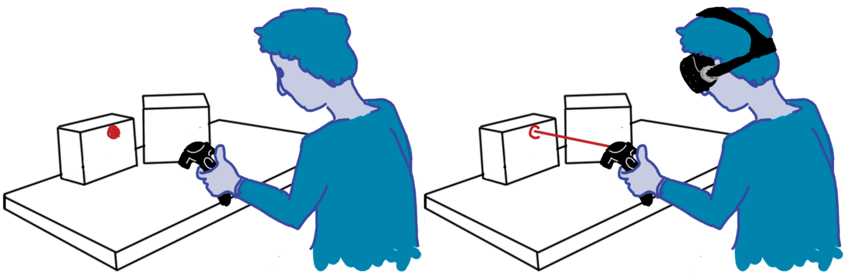


Figure 5. Transition from Spatial Augmented Reality to Virtual Reality

7.4. Augmented Reality and Tangible User Interfaces for Understanding Astronomy

Participants: Robin Gourdel, Benoit Coulais, Jeremy Laviole, Martin Hachet

We have worked with Stephanie Fleck (Université de Lorraine) to improve the leraning environment AIBLE she had imagined (see http://stefleck4.wixsite.com/aible/-propos2-cw4c). We have designed Helios, an augmented reality platform that aims at enhancing the understanding of abstract concepts in astronomy, specifically for primary schools' curriculum with children aged from 8 to 11. In order to provide physical evidence for the influence of sunlight on the Earth and the Moon, and of the consequences of their relative positions, the learning tasks are designed on inquiry-based learning principles. Children have to test their own hypotheses by using tangible props and a set of cards that trigger dedicated pedagogical activities (e.g. seasons and the Earth revolution around the Sun, lunation origin, Earth rotation and time measurement).

Helios basically consists of a standard laptop computer, a webcam, printable AR markers placed on tangible props and on dedicated pedagogical cards (See Figure 6). The (virtual) celestial bodies are displayed on the screen, and many visual feedback help understanding various phenomena (e.g. shadow cones, time zones, and so on). In [13], we discuss how such an approach allows learners to better understand abstract phenomena.

Figure 6. Helios: Manipulation of tangible objects and visualization of an augmented scene to learn astronomy.

7.5. Collaboration in VR

Participants: Damien Clergeaud and Pascal Guitton

The Airbus company regularly uses virtual reality for design, manufacturing and maintenance. We work with them on collaborative interaction in order to enable an efficient collaboration between operators immersed in the same virtual environment from remote locations and with heterogeneous equipment (large displays, CAVE, HMD). More precisely, we have developped tools to point and manipulate 3D objects, to remotely visualize the virtual environment, to be aware of remote manipulations and to describe tools and components trajectories (Figure 7). These tools have been validated by Airbus experts and the next step is to integrate them in their global process.

We are also working on Through-The-Lens Interaction techniques to ease the collaboration in some asymmetric tasks that requires a guide and an operator. Through-The-Lens techniques enable the guide to interact with the surroundings of the operator in order to help him in the task he has to perfrom. A paper presenting such a technique has been (conditionaly) accepted at IEEE 3DUI 2017.

7.6. Interactive 3D Environments for Immersive Musical Performances

Participants: Martin Hachet

Together with Florent Berthaut (Univeristé Lille 3), we presented a set of works that attempts to extent the frontiers of music creation as well as the experience of audiences attending digital performances. Indeed, the power of interactive 3D graphics, immersive displays, and spatial interfaces is still under-explored in domains where the main target is to enhance creativity and emotional experiences. The goal of our work is to connect sounds to interactive 3D graphics that musicians can interact with and the audience can observe [11].

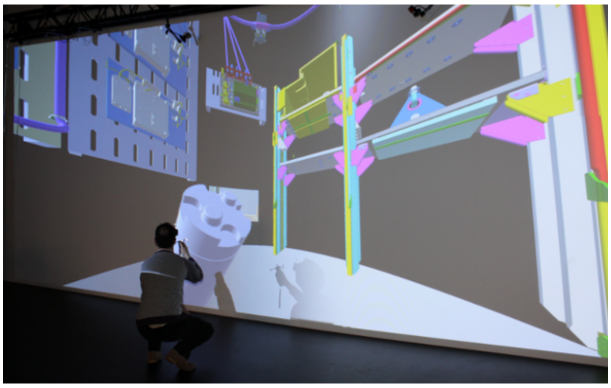


Figure 7. An immersed user has to perform a virtual task in a complex environment. In order to help the user to be fully aware of the VE, another immersed operator may guide him using a Through-The-Lens metaphor.

7.7. Multisensory Maps and 3D Printed Interactive Maps for Visually Impaired People

Participant: Anke Brock

Visually impaired people face important challenges related to orientation and mobility. Accessible geographic maps are helpful for acquiring knowledge of an urban environment. Historically, raised-line paper maps with braille text have been used, but these maps possess significant limitations. For instance, only a small percentage of the visually impaired population reads braille. Recent technological advances have enabled the design of accessible interactive maps with the aim to overcome these limitations. We designed Mappie, an accessible interactive map prototype based on the use of a multi-touch screen with a raised-line map overlay and speech output (Figure 8). Then, we deployed Mappie in a class of seven children and one caretaker during three months. Our formative study showed promising results and allowed insights in the design of accessible interactive maps, such as using olfactory and gustatory modalities to foster reflective learning, and using tangible objects to support storytelling. Following this first study, we designed MapSense as an extension of Mappie. MapSense uses the same hardware and interaction techniques as Mappie, but additionally provides fourteen "Do-It-Yourself" conductive tangibles. Some tangibles could be filled with scented material, such as olive oil, smashed raisins or honey, thus creating a real multi-sensory experience. The map was explored during two classes of three hours separated by a week, taught conjointly by a locomotion trainer and a specialized teacher. We observed that the map and tangible objects triggered strong positive emotions and stimulated spatial learning as well as creativity of the visually impaired students. This work has been conducted as part of the PhD thesis of Emeline Brulé under the supervision of Gilles Bailly and Annie Gentes, and in collaboration with the IRIT research lab in Toulouse. It has been published at CHI'16 [20].

As part of the postdoc of Stephanie Giraud at IRIT Toulouse under the supervision of Christophe Jouffrais, we have investigated how to print entire interactive maps in 3D, allowing users to construct a city like using a puzzle. We have conducted a user study comparing an interactive map that has been entirely 3D printed to a raised-line map with braille text (Figure 8 left). Our results suggest that the interactive map is significantly more effective for providing spatial knowledge than a tactile paper map with braille.

7.8. Navidrone

Participants: Julia Chatain, Anke Brock, Martin Hachet

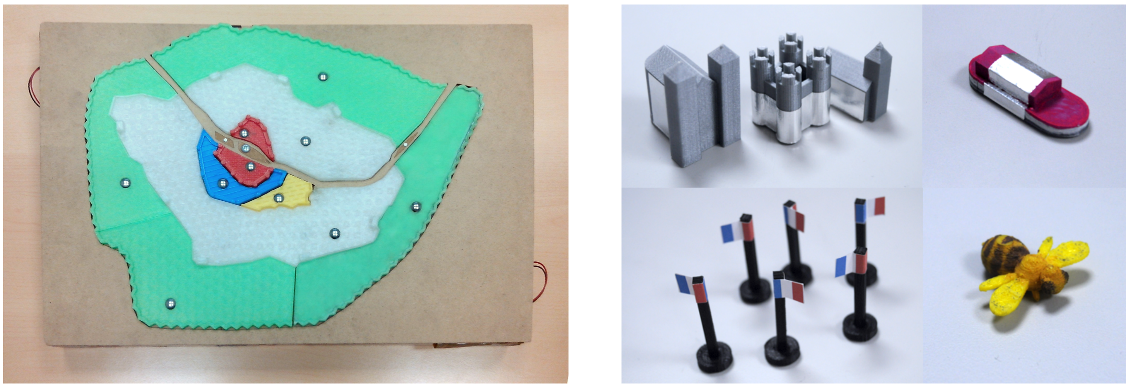


Figure 8. (Left) 3D printed interactive map, (Right) 3D printed tangibles for multisensory maps

With recent technological advances, the shapes of mobile devices are evolving. For example, we now see the emergence of new types of devices in form of autonomous aerial vehicles (drones) that become available in our everyday environment. As drones are becoming increasingly autonomous, it is crucial to understand how interaction with such devices will happen. These new devices, allow us to imagine new contexts of map usage, as for instance drone-based autonomous tour guides ((Figure 9). In order for those scenarios to happen, many problems need to be investigated. From a perspective in Human-Computer Interaction (HCI), the first questions to study are related to suitable input and output techniques. We iteratively designed interaction techniques for Navidrone, a drone-based autonomous tour guide. This work has been done in collaboration with the Prof. James Landay and Dr. Jessica Cauchard from the Stanford HCI Group.

Figure 9. Sketch showing the Navidrone concept: users interact with maps projected from drones by using their phones.

7.9. Accessibility of e-learning systems

Participants: Pierre-Antoine Cinquin and Pascal Guitton

E-learning systems, such as MOOC or serious games, are increasingly taking part in training process. Unfortunately, like most digital systems, they suffer from a lack of accessibility, in particular for people with cognitive disabilities (e.g. who have limited attention and memory). In this project, we develop a framework based on various disciplinary fields (education, cognitive sciences, human factors) but also participatory design research with students with disabilities to design interfaces promoting e-learning accessibility. From this 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. We are currently working on the analysis of this first play in order to enhance Aïana by designing new interaction modalities.

Figure 10. The Aïana MOOC player.

7.10. Teegi, a tangible EEG interface for scientific outreach

Participants : Thibault Lainé, Renaud Gervais, Jérémy Frey, Hugo Germain, Fabien Lotte, Martin Hachet

Teegi is an interactive systems that combines electroencephalographic (EEG) recordings and tangible interaction in order to let novices learn about how their brain works. By displaying EEG activity in real time on a support that is easy to manipulate and to comprehend, Teegi is a good tool for scientific outreach, that raises public interest.

While last year we developed a semi-spherical display based on LEDs, we push the project further during 2016 and built a complete autonomous puppet (Figure 11). The robot can move its two hands independently or its feet, and it can close its eyes. Beside the display of EEG activity, Teegi can react accordingly to the brain patterns recorded in real time from the user.

We demonstrated this new prototype in various occasions over the year, during public events such as “Fête de la Science” in La Cité des Sciences in Paris, a manifestation that gathered thousands of visitors (See Section 10.3 "Popularization").

7.11. Neuroergonomy

Participants : Jérémy Frey, Martin Hachet, Fabien Lotte

3D user interfaces are increasingly used in a number of applications, spanning from entertainment to industrial design. However, 3D interaction tasks are generally more complex for users since interacting with a 3D environment is more cognitively demanding than perceiving and interacting with a 2D one. As such, it is essential that we could finely evaluate user experience, in order to propose seamless interfaces. To do

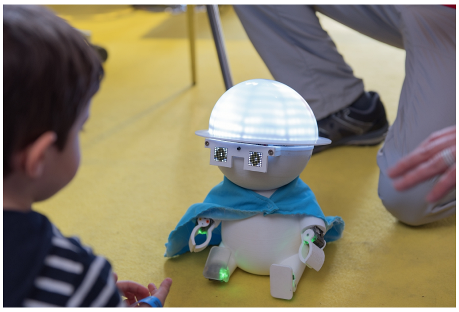


Figure 11. Teegi is a "Tangible EEG interface" that displays cerebral activity in real time by the mean of electroencephalography. A new robotized version can move and react accordingly to the brain patterns of the user, helping to explain to novices how the brain works.

so, a promising research direction is to measure users' inner states based on brain signals acquired during interaction, by following a neuroergonomics approach. Combined with existing methods, such tool can be used to strengthen the understanding of user experience.

In [15][26], we reviewed the related work in this area. We summurized what has been achieved and the new challenges that arise. We described how a mobile brain imaging technique such as electroencephalography (EEG) brings continuous and non-disruptive measures. EEG-based evaluation of users can give insights about multiple dimensions of the user experience, with realistic interaction tasks or novel interfaces. We investigate four constructs: workload, attention, error recognition and visual comfort. Ultimately, these metrics could help to alleviate users when they interact with computers.

Such advance in the understanding of the users will eventually come forward thanks to the increasing dissemination of non-invasive brain imaging devices that record electrical activity onto the scalp. In [24][23] we compared side by side two EEG amplifiers, the consumer grade OpenBCI and the medical grade g.tec g.USBamp. We suggested how an affordable and open-hardware device could facilitate, beside neuroergomomics, the appearance of various brain-computer interfaces applications.

7.12. Physiological computing

7.12.1. Physiological computing

Participants : Jérémy Frey

While physiological sensors enter the mass market and reach the general public, they are still mainly employed to monitor health. Over the course of a thesis that explored the new possibilities offered by physiological computing in terms of communication and social presence, we described several use-cases involving the externalization of inner states through novel user interfaces.

For example, we created an application that uses heart rate feedback as an incentive for social interactions. A traditional board game was "augmented" through remote physiological sensing (Figure 12), using webcams to account for the subtle changes in blood flow that occur with each heartbeat. Projection helped to conceal the technological aspects from users and merged the biofeedback with the physical environment. We detailed how players reacted – stressful situations could emerge when users are deprived from their own signals – and we gave directions for game designers to integrate physiological sensors.

We envisioned a second application, that merges virtual reality, interactive fiction and physiological computing in order to craft *truly* immersive stories; narratives that evolve depending both on the actions and on the inner states of the user/reader, stretching a medium that shaped for ages humanity (Figure 13) [32].

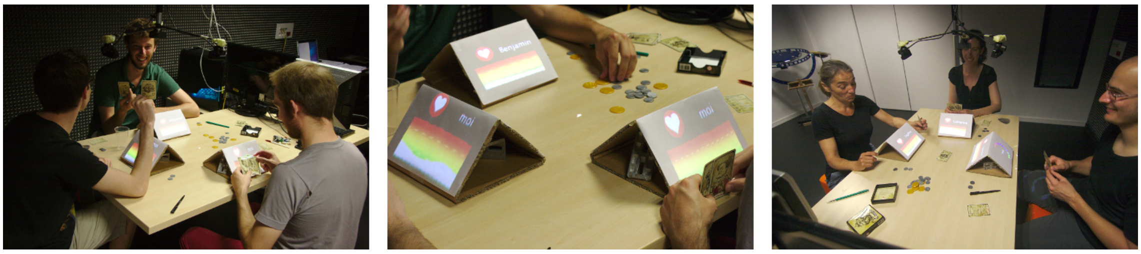


Figure 12. We augmented a traditional board game with remote physiological monitoring and projection to demonstrate how physiological computing could be used to foster new interactions between people and increase social presence.

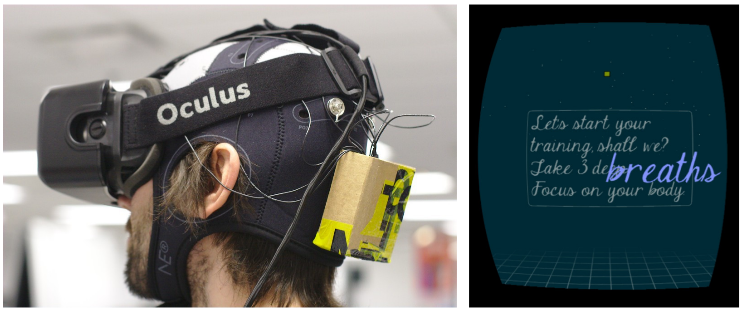


Figure 13. A combination of physiological sensors and head-mounted display (left) is used to immerse the reader in a narrative that reacts to gaze and to bodily activity (right).

7.13. EEG signal classification for BCI based on Riemannian geometry

Participants : Fabien Lotte

Although promising from numerous applications, current Brain-Computer Interfaces (BCIs) still suffer from a number of limitations. In particular, they are sensitive to noise, outliers and the non-stationarity of ElectroEncephaloGraphic (EEG) signals, they require long calibration times and are not reliable. Thus, new approaches and tools, notably at the EEG signal processing and classification level, are necessary to address these limitations. Riemannian approaches, spearheaded by the use of covariance matrices, are such a very promising tool slowly adopted by a growing number of researchers. We proposed a review of how these approaches have been used for EEG-based BCI, in particular for feature representation and learning, classifier design and calibration time reduction. Finally, we also identified relevant challenges and promising research directions for EEG signal classification in BCIs, such as feature tracking on manifold or multi-task learning [18].

7.14. Understanding Mental Imagery-based Brain-Computer Interface user-training

Participants : Camille Jeunet, Fabien Lotte

Mental Imagery-based Brain-Computer Interface (MI-BCI) enable their users to send commands to computer by imagining mental tasks (i.e., by performing MI) that are recognized in their brain signals. This type of BCI requires user training, and this training is currently poorly understood, and we basically do not know, who can learn MI-BCI control, what is to learn and how to learn it efficiently. Moreover, we have shown that current MI-BCI training protocols were both theoretically and practically inappropriate, and that there is a lack of fundamental knowledge on BCI user training, which prevents us from designing better user training approach [12].

In order to address these points, we first proposed a review and classification of cognitive and psychological predictors of MI-BCI performance. Three categories were defined: the user-technology relationship, attention and spatial abilities. The user-technology relationship refers to personality traits and states that influence users' perception of the technology and consequently impact the way they will interact with the technology (i.e., their feeling of being in control, their self-efficacy, etc.). The attention category gathers, among others, users' attentional abilities, motivation and engagement towards the task. These elements are essential to learn, whatever the skill targeted. They are also closely related to the user-technology relationship (for instance, feeling in control will increase users' engagement towards the task, thus they will allocate more attentional resources to the task). Finally, spatial abilities are the ability to produce, manipulate and transform mental images, which is closely related to the ability to control an MI-BCI. The description of these categories and of their neurophysiological correlates enabled us to submit ideas to improve MI-BCI user-training. For instance, we explained how to reduce computer-anxiety and increase the sense of agency, notably through the use of a positively biased feedback for novice users. Also, we proposed solutions to raise and improve attention, e.g., using neurofeedback or meditation. Finally, we argued that spatial abilities could be trained to improve users' capacity to perform mental imagery and consequently, potentially improve their MI-BCI performance [17].

We also did a review of the literature of current training protocols (published as a book chapter in [41]) which suggests that these protocols are, at least theoretically, inappropriate to acquire a skill and thus that they could be one of the factors responsible for inefficient MI-BCI user-training. In particular, participants are most of the time provided with uni-modal and evaluative feedback while literature recommends multi-modal, informative and supporting feedback. Although instructive, these insights only provide theoretical considerations about the flaws associated with the feedback approaches used in MI-BCI. It was therefore necessary to *concretely* assess whether standard MI-BCI feedback is appropriate to train a skill, and to what extent the feedback impacts BCI performance and skill acquisition. In order to experimentally evaluate the extent to which such a feedback has an impact on their ability to acquire a skill, we used it to teach users to perform simple motor tasks. Results (N=53 participants) revealed that with this feedback, 17% did not manage to learn the skill. This suggests that current BCI feedback is most probably suboptimal to teach a skill. A sub-group of our participants (N=20) then took part in a motor-imagery based BCI experiment. Results showed that those who struggled during the first experiment improved in performance during the second, while the others did not. We hypothesised that these results are linked to the considerable cognitive resources required to process this feedback [16].

It should be noted that there are many connections between BCI user training, and neurofeedback training for clinical applications, both field aiming at training their users to perform self regulation of their brain activity. We have therefore shown how these two field share fundamental research questions on BCI user training, and how they can both benefit from each other [10].

7.15. Improving Mental Imagery BCI user-training & feedback

7.15.1. Spatial Ability Training Protocol

Participants : Suzy Teillet, Camille Jeunet, Fabien Lotte

The results of one of our previous studies suggested that users' MI-BCI performance correlates with their spatial abilities [34], which is consistent with the literature. This result was replicated in a second study in a pure motor-imagery based BCI [16]. We thus decided to explore the potential causal relationship between both: would an improvement of spatial abilities lead to better MI-BCI performances? To try to answer this question, we designed and implemented a spatial ability (SA) training procedure (see Figure 14). Then, we performed two user studies to validate the SA training procedure: results suggest that it efficiently improves participants' SA [29]. Consequently, we included this SA training procedure in an MI-BCI protocol. Results (N=24) showed no difference in classification accuracy between participants performing 6 MI-BCI sessions and those performing 3 SA and 3 MI-BCI sessions. Nonetheless, SA training intensity impacted users' progression, and neurophysiological analyses provided us with valuable insights into brain pattern evolution throughout the training process.

Figure 14. One item per exercise included in the Spatial Ability training: the shape on top is the target, and the participant must identify the two shapes that are identical to the target among the four below.From the left to the right are displayed the shapes, matrices, cubes, arms exercises.

7.15.2. PEANUT: Personalised Emotional Agent for Neurotechnology User-Training

Participants : Léa Pillette, Camille Jeunet, Boris Mansencal, Fabien Lotte

Mental-Imagery based Brain-Computer Interfaces (MI-BCI) are neurotechnologies enabling users to control applications using their brain activity alone. These neurotechnologies are very promising. However, existing training protocols do not enable every user to acquire the skills needed to use them. Indeed, those protocols are not consonant with psychological field recommendations. In particular, the current protocols do not provide social presence and emotional support to the user. Therefore, we designed and tested PEANUT, the first Learning Companion dedicated to the improvement of MI-BCI user-training. PEANUT has been designed throughout a combination of recommendations from the literature, the analysis of data from previous experiments and user-studies. He provides emotional support using spoken sentences, such as "'C'est avec la pratique que l'on progresse"', and facial expressions. Experiments were conducted in order to assess his influence on user's performance and experience. The first results indicate that PEANUT improves the user experience. Indeed, people who trained with PEANUT found it was easier to learn and memorize how to use the MI-BCI system and rated themselves more efficient and effective than participants who had no learning

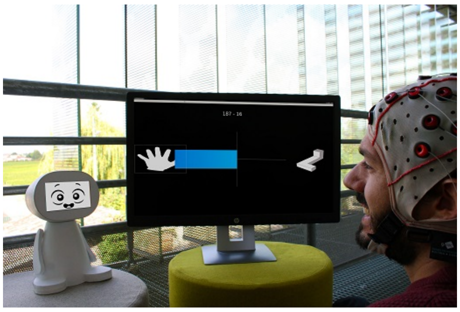


Figure 15. A participant taking part in a Brain-Computer Interface training process during which he learns to perform different mental imagery tasks (here, imagining a left-hand movement) to control the system. Along the training, PEANUT (on the left) provides the user with social presence and emotional support adapted to his performance and progression.

companion. These results indicate that using PEANUT does benefit MI-BCI user training. Future research will keep focusing on how to provide adapted cognitive and emotional feedback to MI-BCI users thanks to the use of learning companions.

7.16. Adaptive BCI training and operation

Participants : Jelena Mladenović, Jérémy Frey, Fabien Lotte

7.16.1. A generic framework for adaptive EEG-based BCI training and operation

There are two main approaches engaged in improving BCI systems: (i) improving the machine learning techniques, and the newly introduced (ii) improving human learning, by using the knowledge from instructional design and positive psychology. Both agree that the system needs to be adapted to the user but rely on different sources of adaptation: the machine for the former and the brain for the latter. In particular, machine learning algorithms should adapt to non-stationary brain signals, while human learning approaches should adapt the system to the various users' skills and profiles. Including both aspects of adaptation would give rise to a system ready to be used in real life conditions. However, a major obstacle lies in the large spectrum of sources of variability during BCI use, ranging from (i) imperfect recording conditions (e.g., environmental noise, humidity, static electricity etc. to (ii) the fluctuations in the user's psychophysiological states, due to e.g., fatigue, motivation or attention. For these reasons a BCI has not yet proved to be reliable enough to be used outside the laboratory. Particularly, it is still almost impossible to create one BCI design effective for every user, due to large inter subject variability. Therefore, the main concerns are to create a more robust system with the same high level of success for everyone, at all times, and to improve the current usability of the system. This calls for adaptive BCI training and operation.

We propose a conceptual framework which encompasses most important approaches to fit them in such a way that a reader can clearly visualize which elements can be adapted and for what reason. In the interest of having a clear review of the existing adaptive BCIs, this framework considers adaptation approaches for both the user and the machine, i.e., referring to instructional design observations as well as the usual machine learning techniques. It provides not only a coherent review of the extensive literature but also enables the reader to perceive gaps and flaws in current BCI systems, which would, hopefully, bring novel solutions for an overall improvement.

The framework (see Figure 16) has a hierarchical structure, from the lowest level elements which endure rapid changes, to the highest level elements which change at a much slower rate. It is comprised of: (i) one or several BCI systems/pipelines; (ii) a user model, whose elements are arranged according to different time scales ; (iii) a task model, enabling the system adaptation with respect to the user model; (iv) the conductor, an intelligent agent which implements the adaptive control of the whole system. A book chapter on this framework was submitted to a handbook on BCI.

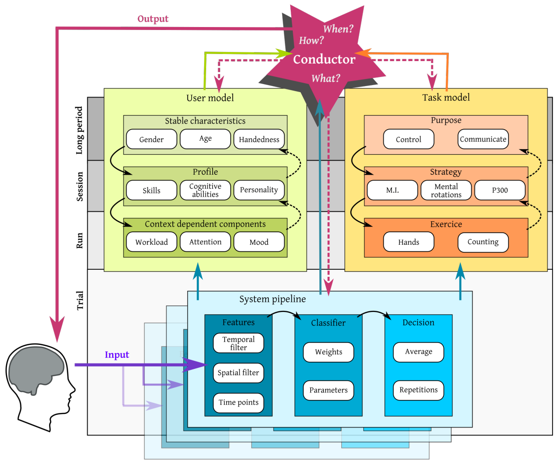


Figure 16. Multiple signals (input) maybe observed and processed in parallel in order to infer complementary states or intents, at the trial wise time scale. All the information extracted from these parallel pipelines may trigger the up-dating of the user or task model, which in turn might yield a decision from the conductor to take action, such as adapting one of the systems or the output, or modifying the task or the user model.

7.16.2. Adapting BCI Feedback based on Flow Theory

Using BCI systems can be very frustrating for people because it is not trivial and so it takes time to master. Differently from other learning procedures, BCIs do not have enough, if any explanatory feedback in assisting the learning of users. Also, as the feedback is not engaging the user's mind might easily wander off, which highly affects the system's accuracy as well as the person's learning pace. For this reason it takes more time to train a user to understand the procedure and have control over the system. Hence, we want to create an immersive and playful environment to attract the user's attention and help them learn with less effort and frustration.

We rely on the theory of Flow, introduced by Csikszentmihalyi in the 1970s. Flow is a state of enjoyment while effortlessly focused on a task so immersive that one looses the perception of time. In order to fulfil these requirements, we choose the users to be involved in an open-source video game called Tux Racer. Also, to ensure the maximal attention of the users, the game difficulty adapts according to users performance in real-time.

7.17. Brain-Computer Interfaces 1 and 2: foundations, methods, practice and applications

Participants : Jérémy Frey, Camille Jeunet, Fabien Lotte

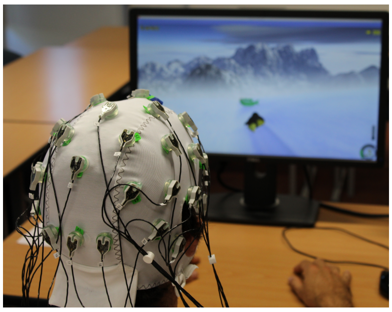


Figure 17. A subject using motor imagery, i.e., imagining left or right hand movements to manipulate Tux to catch fish.

Together with Maureen Clerc (Inria Sophia) and Laurent Bougrain (Inria Nancy), we co-edited the first book on Brain-Computer Interfaces in French [50], [51], this book being also translated into English [48], [49]. It is published in two volumes, and co-written with researchers from all over France from many disciplines related to BCI. It covers both theoretical and practical aspects, as well as the neuroscience, mathematics, psychology, computer science, engineering, and ethical aspects of BCI. It is aimed at being a key resource for anyone who wants to start BCI research or want to deepen their knowledge in the many aspects of this exciting discipline.

Figure 18. The two volumes of the French version of the BCI book we edited.

TITANE Project-Team

7. New Results

7.1. Analysis

7.1.1. Object Classification via Planar Abstraction

Participants: Sven Oesau, Florent Lafarge, Pierre Alliez.

In collaboration with EADS ASTRIUM

We contributed a supervised machine learning approach for classification of objects from sampled point data. The main idea consists in first abstracting the input object into planar parts at several scales, then discriminate between the different classes of objects solely through features derived from these planar shapes. Abstracting into planar shapes provides a means to both reduce the computational complexity and improve robustness to defects inherent to the acquisition process. Measuring statistical properties and relationships between planar shapes offers invariance to scale and orientation. A random forest is then used for solving the multiclass classification problem. We demonstrate the potential of our approach on a set of indoor objects from the Princeton shape benchmark and on objects acquired from indoor scenes and compare the performance of our method with other point-based shape descriptors [9]. This work was published in the proceedings of ISPRS.

7.1.2. Fidelity vs. Simplicity: a Global Approach to Line Drawing Vectorization

Participants: Jean-Dominique Favreau, Florent Lafarge.

In collaboration with Adrien Bousseau (GraphDeco Inria team)

Vector drawing is a popular representation in graphic design because of the precision, compactness and editability offered by parametric curves. However, prior work on line drawing vectorization focused solely on faithfully capturing input bitmaps, and largely overlooked the problem of producing a compact and editable curve network. As a result, existing algorithms tend to produce overly-complex drawings composed of many short curves and control points, especially in the presence of thick or sketchy lines that yield spurious curves at junctions. We propose the first vectorization algorithm that explicitly balances fidelity to the input bitmap with simplicity of the output, as measured by the number of curves and their degree. By casting this trade-off as a global optimization, our algorithm generates few yet accurate curves, and also disambiguates curve topology at junctions by favoring the simplest interpretations overall. We demonstrate the robustness of our algorithm on a variety of drawings, sketchy cartoons and rough design sketches (See Figure 1). This work was published at ACM SIGGRAPH 2016 [4].

7.1.3. High-Resolution Semantic Labeling with Convolutional Neural Networks

Participants: Emmanuel Maggiori, Yuliya Tarabalka, Pierre Alliez.

In collaboration with Guillaume Charpiat (Inria TAO team)

Convolutional neural networks (CNNs) were initially conceived for image categorization, i.e., the problem of assigning a semantic label to an entire input image. We have address the problem of dense semantic labeling, which consists in assigning a semantic label to *every* pixel in an image. Since this requires a high spatial accuracy to determine *where* labels are assigned, categorization CNNs, intended to be highly robust to local deformations, are not directly applicable. By adapting categorization networks, many semantic labeling CNNs have been recently proposed. Our first contribution is an in-depth analysis of these architectures. We establish the desired properties of an ideal semantic labeling CNN, and assess how those methods stand with regard to these properties. We observe that even though they provide competitive results, these CNNs often do not leverage properties of semantic labeling that could lead to more effective and efficient architectures. Out of these observations, we then derive a CNN framework specifically adapted to the semantic labeling problem [12]. In addition to learning features at different resolutions, it learns how to combine these features. By integrating local and global information in an efficient and flexible manner, it outperforms previous techniques. We evaluate the proposed framework and compare it with state-of-the-art architectures on public benchmarks of high-resolution aerial image labeling.

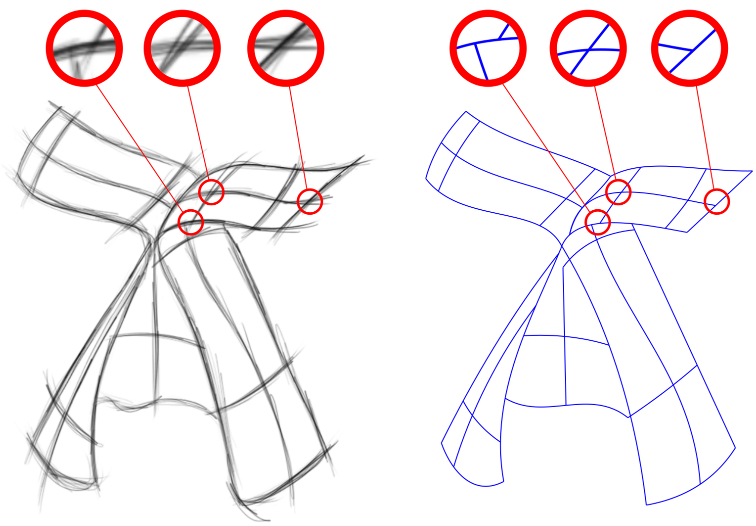


Figure 1. Line Drawing Vectorization. Rough sketches often contain overlapping strokes (left). Since existing algorithms analyze junctions locally, they cannot recover the proper topology of these seemingly similar line configurations. By adopting a global formulation that optimizes for both fidelity to the input sketch and simplicity of the output curve network, our algorithm recovers proper topology while significantly reducing the overall number of curves and control points (right). Design sketch after Sori Yanagi Butterfly stool.

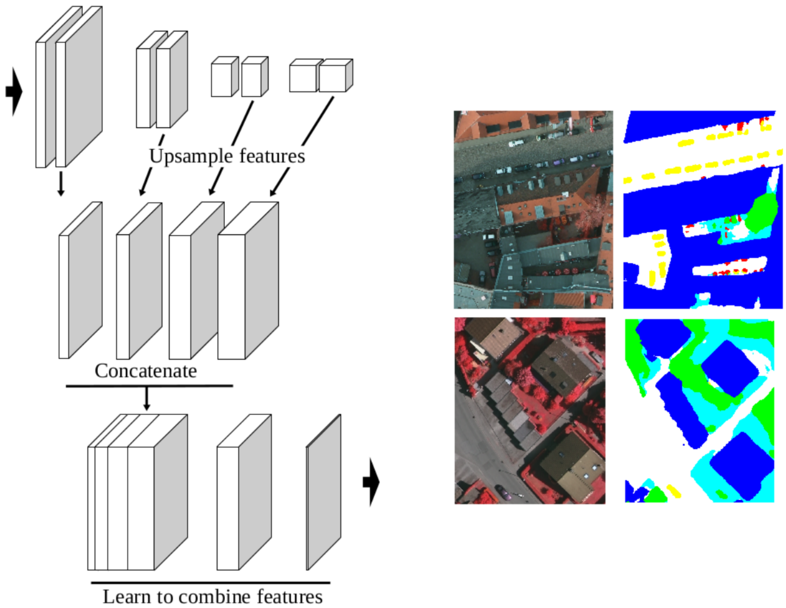


Figure 2. Our MLP network architecture (left) learns features at different resolutions and also learns how to combine those features. The technique was evaluated on the ISPRS 2D Semantic Segmentation Contest (right), providing competitive results.

7.1.4. Learning Iterative Processes with Recurrent Neural Networks to Correct Satellite Image Classification Maps

Participants: Emmanuel Maggiori, Yuliya Tarabalka, Pierre Alliez.

In collaboration with Guillaume Charpiat (Inria TAO team)

While initially devised for image categorization, convolutional neural networks (CNNs) are being increasingly used for the pixelwise semantic labeling of images. However, the proper nature of the most common CNN architectures makes them good at recognizing but poor at localizing objects precisely. This problem is magnified in the context of aerial and satellite image labeling, where a spatially fine object outlining is of paramount importance.

Different iterative enhancement algorithms have been presented in the literature to progressively improve the coarse CNN outputs, seeking to sharpen object boundaries around real image edges. However, one must carefully design, choose and tune such algorithms. Instead, our goal is to directly learn the iterative process itself. For this, we formulate a generic iterative enhancement process inspired from partial differential equations, and observe that it can be expressed as a recurrent neural network (RNN). Consequently, we train such a network from manually labeled data for our enhancement task. In a series of experiments we show that our RNN effectively learns an iterative process that significantly improves the quality of satellite image classification maps [11].

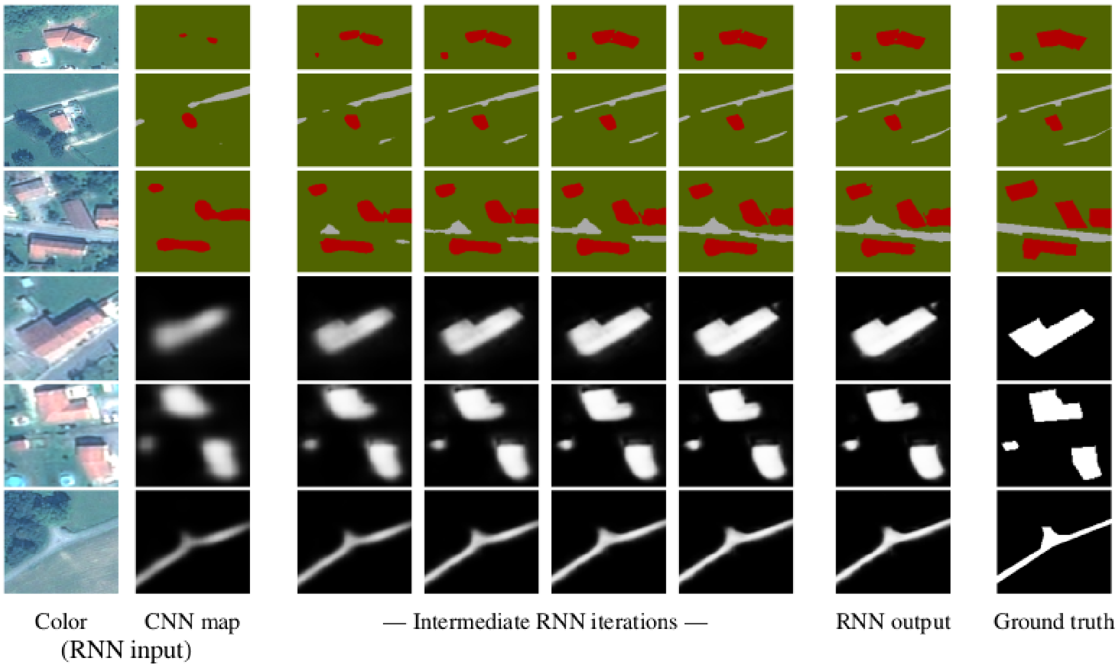


Figure 3. A recurrent neural network (RNN) learns an algorithm to iteratively correct the output of a coarse classification map. As a result, the satellite image classification maps become finer and better aligned to the real objects.

7.1.5. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification

Participants: Emmanuel Maggiori, Yuliya Tarabalka, Pierre Alliez.

In collaboration with Guillaume Charpiat (Inria TAO team)

We propose an end-to-end framework for the dense, pixelwise classification of satellite imagery with convolutional neural networks (CNNs). In our framework, CNNs are directly trained to produce classification maps out of the input images. We first devise a *fully convolutional* architecture and demonstrate its relevance to the dense classification problem. We then address the issue of imperfect training data through a two-step training approach: CNNs are first initialized by using a large amount of possibly inaccurate reference data, then refined on a small amount of accurately labeled data. To complete our framework we design a multi-scale neuron module that alleviates the common trade-off between recognition and precise localization. A series of experiments show that our networks take into account a large amount of context to provide fine-grained classification maps. This work was published in IEEE Transactions on Geoscience and Remote Sensing (TGRS) [5].

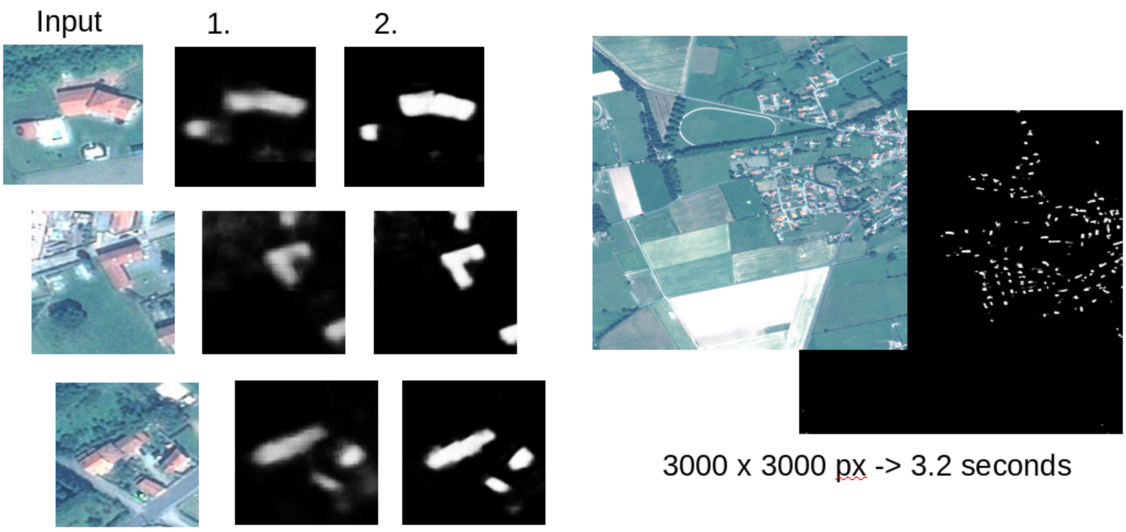


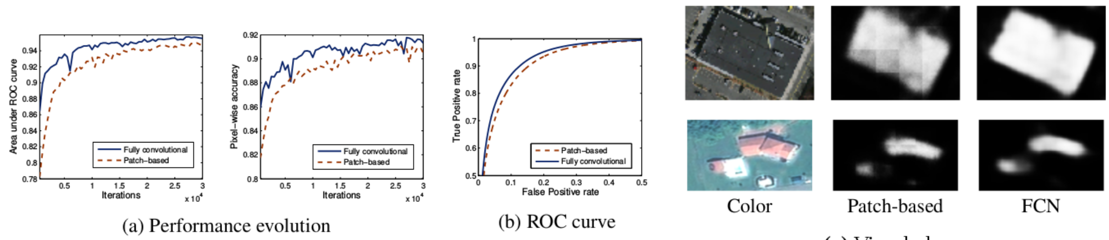
Figure 4. We train in a two-step scheme: first we train a fully convolutional network on large amounts of imperfect training data, to capture the generalities of the problem, which leads to coarse classification maps (1). In a second stage we fine-tune the network for few iterations on a precise manually labeled dataset, outputting fine classification maps as a results (2). The overall system is efficient and scalable.

7.1.6. Fully Convolutional Neural Networks for Remote Sensing Image Classification

Participants: Emmanuel Maggiori, Yuliya Tarabalka, Pierre Alliez.

In collaboration with Guillaume Charpiat (Inria TAO team)

We propose a convolutional neural network (CNN) model for remote sensing image classification, i.e. the assignment of a class to every pixel in an image. Using CNNs provides us with a means of learning contextual features for large-scale image labeling. Our network consists of four stacked convolutional layers that downsample the image and extract relevant features. On top of these, a deconvolutional layer upsamples the data back to the initial resolution, producing a final dense image labeling. Contrary to previous frameworks, our architecture is a fully convolutional network (FCN), contains only convolution and deconvolution operations and no fully connected layers as in previous work. The fact of being fully convolutional removes the artifacts present in previous work by construction and is considerably more efficient. Experiments on aerial images show that our network produces more accurate classifications in lower computational time. This work was published in the proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) [8].



(a) Performance evolution (b) ROC curve (c) Visual closeups

Figure 5. Our fully convolutional network (FCN) provides a better accuracy compared to a previous method (the "patch-based" network), as observed by the evolution of the accuracy through the training iterations (a) and the final precision/recall curve (b). We can also observe, visually, that our FCN network removes the artifacts at the border of patches (c). Besides the improved performance, the architecture drastically reduces the number of trainable parameters, being 10 times faster to run compared to the patch-based counterpart.

7.1.7. Large-scale Remote Sensing Image Segmentation and Classification

Participants: Chunlin Xiao, Emmanuel Maggiori, Yuliya Tarabalka.

In collaboration with Guillaume Charpiat (Inria TAO team)

The representation of images with binary partition trees (BPTs) has proven to be very efficient for multiscale analysis, object detection and classification of high-resolution images. We propose a new framework for multi-class image segmentation using a binary partition tree. The region model is composed of three components : color component, probability component and shape component, some of which can be used or omitted depending on the information available and the application itself. The problem to extract a segmentation is formulated as the minimization of an energy function which can be solved with dynamical programming efficiently. However, BPT represents a hierarchy of the image regions at different scales. For large-scale images such representation can be demanding in terms of both memory and computation resources. We propose a tile-based scheme to extend the framework for processing arbitrarily large images. Experiments (see Fig. 6) prove that the algorithm can segment large images efficiently while ensuring quite similar results with respect to processing the whole image at once. This work has not been published yet.

7.2. Reconstruction

7.2.1. Towards Large-scale City Reconstruction from Satellites

Participants: Liuyun Duan, Florent Lafarge.

In collaboration with Geoimage.

Automatic city modeling from satellite imagery is one of the biggest challenges in urban reconstruction. Existing methods produce at best rough and dense Digital Surface Models. Inspired by recent works on semantic 3D reconstruction and region-based stereovision, we propose a method for producing compact, semantic-aware and geometrically accurate 3D city models from stereo pair of satellite images [7]. Our approach relies on two key ingredients. First, geometry and semantics are retrieved simultaneously bringing robustness to occlusions and to low image quality. Second, we operate at the scale of geometric atomic region which allows the shape of urban objects to be well preserved, and a gain in scalability and efficiency. We demonstrate the potential of our algorithm by reconstructing different cities around the world in a few minutes (See Figure 7). This work has been published in the proceedings of the European Conference on Computer Vision (ECCV).

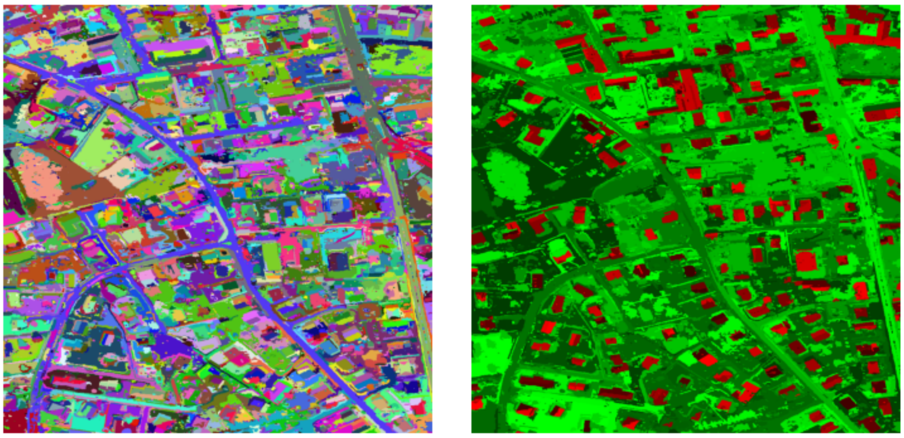


Figure 6. Results of (left) unsupervised segmentation and (right) supervised segmentation of the image into building (red) and non-building (green) regions, using 4×4 tiling scheme.

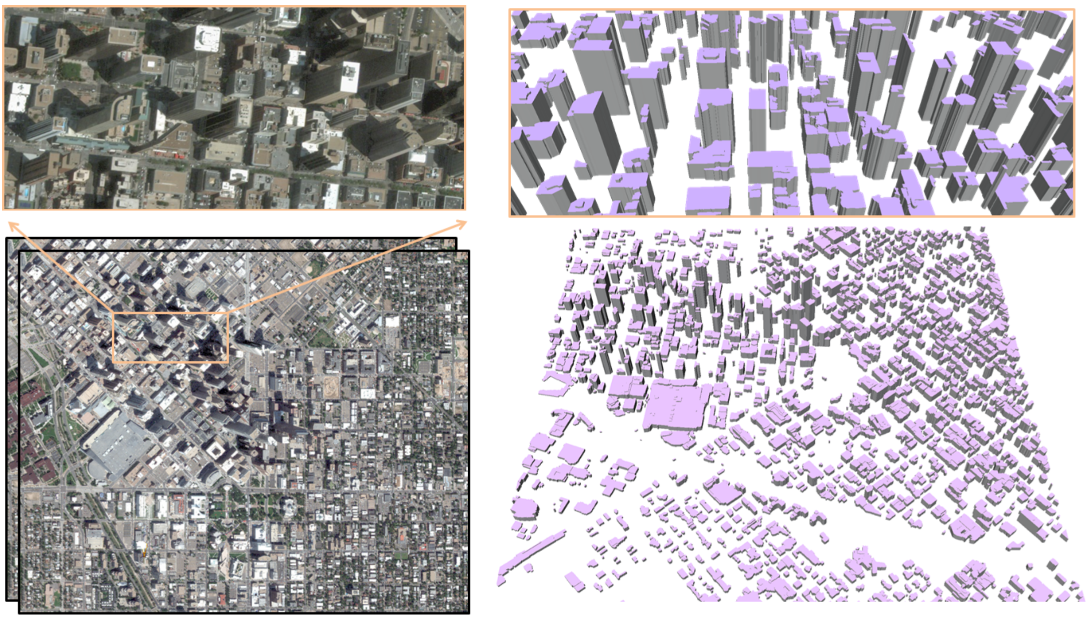


Figure 7. City reconstruction from satellites. Starting from a stereo pair of satellite images (left), our algorithm produces a compact and semantic-aware 3D model (right) in a few minutes.

7.2.2. A Survey of Surface Reconstruction from Point Clouds

Participant: Pierre Alliez.

In collaboration with Matthew Berger, Andrea Tagliasacchi, Lee Seversky, Gael Guennebaud (Inria MANAO), Joshua Levine, Andrei Sharf and Claudio Silva.

The area of surface reconstruction has seen substantial progress in the past two decades. The traditional problem addressed by surface reconstruction is to recover the digital representation of a physical shape that has been scanned, where the scanned data contains a wide variety of defects. While much of the earlier work has been focused on reconstructing a piece-wise smooth representation of the original shape, recent work has taken on more specialized priors to address significantly challenging data imperfections, where the reconstruction can take on different representations – not necessarily the explicit geometry. We survey the field of surface reconstruction, and provide a categorization with respect to priors, data imperfections, and reconstruction output. By considering a holistic view of surface reconstruction, we show a detailed characterization of the field, highlight similarities between diverse reconstruction techniques, and provide directions for future work in surface reconstruction. This survey was published in Computer Graphics Forum [2].

7.3. Approximation

7.3.1. A Line/Trimmed NURBS Surface Intersection Algorithm Using Matrix Representations

Participant: Pierre Alliez.

In collaboration with Laurent Busé from Inria AROMATH, and Jingjing Shen and Neil Dodgson from Cambridge University (UK).

We contribute a reliable line/surface intersection method for trimmed NURBS surfaces, based on a novel matrix-based implicit representation and numerical methods in linear algebra such as singular value decomposition and the computation of generalized eigenvalues and eigenvectors. A careful treatment of degenerate cases makes our approach robust to intersection points with multiple pre-images. We then apply our intersection algorithm to seamlessly mesh NURBS surfaces through Delaunay refinement (see Figure 8). We demonstrate the added value of our approach in terms of accuracy and treatment of degenerate cases, by providing comparisons with other intersection approaches as well as a variety of meshing experiments. This work was published in Computer Aided Geometric Design [6].

7.3.2. Optimal Voronoi Tessellations with Hessian-based Anisotropy

Participants: Pierre Alliez, Mathieu Desbrun.

In collaboration with Max Budninskiy and Beibei Liu from Caltech, Fernando de Goes from Pixar and Yiying Tong from Michigan State University.

We contribute a variational method to generate cell complexes with local anisotropy conforming to the Hessian of any given convex function and for any given local mesh density. Our formulation builds upon approximation theory to offer an anisotropic extension of Centroidal Voronoi Tessellations which can be seen as a dual form of Optimal Delaunay Triangulation. We thus refer to the resulting anisotropic polytopal meshes as Optimal Voronoi Tessellations. Our approach sharply contrasts with previous anisotropic versions of Voronoi diagrams as it employs first-type Bregman diagrams, a generalization of power diagrams where sites are augmented with not only a scalar-valued weight but also a vector-valued shift. As such, our OVT meshes contain only convex cells with straight edges (Figure 9), and admit an embedded dual triangulation that is combinatorially-regular. We show the effectiveness of our technique using off-the-shelf computational geometry libraries. This work was published at ACM SIGGRAPH Asia [3].

7.3.3. Symmetry and Orbit Detection via Lie-Algebra Voting

Participants: Pierre Alliez, Mathieu Desbrun.

In collaboration with Zeyun Shi, Hujun Bao and Jin Huang from Zhejiang University.

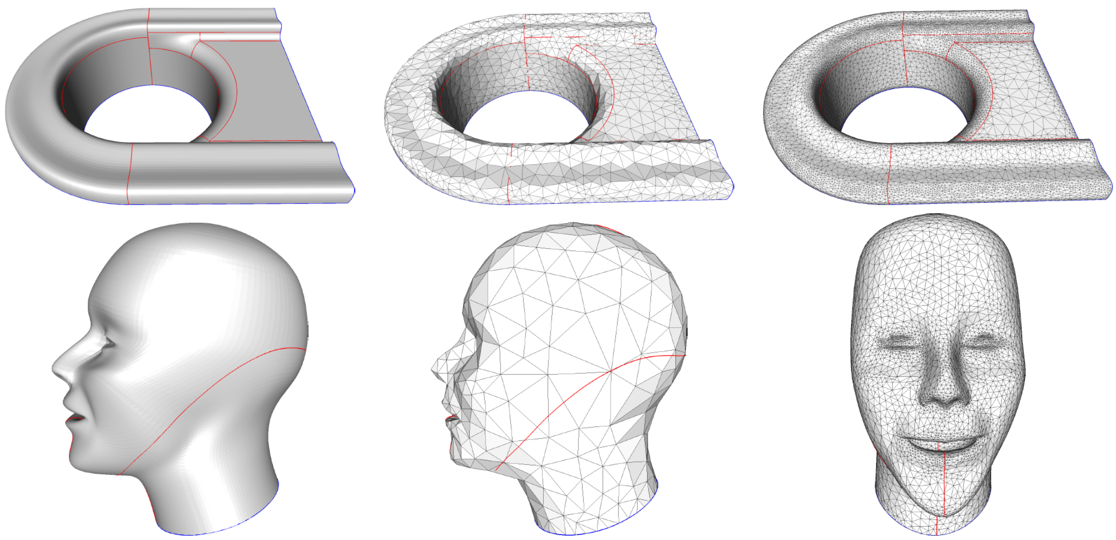


Figure 8. Seamless meshing. Top: meshing with two mesh sizing values. The initial point set is generated by sampling along the open boundary. Bottom: meshing across smooth edges (red). The initial point set is generated by sampling along the boundary edge (blue). Meshes generated with two sizing values (side and front view).

We formulate an automatic approach to the detection of partial, local, and global symmetries and orbits in arbitrary 3D datasets. We improve upon existing voting-based symmetry detection techniques by leveraging the Lie group structure of geometric transformations. In particular, we introduce a logarithmic mapping that ensures that orbits are mapped to linear subspaces, hence unifying and extending many existing mappings in a single Lie-algebra voting formulation (Figure 10). Compared to previous work, our resulting method offers significantly improved robustness as it guarantees that our symmetry detection of an input model is frame, scale, and reflection invariant. As a consequence, we demonstrate that our approach efficiently and reliably discovers symmetries and orbits of geometric datasets without requiring heavy parameter tuning. This work was published in the proceedings of the EUROGRAPHICS Symposium on Geometry Processing [].

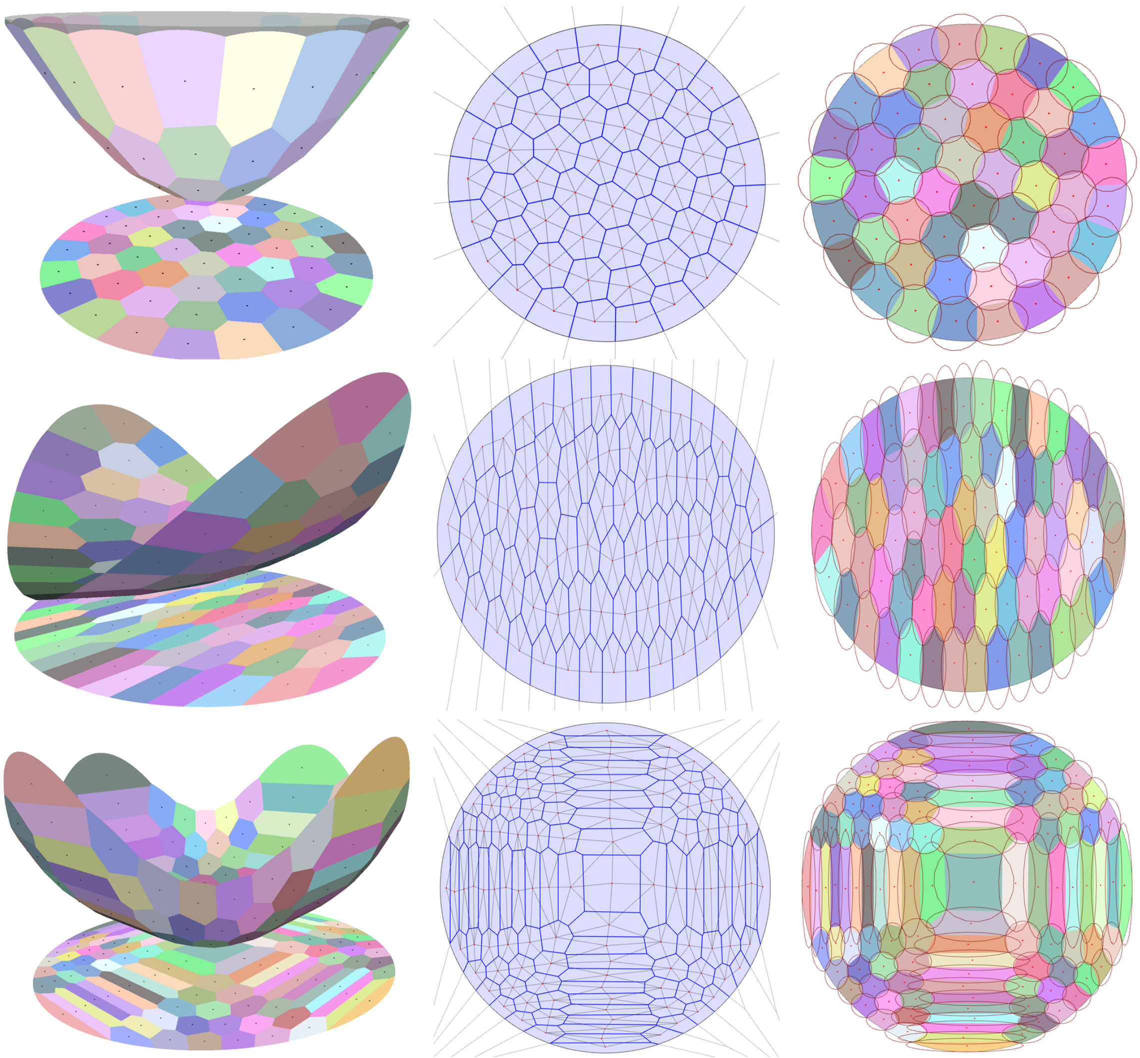


Figure 9. Optimal Voronoi Tessellations with Hessian-based Anisotropy. We show that the construction of an optimal piecewise-linear approximation of a function over a cell complex (left) extends the isotropic notion of Centroidal Voronoi Tessellations (CVT, top) to an anisotropic variant (middle and bottom) we call Optimal Voronoi Tessellation (OVT), to stress its duality to Optimal Delaunay Triangulation (ODT). Cell anisotropy (indicated by tightest ellipses) and density are independently controlled, and the dual triangulation based on cell barycenters is embedded and combinatorially-regular.

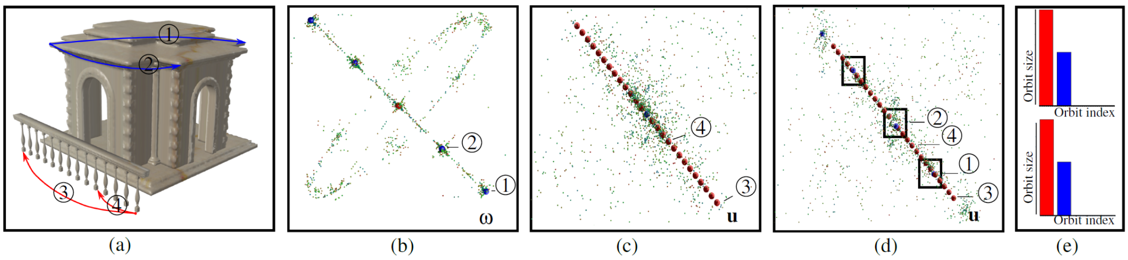


Figure 10. Our Lie algebra voting approach to symmetry and orbit detection maps SE(3) transformations into points in a logarithmic space composed of a rotation part and a translation part. The rotational orbit of the church and the translational orbit of the side railing (a) are mapped into collinear blue and red spheres respectively (a few transformations within these two orbits are marked with circled numbers to enhance comprehension). When the scene is centered, the two lines are orthogonal to each other and easy to distinguish (b). However, after a rigid translation of the scene, the rotational orbit now has translation-values near the translation orbit points, making it impossible to automatically distinguish these two orbits using a Euclidean distance (d), while our adjoint invariant distance for orbit shows no discernible difference in results as evidenced by a binning of detected orbit sizes for both situations (e).

ALPAGE Project-Team

6. New Results

6.1. Deep syntactic parsing

Participants: Corentin Ribeyre, Marie-Hélène Candito.

Syntax plays an important role in the task of predicting the semantic structure of a sentence. But syntactic phenomena such as alternations, control and raising tend to obfuscate the relation between syntax and semantics. We have investigated how to predict the semantic structure of a sentence, encoded using the FrameNet model, taking advantage of deeper syntactic information than what is usually used. This deep syntactic representation abstracts away from purely syntactic phenomena and proposes a structural organization of the sentence that is closer to the semantic representation, by normalising the syntactic paths between a verb and its arguments. This reduces the variety of the syntactic realization of semantic roles, as shown by a decrease of the entropy of the syntactic paths of a given role.

Experiments conducted on a French corpus annotated with semantic frames showed that a FrameNet semantic parser reaches better performances with such a deep syntactic information [31]. For instance, switching from surface to deep syntactic information leads to a significant gain in FrameNet role identification, especially when this information is predicted (rather than reference information): $+5.1$ points (56.7 to 61.7) on all triggers [0] and $+6.7$ points (61.3 to 68.0) on verbal triggers only. These results clearly show the benefit of using deep syntactic features.

6.2. Multilingual POS-tagging

Participant: Benoît Sagot.

Morphosyntactic lexicons and word vector representations have both proven useful for improving the accuracy of statistical part-of-speech taggers. We compare the performances of four systems on datasets covering 16 languages, two of these systems being feature-based (MEMMs —in the case of our own system MElt— and CRFs) and two of them being neural-based (bi-LSTMs). We show that, on average, all four approaches perform similarly and reach state-of-the-art results. Yet we obtained better performances with feature-based models on lexically richer datasets (e.g. for morphologically rich languages), whereas neural-based results are higher on datasets with less lexical variability (e.g. for English). These conclusions hold in particular for the MEMM models relying on our system MElt, which benefited from newly designed features [32], [44]. Thus we have shown that, under certain conditions, feature-based approaches enriched with morphosyntactic lexicons are competitive with respect to neural methods.

6.3. Transition-based constituency parsing with HyParse

Participants: Benoît Crabbé, Maximin Coavoux.

Transition-based parsing reduces the parsing task to predict a sequence of atomic decisions. These decisions are taken while sequentially reading words from a buffer and combining them incrementally into syntactic structures. The resulting structures are often dependency structures but can also be constituents, as is the case for our parser HyParse. Such an approach is therefore linear in the length of the input sentence, making transition-based parsing computationally efficient relative to other approaches. The challenge in transition-based parsing is modelling which action should be taken in each state it encounters as it progresses in a sentence provided as an input.

[0] In the sense of FrameNet, i.e. predicative lexial units, which should be assigned a frame.

Training of a transition-based parser therefore consists in training a function that maps each of the unboundedly many states the parser might encounter to the best possible action, or transition, it should take. This function generally relies on a huge set of features, often conveniently grouped in the form of more abstract feature templates. Yet selecting the optimal subset of feature(template)s remains a challenge.

The training procedure therefore requires the help of an “oracle”, that is a function that returns the action that the parser should take in a given parser state given the gold parse. If the oracle assumes that the next action is necessarily the one given in the gold parse, it is said to be "static” and the oracle is deteminist. In order to train the parser to take relevant decisions when in an erroneous state, we can introduce some non determinism in the oracle in order to explore not only gold transition sequences but also near-gold transition sequences. This is the purpose of a dynamic oracle. Dynamic oracle training has shown substantial improvements for dependency parsing in various settings, but had not previously been explored for constituent parsing.

The two research directions we have investigated reflect the two above-mentioned challenges.

First, in collaboration with Rachel Bawden, now PhD student at LIMSI, we resumed our work on developing an efficient, language-independent model selection method for our parser HyParse [61]. It is designed for model selection when faced with a large number of possible feature templates, which is typically the case for morphologically rich languages, for which we want to exploit morphological information. The method we proposed uses multi-class boosting for iterative selection in constant time, using virtually no *a priori* constraints on the search space. We did however use a pre-ranking step before selection in order to guide the selection process. Our experiments have illustrated the feasibility of the method for our working language, French and resulted in high-performing, compact models much more efficiently than naive methods [22].

Second, we developed a dynamic oracle for HyParse. First, we replaced the traditional feature-based approach used in the above-described experiments by a neural approach. This is a way to overcome the feature selection issue addressed in the above-described work. The neural network weighting function we developed uses a non-linear hidden layer to automatically capture interactions between variables, and embeds morphological information in a vector space, as is usual for words and other symbols. Then, we developed our dynamic oracle based on this neural function and conducted experiments on the 9 languages of the SPMRL dataset in order to assess the impact of this oracle [25]. The experiments have shown that a neural greedy parser with morphological features, trained with a dynamic oracle, leads to accuracies comparable with the best currently available non-reranking and non-ensemble parsers.

6.4. French FrameNet

Participants: Marie-Hélène Candito, Marianne Djemaa.

In 2016 we have continued the development of a French FrameNet, within the ASFALDA project. While the first phase of the project focused on the development of a French set of frames and corresponding lexicon (Candito et al., 2014), we have focused this year on the subsequent corpus annotation phase, which targeted four notional domains (commercial transactions, cognitive stances, causality and verbal communication). Given full coverage is not reachable for a relatively “new” FrameNet project such as ours, focusing on specific notional domains allowed us to obtain full lexical coverage for the frames of these domains, while partially reflecting word sense ambiguities. Furthermore, as frames and roles were annotated on two main French Treebanks (the French Treebank and the Sequoia Treebank), we were able to extract a syntactico-semantic lexicon from the annotated frames. In the resource’s current status [28], there are 98 frames, 662 frame-evoking words or “triggers”, 872 senses, and about 13,000 annotated frames, with their semantic roles assigned to portions of text [0]

During this year’s resource development efforts, we have put a specific emphasis on the causality domain (about 4000 instances of causal lexical items with their corresponding semantic frames are included in our resource). In the process of building the French lexicon and preparing the annotation of the corpus, we had to remodel some of the frames proposed in FrameNet based on English data, with hopefully more precise frame definitions to facilitate human annotation. This includes semantic clarifications of frames and frame elements,

[0]The French FrameNet is freely available at http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml.

redundancy elimination, and added coverage. The result is arguably a significant improvement of the treatment of causality in FrameNet itself [34].

6.5. Verb∋net

Participants: Lucie Barque, Laurence Danlos.

VerbNet is a lexical resource for English verbs in which verbs are grouped together based on their ability to appear in similar sets of syntactic frames that correspond as well to alternations exhibited by verbs as to alternative syntactic realizations (Kipper et al. 2004). A French Verbnet, named Verb∋net, was first automatically derived from English VerbNet (Pradet et al., 2014) and is still under development. [13] details how Verb∋net was developed from the English VerbNet while using as far as possible the available lexical resources for French and how the various French alternations are coded, focusing on differences with English (e.g. existence of pronominal forms). One difficulty encountered in the development of Verb∋net springs from the fact that the list of (potentially numerous) frames has no internal organization in VerbNet. [26] proposes a type system for frames that shows whether two frames are variants of a given alternation. Frame typing facilitates coherence checking of the resource in a “virtuous circle”.

6.6. French FrameNet

Participant: Benoît Crabbé.

Elaborating on our previous work on Medieval French in collaboration with Sasha Simonenko (McGill) and Sophie Prévost (LATTICE), we have conducted the first large-scale quantitative investigation of the syncretisation of verbal subject agreement in this language and test a classic analysis which relates non-syncretic agreement and null subjects as parts of the same grammar. We have shown that agreement syncretisation and the emergence of overt pronominal subjects proceeded at the same rate. Under the Constant Rate Hypothesis of Kroch (1989), which states that a grammatical change has the same rate in different contexts, these results are compatible with the traditional analysis [40], [39], [33]. However, we show that this analysis also generates a number of predictions which are not borne out by the quantitative data. We conclude that a more complex model of interaction of subject and inflection parameters is needed. Such a model may for instance be one where the type of an ending (non-syncretic vs. syncretic), presumably dependent on some unrelated phonological mechanism, presents a parsing difficulty for a null subject-licensing grammar and thus lowers its probability to be chosen by the speaker, which eventually drives it to extinction, similarly to the grammar competition model proposed in Yang (2010).

We have also investigated the effects of the text form (prose vs. verse) on diachronic grammatical changes in Medieval French using parsed treebanks and (1 million words with PTB-like annotations). Despite the common intuition that the prose is somehow more “advanced” than the verse contemporary to it with respect to grammatical changes, the magnitude of the difference has remained unknown in the absence of quantificational evaluations. At the same time, the prevalence of verse in the earliest periods of documented French (i.e. X–XII c.) results in a strong and unavoidable correlation between time and form, which potentially undermines the results of the studies attempting to formally model Medieval French evolution. We have compared two historical changes across text forms (namely the loss of pro-drop and that of OV_{finite} order), and shown that verse and prose behave differently, at least regarding the OV_{finite} order, thus contradicting Kroch’s (1989) Constant Rate Hypothesis [38].

6.7. Modelling discourse-level information

Participants: Laurence Danlos, Timothée Bernard.

We have continued our work on the formalisation of discourse-level information. First, we have proposed in [24] a new model in STAG syntax and semantics for subordinate conjunctions (SubConjs) and attributing phrases —attitude/reporting verbs (AVs; *believe*, *say*) and attributing prepositional phrase (APPs; *according to*). This discourse-oriented model is based on the observation that SubConjs and AVs are not homogeneous categories. Indeed, previous work has shown that SubConjs can be divided into two classes according to their syntactic and semantic properties. Similarly, AVs have two different uses in discourse: evidential and intentional. While evidential AVs and APPs have strong semantic similarities, they do not appear in the same contexts when SubConjs are at play. Our proposition aims at representing these distinctions and capturing these various discourse-related interactions.

We have also investigated how sentential and discourse TAG-based grammars can be interfaced, in collaboration with Aleksandre Maskharashvili and Sylvain Pogodalla (LORIA). Tree-Adjoining Grammars (TAG) have been used both for syntactic parsing, with sentential grammars, and for discourse parsing, with discourse grammars (see for example our D-STAG model or the D-LTAG model). Yet the modelling of discourse connectives (coordinate conjunctions, subordinate conjunctions, adverbs...) in TAG-based formalisms for discourse differ from their modelling in sentential grammars. Because of this mismatch, an intermediate processing step is required between the sentential and the discourse processes, both in parsing and in generation [27]. We have developed a method to smoothly interface sentential and discourse TAG grammars, without using such an intermediate processing step. This method, based on Abstract Categorial Grammars (ACG), allows for building D-STAG discourse structures that are direct acyclic graphs (DAG) and not only trees.

6.8. Detecting omissions in journalistic texts

Participants: Héctor Martínez Alonso, Benoît Sagot.

In the journalistic genre that is characteristic of online news, editors make frequent use of citations as prominent information; yet these citations are not always in full. The reasons for leaving information out are often motivated by the political leaning of the news platform.

Existing approaches to the detection of political bias rely on bag-of-words models that examine the words present in the writings. In the context of the VerDI project (see below), we have initiated work aimed at going beyond such approaches, which focus on what is said, by instead focusing on what is *ommited*. Thus, this method requires a pair of statements; an original one, and a shortened version with some deleted words or spans. The task is then to determine whether the information left out in the second statement conveys *substantial* additional information. If so, we consider that a certain statement pair presents an omission. To tackle this question, we used a supervised classification framework, for which we require a dataset of sentence pairs, each pair manually annotated for omission.

We have developed a small reference corpus for evaluation purposes, using and comparing both crowd and expert annotation. This corpus has allowed us to examine which features help automaticallly identify cases of omission. In addition to straightforward measures of word overlap (the Dice coefficient), we also determined that there is a good deal of lexical information that determines whether there is an omission. This work is, to the best of our knowledge, the first empirical study on omission identification in statement pairs. We shall make all data and annotations freely available upon publication.

6.9. Models for interoperable lexical data

Participants: Mohamed Khemakhem, Laurent Romary.

Lexical data play an essential role in computational linguistic in two complementary ways:

- They serve as basic resources with which computational linguistic process can be parameterized. Such lexical resources are usually automatically or semi-automatically produced, are highly structured and may cover various levels of linguistic description from basic morpho-syntactic content to semantic representations;

- When created manually either for the purpose of describing a language (mono- or multilingual dictionary) or as a by product other language based activities (e.g. technical writing, translation), they may serve as a primary source of observation to analyse the way the lexicon of a language is organized, is used in domain oriented content, or how languages vary across time, space and usage.

The Alpage team has a specific expertise in the domain of lexical data, having been involved in the recent years in the creation of reference resources for the French language in particular, but also as driving force in the definition of international standards for the modelling and representation of both semasiological (word to sense) and onomasiological (concept to term) lexical information:

- ISO 16642 (TMF, Terminological Markup framework) and ISO 30042 (TBX, TermBase eXchange) as reference standards for the interchange of terminological data, for instance between translators' workbenches, but also for the modelling of dialectal information in linguistics;
- ISO 24613 (LMF, Lexical Markup Framework), a modular modelling framework for the representation of both machine and human semasiological resources;
- The Text Encoding Initiative (TEI), which since its inception has provided an XML based format for human readable dictionaries, widely used in most last scale dictionary projects worldwide.

One of the difficulties in lexical modelling is to identify the proper modelling framework for a given lexical resource but also to ensure maximal interoperability across heterogeneous lexical content. In the recent period, we have been working on the following aspects:

- Participation in the on going revision of ISO 30046, and planning of a possible integration of a TBX dialect in the TEI guidelines;
- Setting up the revision of ISO 24613 as a multi-part standard. Alpage is now involved in the provision of a reference TEI based serialisation of LMF and the part dedicated to etymological/diachronical information;
- Proposing an extension to the TEI guidelines for the representation of etymological information in dictionaries thus offering a formal basis for the study of diachronical phenomena across dictionaries [46];
- Organising a workshop in the context of the COST action eNEL that brought together the most relevant experts in the field in order to provide a set of constraints to apply the TEI guidelines in a more interoperable way across dictionary projects;
- Starting working on a machine learning based process to extract lexical content and structure automatically from digitized legacy dictionaries, This activity, base don the architecture of the Grobid library, is the basis of the PhD work by Mohamed Khemakhem.

6.10. Open data in the arts and humanities

Participants: Luca Foppiano, Marie Puren, Charles Riondet, Laurent Romary, Dorian Seillier.

The issue of open data has become increasingly important in various scholarly domains for it impacts on the visibility of the corresponding works, the capacity to provide evidence for reported facts and results, but also let other scholars build up new research on existing data sets. This is particularly acute in the humanities where primary sources play an essential role in providing the core material of scholarly results and for which the digital turn has offered a unique perspective of building up a wealth of structure information about human traces at large.

Based upon the experience gained in the definition of the open access policy at Inria [42], [50], [43], we have pursued various activities leading to a better understanding of the technical, editorial and political factors that may improve the wide dissemination of scholarly data sets in the humanities:

- Carry out a large scale questionnaire on data re-use within the partnership of the Iperion projects, which showed the lack of a coherent data management policy across cultural heritage laboratories in Europe from the points of view of documentation, archiving, licencing and re-use [49];

- Design a concept [16], [41] to improve the general fluidity of research results in the humanities based on data quality assessment, data journals and above all the setting of of a data re-use charter between scholars and cultural research institutions in the humanities. This action, carried out in the context of the Parthenos project has started with the organisation of two high level workshops in Berlin and Paris with representatives of major cultural research institutions;
- Coordinate as leader of WP 4 (Standards) in the Parthenos project a major overview of the needs and possible deployment of standards in the humanities based of an in depth survey of possible research scenario and associated practices in the domain of standards (Deliverable 4.1 published in October 2016). This has been accompanied by specific technical developments such as the proposition of an extension to the TEI guidelines for the representation of embedded stand-off annotations [45], [51];
- Develop specific modules for mining digital sources in the humanities, in particular in the domain of named entity recognition as an improvement of the NERD software initially developed in the European Cendari project.

MULTISPEECH Project-Team

7. New Results

7.1. Explicit Modeling of Speech Production and Perception

Participants: Yves Laprie, Slim Ouni, Vincent Colotte, Anne Bonneau, Agnès Piquard-Kipffer, Denis Jouvet, Odile Mella, Dominique Fohr, Benjamin Elie, Sucheta Ghosh, Anastasiia Tsukanova, Yang Liu, Sara Dahmani, Valérian Girard, Aghilas Sini.

7.1.1. Articulatory modeling

7.1.1.1. Acoustic simulations

The acoustic simulations play a central role in articulatory synthesis and should enable the production of all classes of sounds in a realistic manner. The production of voiced fricatives relies on a partial closure of the glottis which simultaneously creates an airflow which generates turbulence downwards from the constriction and the vibration of the vocal folds. Our acoustic simulation framework [14] has been extended to incorporate a glottal chink [29] in a self-oscillating vocal fold model. The glottis is then made up of two main separated components: a self-oscillating part and a constantly open chink. This feature allows the simulation of voiced fricatives, thanks to a self-oscillating model of the vocal folds to generate the voiced source, and the glottal opening that is necessary to generate the frication noise.

The acoustic propagation paradigm is appropriately chosen so that it can deal with complex geometries and a time-varying length of the vocal tract. Temporal scenarios for the dynamic shapes of the vocal tract and the glottal configurations were derived from the simultaneous acquisition of X-ray or MRI images and audio recording. Copy synthesis of a few French sentences [30], [31], [53] shows the accuracy of the simulation framework to reproduce acoustic cues of phrase-level utterances containing most of French phone (sound) classes while considering the real geometric shape of the speaker. For this purpose the articulatory model has been extended to offer a better precision of the epiglottis and of lips.

7.1.1.2. Acquisition of articulatory data

The acquisition of dynamic data is a key objective since speech production gestures involve the anticipation of the articulatory targets of the coming sounds. Cine-MRI represents an invaluable tool since it can image the whole vocal tract. However, speech requires a sampling frequency above 30 Hz to capture interesting information. Compressive sampling relies on partially collecting data in the Fourier space of the images acquired via MRI. The combination of compressed sensing technique, along with homodyne reconstruction, enables the missing data to be recovered [32]. The good reconstruction is guaranteed by an appropriate design of the sampling pattern. It is based on a pseudo-random Cartesian scheme, where each line is partially acquired for use of the homodyne reconstruction, and where the lines are pseudo-randomly sampled: central lines are constantly acquired and the sampling density decreases as the lines are far from the center.

7.1.1.3. Markerless articulatory acquisition techniques

With the spread of depth cameras (kinect-like systems), many researchers consider using these systems to track the movement of some speech articulators as lips and jaw. We are considering using this kind of system if it is suitable for speech production studies. For this reason, we have assessed the precision of markerless acquisition techniques when used to acquire articulatory data for speech production studies [19]. Two different markerless systems have been evaluated and compared to a marker-based one. The main finding is that both markerless systems provide reasonable results during normal speech and the quality is uneven during fast articulated speech. The quality of the data is dependent on the temporal resolution of the markerless system.

7.1.2. Expressive acoustic-visual synthesis

7.1.2.1. Expressive speech

A comparison between emotional and neutral speech was conducted using a small database containing utterances recorded in six emotional types (anger, fear, sadness, disgust, surprise and joy) as well as in a neutral pronunciation. The prosodic analysis focused on the main prosodic parameters such as vowel duration, energy and fundamental frequency (F0) level, and pause occurrences. The values of prosodic parameters were compared among the various emotional styles, as well as between emotional style and neutral style utterances. Moreover, the structuration of the sentences, in the various emotional styles, was particularly studied through a detailed analysis of pause occurrences and their length, and of the length of prosodic groups [23].

7.1.2.2. Expressive acoustic and visual speech

Concerning expressive audiovisual speech synthesis, a case study of a semi-professional actor who uttered a set of sentences for 6 different emotions in addition to neutral speech was conducted. Our purpose is to identify the main characteristics of audiovisual expressions that need to be integrated during synthesis to provide believable emotions to the virtual 3D talking head. We have recorded concurrently audio and motion capture data. The acoustic and the visual data have been analyzed. The main finding is that although some expressions are not well identified, some expressions were well characterized and tied in both acoustic and visual space [40]. The acquisition of the corpus was done with the platform software PLAVIS (cf. 9.2.12).

7.1.3. Categorization of sounds and prosody for native and non-native speech

7.1.3.1. Categorization of sounds for native speech

We examined the schooling experiences of 166 young people with disabilities, aged from 6 to 20 years old. These children and teenagers had specific language impairment : SLI (severe language impairment), dyslexia, dysorthographia. The phonemic discrimination, phonological and phonemic analysis difficulties faced in their childhoods had raised reading difficulties which constituted a major obstacle, which the pupils did not overcome. Consequently, this led them to repeat one or more grades. This rate is 18 times higher than the French average. The importance of this cycle of learning can be better understood through this data, which could also enable, if not overcoming the handicap, to at least improving their learning possibilities [64].

7.1.3.2. Digital books for language impaired children

Three digital albums for language impaired children were designed within the Handicom (ADT funded by Inria). These three prototypes focus on the importance of multimodal speech combining written words and visual clues: a 3D avatar telling the stories and coding oral language in LPC (french cued speech) for hearing impaired children. Eight speech and language therapists used one of these albums (the digital prototype *Nina fête son anniversaire* !) with 8 children who are aged 5 years: 4 hearing impaired children, 2 children with SLI and 2 children with autism. The training they experienced with these children showed that the use of the digital book can foster some capacities involved in language learning [41].

7.1.3.3. Analysis of non-native pronunciations

The IFCASL corpus is a French-German bilingual phonetic learner corpus designed, recorded and annotated in the IFCASL project (cf. 9.2.6). It incorporates data for a language pair in both directions, i.e. in our case French learners of German, and German learners of French. In addition, the corpus is complemented by two sub-corpora of native speech by the same speakers. The corpus has been finalized, and provides spoken data by about 100 speakers with comparable productions, annotated and segmented at the word and phone levels, with more than 50% of manually checked and corrected data [51].

We investigated the correct placement of lexical (German) or post-lexical (French) accents [52]. French and German differ with respect to the representation and implementation of prominence. French can be assumed to have no prominence represented in the mental lexicon and accents are regularly assigned post-lexically on the last full vowel of an accentual group. In German, prominence is considered to be represented lexically. This difference may give rise to interferences when German speakers learn French and French speakers learn German. Results of a judgment task (conducted with 3 trained phoneticians) of native and nonnative productions of French learners of German and German learners of French, all of them beginners, show that both groups have not completely acquired the correct suprasegmental structures in the respective L2 [0], since both groups are worse concerning the correct placement of prominence than the native speakers. Furthermore, the results suggest that the native pattern is one of the most important factors for wrong prominence placements in the foreign language, e.g., if the prominence placement of L1 and L2 coincide, speakers produce the smallest amount of errors. Finally, results indicate that visual display of accented syllables increases the likelihood of a correct accent placement.

7.1.3.4. Implementation of acoustic feedback for devoicing of final fricatives

In view of implementing acoustic feedback in foreign language learning we analyzed acoustic cues which could explain that final fricatives are perceived as voiced or unvoiced. The ratio of unvoiced frames in the consonantal segment and also the ratio between consonantal duration and vowel duration were measured. As expected, we found that beginners face more difficulties to produce voiced fricatives than advanced learners. Also, the production becomes easier for the learners, especially for beginners, if they practice repetition after a native speaker. We use these findings to design and develop feedback via speech analysis/synthesis technique TD-PSOLA using the learner's own voice and voiced fricatives uttered by French speakers [36]. We selected fully voiced exemplars and evaluated whether the presence of an additional schwa fosters the perception of voicing by native French speakers.

7.2. Statistical Modeling of Speech

Participants: Antoine Liutkus, Emmanuel Vincent, Irène Illina, Dominique Fohr, Denis Jouvet, Vincent Colotte, Ken Deguernel, Mathieu Fontaine, Amal Houidhek, Aditya Nugraha, Imran Sheikh, Imene Zangar, Mohamed Bouallegue, Sunit Sivasankaran.

7.2.1. Source separation

7.2.1.1. Deep neural models for source separation

We pursued our research on the use of deep learning for multichannel source separation [18]. Our technique exploits both the spatial properties of the sources as modeled by their spatial covariance matrices and their spectral properties as modeled by a deep neural network. The model parameters are alternately estimated in an expectation-maximization (EM) fashion. We used this technique for music separation in the context of the 2016 Signal Separation Evaluation Campaign (SiSEC) [39]. We also used deep learning to address the fusion of multiple source separation techniques and found it to perform much better than the variational Bayesian model averaging techniques previously investigated [17].

We wrote an article about music source separation for the general public [59].

7.2.1.2. α-stable modeling of audio signals

The alpha-harmonizable model has recently been proposed by A. Liutkus et al. [66] as the only available probabilistic framework to account for signal processing methods manipulating fractional spectrograms instead of more traditional power spectrograms. Indeed, they generalize the classical Gaussian formulation and permit to handle large uncertainties or signal dynamics, which are both common in audio.

Our work on this topic this year has notably focused on its extension to the multichannel setting, which is important for music processing and source localization. Since inference in multivariate alpha-stable distribution is a very intricate issue, the approach undertaken has focused on analysing the multichannel signals through the joint analysis of multiple scalar projections on the real line. This results in an original algorithm called PROJET that combines computational tractability with the inherent robustness of alpha-stable models [15], [34].

[0] L2 indicates the non-native language, whereas L1 indicates the native language

7.2.2. Acoustic modeling

7.2.2.1. Noise-robust acoustic modeling

In many real-world conditions, the target speech signal is reverberated and noisy. In order to motivate further work by the community, we created an international evaluation campaign on that topic in 2011: the CHiME Speech Separation and Recognition Challenge. After three successful editions [11], [55], we organized the fourth edition in 2016. We also summarized the speech distortion conditions in real scenarios for speech processing applications [42] and collected a French corpus for distant-microphone speech processing in real homes [24].

Speech enhancement and automatic speech recognition (ASR) are most often evaluated in matched (or multi-condition) settings where the acoustic conditions of the training data match (or cover) those of the test data. We conducted a systematic assessment of the impact of acoustic mismatches (noise environment, microphone response, data simulation) between training and test data on the performance of recent DNN-based speech enhancement and ASR techniques [21]. The results show that most algorithms perform consistently on real and simulated data and are barely affected by training on different noise environments. This suggests that DNNs generalize more easily than previously thought.

7.2.2.2. Environmental sounds

We explored acoustic modeling for the classification of environmental sound events and sound scenes and submitted our system to the DCASE 2016 Challenge [33].

7.2.3. Linguistic modeling

7.2.3.1. Out-of-vocabulary proper name retrieval

The diachronic nature of broadcast news causes frequent variations in the linguistic content and vocabulary, leading to the problem of Out-Of-Vocabulary (OOV) words in automatic speech recognition. Most of the OOV words are found to be proper names whereas proper names are important for automatic indexing of audio-video content as well as for obtaining reliable automatic transcriptions. New proper names missed by the speech recognition system can be recovered by a dynamic vocabulary multi-pass recognition approach in which new proper names are added to the speech recognition vocabulary based on the context of the spoken content [47]. The goal of this work is to model the semantic and topical context of new proper names in order to retrieve OOV words which are relevant to the spoken content in the audio document. Probabilistic topic models [44] and word embeddings from neural network models are explored for the task of retrieval of relevant proper names. Neural network context models trained with an objective to maximise the retrieval performance are proposed. A Neural Bag-of-Words (NBOW) model trained to learn context vector representations at a document level is shown to outperform the generic representations. The proposed Neural Bag-of-Weighted-Words (NBOW2) model learns to assign a degree of importance to input words and has the ability to capture task specific key-words [46] [45]. Experiments on automatic speech recognition of French broadcast news videos demonstrate the effectiveness of the proposed models. Further evaluation of the NBOW2 model on standard text classification tasks, including movie review sentiment classification and newsgroup topic classification, shows that it learns interesting information about the task and gives the best classification accuracies among the bag-of-words models.

7.2.3.2. Adding words in a language model

Out-of-vocabulary (OOV) words can pose a particular problem for automatic speech recognition of broadcast news. The language models (LMs) of ASR systems are typically trained on static corpora, whereas new words (particularly new proper nouns) are continually introduced in the media. Additionally, such OOVs are often content-rich proper nouns that are vital to understanding the topic. We explore methods for dynamically adding OOVs to language models by adapting the n-gram language model used in our ASR system. We propose two strategies: the first one relies on finding in-vocabulary (IV) words similar to the OOVs, where word embeddings are used to define similarity. Our second strategy leverages a small contemporary corpus to estimate OOV probabilities. The models we propose yield improvements in perplexity over the baseline; in addition, the corpus-based approach leads to a significant decrease in proper noun error rate over the baseline in recognition experiments [26].

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. We published two articles that summarize our work on the System & Contrast model for the characterization of the mid-term and long-term structure of music [12] and on the structural segmentation of popular music pieces using a regularity constraint that naturally stems from this model [20], [58]. We also proposed a new model for automatic music improvisation that combines a multi-dimensional probabilistic model encoding the musical experience of the system and a factor oracle encoding the local context of the improvisation [27].

7.2.4. Speech generation by statistical methods

Work on HMM-based Arabic speech synthesis was carried out within a CMCU PHC project with ENIT (Engineer school at Tunis-Tunisia; cf. 9.4.2.2). A first version of the system, based on the HTS toolkit (HMM-based Speech Synthesis System), is now working; and the study of the impact of some parameters is ongoing.

In parallel, the HTS system is also applied to the French language.

7.3. Uncertainty Estimation and Exploitation in Speech Processing

Participants: Emmanuel Vincent, Odile Mella, Dominique Fohr, Denis Jouvet, Baldwin Dumortier, Juan Andres Morales Cordovilla, Karan Nathwani, Ismaël Bada.

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 in order to exploit it for subsequent feature extraction and decoding. A number of uncertainty estimators have been proposed in the literature, which are typically based on fixed mathematical approximations or heuristics. We finalized our work on a principled variational Bayesian approach to uncertainty estimation and showed its benefit w.r.t. other estimators for speech and speaker recognition [9]. We also pursued our work on the propagation of uncertainty in deep neural network acoustic models.

7.3.1.2. Uncertainty in other applications

Besides the above applications, we pursued our exploration of uncertainty modeling for robot audition and wind turbine control. In the first context, uncertainty arises about the location of acoustic sources and the robot is controlled to locate the sources as quickly as possible [38]. In the second context, uncertainty arises about the noise intensity of each wind turbine and the turbines are controlled to maximize electrical production under a maximum noise threshold [62].

7.3.2. Uncertainty and phonetic segmentation

7.3.2.1. Speech-text alignment

We have continued our work on determining more accurate phonetic boundaries with two new approaches based on DNN. The first approach proposes to find phonetic boundaries directly from the parameterized speech signal using an LSTM (Long Short-Term Memory) neural network. The aim of the second approach is twofold: provide confidence measures for evaluating speech-text alignment outputs and refine these outputs. One of these studies was done with the Synalp team of LORIA in the framework of the project ORFEO (cf. 9.2.5). The achieved confidence measure outperforms a confidence score (based on acoustic posterior probability) derived from a state-of-the-art text-to-speech aligner [43].

Within the IFCASL project (cf. 9.2.6), we have also developed a speech-text alignment system for German which will be integrated into the ASTALI software.

7.3.3. Uncertainty and prosody

The study of discourse particles that was initiated last year, has continued in the framework of the CPER LCHN (cf. 9.1.2). A larger set of words and expressions that can be used either as normal lexical words or as discourse particles (as for example *quoi* (what), *voilà* (there it is), ...) has been considered. For each of these words/expressions and for each speech corpus that was aligned in the ORFEO project (cf. 9.2.5), a subset of about one hundred occurrences were selected. Thanks to the CPER LCHN support, a part of these occurrences have been annotated as "discourse particle" or "non discourse particle". Detailed analysis is in progress, with respect to the function (discourse particle or not), the type of speech corpus, and the associated prosodic features.

The fundamental frequency is one of the prosodic features. Numerous approaches exist for the computation of F0. Most of them lead to good performance on good quality speech. The performance degradation with respect to noise level has been studied on reference databases, for several (about ten) F0 detection approaches. It was observed that for each algorithm, a large part of the errors are due to incorrect voiced/unvoiced decision. Studies have also been initiated for computing a confidence measure on the estimated F0 values through the use of neural network approaches.

PANAMA Project-Team

7. New Results

7.1. Recent results on Sparse Representations, Inverse Problems, and Dimension Reduction

Sparsity, low-rank, dimension-reduction, inverse problem, sparse recovery, scalability, compressive sensing

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the fields of sparse representations, inverse problems, and dimension reduction, which is at the core of the ERC project PLEASE (Projections, Learning and Sparsity for Efficient Data Processing, see Section 9.2.1.1).

7.1.1. Theoretical results on Sparse Representations, Graph Signal Processing, and Dimension Reduction

Participants: Rémi Gribonval, Yann Traonmilin, Gilles Puy, Nicolas Tremblay, Pierre Vandergheynst.

Main collaboration: Mike Davies (University of Edinburgh), Pierre Borgnat (ENS Lyon), and members of the LTS2 lab of Pierre Vandergheynst at EPFL

Stable recovery of low-dimensional cones 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 established generic RIP-based guarantees for the stable recovery of cones (positively homogeneous model sets) with arbitrary regularizers. These guarantees were illustrated on selected examples. For block structured sparsity in the infinite dimensional setting, we used the guarantees for a family of regularizers which efficiency in terms of RIP constant can be controlled, leading to stronger and sharper guarantees than the state of the art. This has been published in a journal paper [21].

Recipes for stable linear embeddings from Hilbert spaces to $\mathbb{R}^m$: We considered the problem of constructing a linear map from a Hilbert space (possibly infinite dimensional) to $\mathbb{R}^m$ that satisfies a restricted isometry property (RIP) on an arbitrary signal model set. We obtained a generic framework that handles a large class of low-dimensional subsets but also *unstructured* and *structured* linear maps. We provided a simple recipe to prove that a random linear map satisfies a general RIP on the model set with high probability. We also described a generic technique to construct linear maps that satisfy the RIP. Finally, we detailed how to use our results in several examples, which allow us to recover and extend many known compressive sampling results. This has been presented at the conference EUSIPCO 2015 [90], and a journal paper is under revision [91].

Signal processing on graphs: from filtering to random sampling and robust PCA: Graph signal processing is an emerging field aiming at extending classical tools from signal processing (1D time series) and image processing (2D pixel grids, 3D voxel grids) to more loosely structured numerical data: collections of numerical values each associated to a vertex of a graph, where the graph encodes the underlying “topology” of proximities and distances. Since our pioneering contributions on this topic [4], the team regularly works on various aspects of graph signal processing, in collaboration with the LTS2 lab of Pierre Vandergheynst at EPFL. This year, we studied the problem of sampling k-bandlimited signals on graphs. We proposed two sampling strategies that consist in selecting a small subset of nodes at random. The first strategy is non-adaptive, i.e., independent of the graph structure, and its performance depends on a parameter called the graph coherence. On the contrary, the second strategy is adaptive but yields optimal results. Indeed, no more than $O(klog(k))$ measurements are sufficient to ensure an accurate and stable recovery of all k-bandlimited signals. This second strategy is based on a careful choice of the sampling distribution, which can be estimated quickly. Then, we proposed a computationally efficient decoder to reconstruct k-bandlimited signals from their samples. We proved that it yields accurate reconstructions and that it is also stable to noise. Finally, we conducted several experiments to test these techniques. A journal paper has been published [17] accompanied by a toolbox for reproducible research (see Section 6.14). Other contributions from this year on the topic of graph signal processing include new subgraph-based filterbanks for graph signals [22], and new accelerated and robustified techniques for PCA on graphs [19], [20] (see also below our contributions in terms of new algorithms to obtain approximate Fast Graph Fourier Transforms [32], [53]).

Accelerated spectral clustering: We leveraged the proposed random sampling technique to propose a faster spectral clustering algorithm. Indeed, classical spectral clustering is based on the computation of the first k eigenvectors of the similarity matrix' Laplacian, whose computation cost, even for sparse matrices, becomes prohibitive for large datasets. We showed that we can estimate the spectral clustering distance matrix without computing these eigenvectors: by graph filtering random signals. Also, we took advantage of the stochasticity of these random vectors to estimate the number of clusters k. We compared our method to classical spectral clustering on synthetic data, and showed that it reaches equal performance while being faster by a factor at least two for large datasets of real data. Two conference papers have been presented, at ICASSP 2016 [39] and ICML 2016 [40] and a toolbox for reproducible research has been released (see Section 6.4).

7.1.2. An Alternative Framework for Sparse Representations: Sparse “Analysis” Models

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic, Clément Gaultier.

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 [7] [73], [87] successful applications of this approach to sound source localization, audio declipping and brain imaging have been developed in 2015 and 2016. In addition, new applications to audio denoising were also introduced this year.

Versatile cosparse regularization: Digging the groove of previous years' results (comparison of the performance of several cosparse recovery algorithms in the context of sound source localization [77], demonstration of its efficiency in situations where usual methods fail ([79], see paragraph 7.4.2), applicability to the hard declipping problem [78], application to EEG brain imaging [56]), a journal paper embedding the latest algorithms and results in sound source localization and brain source localization in a unified fashion was published this year [5]. This framework was also exploited to extend results on audio inpainting (see Section 7.3.2).

New results include experimental confirmation of robustness and versatility of the proposed scheme, and of its computational merits (convergence speed increasing with the amount of data). In a work presented in a workshop [44], we also proposed a multiscale strategy that aims at exploiting computational advantages of both sparse and cosparse regularization approaches, thanks to the simple yet effective all-zero initialization which the synthesis-based optimization can benefit from, while retaining the computational properties of the analysis-based approach for huge scale optimization problems arising in physics-driven settings.

Parametric operator learning for cosparse calibration: In many inverse problems, a key challenge is to cope with unknown physical parameters of the problem such as the speed of sound or the boundary impedance. In the sound source localization problem, we previously showed that the unknown speed of sound can be learned jointly in the process of cosparse recovery, under mild conditions [58], [81]. This year, we extended the formulation to the case of unknown boundary impedance, and showing that a similar biconvex formulation and optimization could solve this new problem efficiently (conference paper published in ICASSP 2016 [29], see also Section 7.3.3).

7.1.3. Algorithmic and Theoretical results on Computational Representation Learning

Participants: Rémi Gribonval, Luc Le Magoarou, Nicolas Bellot, Adrien Leman, Cassio Fraga Dantas, Igal Rozenberg.

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. Inspired by usual fast transforms, we proposed a general dictionary structure that allows cheaper manipulation, and an algorithm to learn such dictionaries together with their fast implementation. The principle and its application to image denoising appeared at ICASSP 2015 [84] and an application to speedup linear inverse problems was published at EUSIPCO 2015 [83]. A Matlab library has been released (see Section 6.6) to reproduce the experiments from the comprehensive journal paper published this year [16], which additionally includes theoretical results on the improved sample complexity of learning such dictionaries. Pioneering identifiability results have been obtained in the Ph.D. thesis of Luc Le Magoarou on this topic [85].

We further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms. A conference paper on this latter topic appeared in ICASSP 2016 [32], and a journal paper has been submitted [53] where 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.

A C++ software library has been developed (see Section 6.6) to release the resulting algorithms.

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. Characterizing and designing multi-carrier waveform systems with optimum PAPR

Participant: Rémi Gribonval.

Main collaboration: Marwa Chafii, Jacques Palicot, Carlos Bader (Equipe SCEE, Supelec, 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 [64], defended in October this year and co-supervised by R. Gribonval), 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, leading to a journal paper [67] and several conference communications [65], [66]. Our characterization of waveforms with optimum PAPR [68] has been published in a journal this year [14]. The work this year has concentrated on designing new adaptive multi-carrier waveform systems able to cope with frequency-selective channels while minimizing PAPR. This has given rise to a patent [49] and a journal paper is in preparation.

7.3. Emerging activities on Compressive Learning and Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

7.3.1. Phase Estimation in Multichannel Mixtures

Participants: Antoine Deleforge, Yann Traonmilin.

The problem of estimating source signals given an observed multichannel mixture is fundamentally ill-posed when the mixing matrix is unknown or when the number of sources is larger that the number of microphones. Hence, prior information on the desired source signals must be incorporated in order to tackle it. An important line of research in audio source separation over the past decade consists in using a model of the source signals' magnitudes in the short-time Fourier domain [8]. Such models can be inferred through, *e.g.*, non-negative matrix factorization [89] or deep neural networks [88]. Magnitudes estimates are often interpreted as instantaneous variances of Gaussian-process source signals, and are combined with Wiener filtering for source separation. In [50], we introduced a shift of this paradigm by considering the *Phase Unmixing* problem: how can one recover the instantaneous phases of complex mixed source signals when their magnitudes and mixing matrix are known? This problem was showed to be NP-hard, and three approaches were proposed to tackle it: a heuristic method, an alternate minimization method, and a convex relaxation into a semi-definite program. The last two approaches were showed to outperform the oracle multichannel Wiener filter in under-determined informed source separation tasks. The latter yielded best results, including the potential for exact source separation in under-determined settings.

7.3.2. Audio Inpainting and Denoising

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic.

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 (definition of the audio inpainting problem as a general framework for many audio processing tasks, application to the audio declipping or desaturation problem, formulation as a sparse recovery problem [55]), we proposed over the last two years a series of algorithms leveraging the competitive cosparse approach, which offers a very appealing trade-off between reconstruction performance and computational time [78], [80], [81]. The work on cosparse audio declipping which was awarded the Conexant best paper award at the LVA/ICA 2015 conference [80], together with the associated toolbox for reproducible research (see Section 6.8) draw the attention of a world leading company in professional audio signal processing, with which some transfer has been negotiated. In 2016, real-time implementation of the A-SPADE algorithm was obtained and demonstrated at various events (HCERES evaluation, TechnofÃ©rence # 18 « Nouvelles expériences son et vidéo », ...).

Current and future works deal with developing advanced (co)sparse decomposition for audio inpainting, including several forms of structured sparsity (*e.g.* temporal and multichannel joint-sparsity), dictionary learning for inpainting, and several applicative scenarios (declipping, denoising, time-frequency inpainting, joint source separation and declipping). In particular, we investigated the incorporation of the so-called "social" structure constraint [82] into problems regularized by a cosparse prior, including declipping and denoising. Publication of this work is currently under preparation.

7.3.3. Blind Calibration of Impedance and Geometry

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic.

Main collaborations: Laurent Daudet, Thibault Nowakowski, Julien de Rosny (Institut Langevin)

Last year, we also investigated extended inverse problem scenarios where a "lack of calibration" may occur, i.e., when some physical parameters are needed for reconstruction but apriori unknown: speed of sound, impedance at the boundaries of the domain where the studied phenomenon propagates, or even the shape of these boundaries. In a first approach, based on our physics-driven cosparse regularization of the sound source localization problem [5] (see section 7.1.2), we managed to preserve the sound source localization performance when the speed of sound is unknown, or, equally, when the impedance is unknown, provided the shape is and under some smoothness assumptions. Unlike the previous case (gain calibration), the arising problems are not convex but biconvex, and can be solved with proper biconvex formulation of ADMM algorithm. In a second approach based on eigenmode decomposition (limited to a 2D membrane), we showed that impedance learning with known shape, or shape learning with known impedance can be expressed as two facets of the same problem, and solved by the same approach, from a small number of measurements. Two papers presenting these two sets of results appeared at ICASSP 2016 [29], [37].

7.3.4. Sketching for Large-Scale Mixture Estimation

Participants: Rémi Gribonval, Nicolas Keriven.

Main collaborations: Patrick Perez (Technicolor R&I France) Anthony Bourrier (formerly Technicolor R&I France, then GIPSA-Lab)

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 [60] a framework aimed at fitting a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical 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 proposed algorithm provided good reconstruction and scaled to higher dimensions than previous probability mixture estimation algorithms, while consuming less memory in the case of voluminous datasets. It also provided a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum it is based on [63], [61], [62]. Last year, we consolidated our extensions to non-isotropic Gaussians, with new algorithms [76] and conducted large-scale experiments demonstrating its potential for speaker verification. A conference paper appeared at ICASSP 2016 [31] and a journal version has been submitted [52], accompanied by a toolbox for reproducible research (see Section 6.12).

This year the work concentrated on extending the approach beyond the case of Gaussian Mixture Estimation. First, we showed empirically that the algorithm can be adapted to sketch a training collection while still allowing to compute clusters. The approach, called "Compressive K-means", is described in a paper accepted at ICASSP 2017 [27]. Then, we expressed a theoretical framework for sketched learning, encompassing statistical learning guarantees as well as dimension reduction guarantees. The framework already covers compressive K-means as well as compressive Principal Component Analysis (PCA), and a conference paper has been submitted. A comprehensive journal paper is under preparation, and future work will include expliciting the impact of the proposed framework on a wider set of concrete learning problems.

7.4. Source Separation and Localization

Source separation, sparse representations, probabilistic model, source localization

Source separation is the task of retrieving the source signals underlying a multichannel mixture signal.

About a decade ago, state-of-the-art approaches consisted of representing the signals in the time-frequency domain and estimating the source coefficients by sparse decomposition in that basis. These approaches rely only on spatial cues, which are often not sufficient to discriminate the sources unambiguously. Over the last years, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [8], which generalizes a number of existing techniques including our former study on spectral GMMs [57]. 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, and it became the basis of a large number of works in the field, including our own. In the last years, improvements were obtained through the use of prior knowledge about the source spatial covariance matrices [71], [75], [74], knowledge on the source positions and room characteristics [72], or a better initialization of parameters thanks to specific source localization techniques [59].

This accumulated progress lead, in 2015, to two main achievements: a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released [92] 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 devoted to source separation and its applications [10]. This two achievements formed the basis of our work in 2016, exploring intensively the concrete use of these tools and principles in real-world scenarios, in particular within the voiceHome project (see Section 6.13).

7.4.1. Towards Real-world Separation and Remixing Applications

Participants: Nancy Bertin, Frédéric Bimbot, Ewen Camberlein, Romain Lebarbenchon.

In 2015, we began a new industrial collaboration, in the context of the VoiceHome project, aiming at another challenging real-world application: natural language dialog in home applications, such as control of domotic and multimedia devices. As a very noisy and reverberant environment, home is a particularly challenging target for source separation, used here as a pre-processing for speech recognition (and possibly with stronger interactions with voice activity detection or speaker identification tasks as well). In 2016, we publicly released a realistic corpus of room impulse responses and utterances recorded in real homes, and presented it during the Interspeech conference [28]. We also continued benchmarking and adapting 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 investigated new means to reduce the latency and computational burden of the currently available tools (low-resolution source separation preserving speech recognition improvement, automatic selection of the best microphones, joint localization and multichannel speech / non speech classification prior to any separation).

In november 2016, we started investigating a new application of source separation to sound respatialization from Higher Order Ambisonics (HOA) signals, in the context of free navigation in 3D audiovisual contents. This work is conducted in a collaboration with the IRT b<>Com, through the Ph.D. of Mohammed Hafsati (co-supervised by Nancy Bertin, RÃ©mi Gribonval).

7.4.2. Implicit Localization through Audio-based Control for Robotics

Participant: Nancy Bertin.

Main collaborations (audio-based control for robotics): Aly Magassouba and François Chaumette (Inria, EPI LAGADIC, France)

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. Common techniques, including beamforming, only provides the *direction of arrival* of the sound, estimated from the *Time Difference of Arrival* (*TDOA*) [59]. This year, we have particularly investigated alternative approaches, either where the explicit localization is not needed (audio-based control of a robot) or, on the contrary, where the exact location of the source is needed and/or TDOA is irrelevant (cosparse modeling of the acoustic field, see Section 7.1.2).

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 investigate over the last three years focuses on a sensor-based control approach. We proposed a new line of work, by considering the hearing sense as a direct and real-time input of a closed loop control scheme for a robotic task. Thus, and unlike most previous works, this approach does not necessitate any explicit source localization: instead of solving the localization problem, we focus on developing an innovative modeling based on sound features. To address this objective, we placed ourselves in the sensor-based control framework, especially visual servoing (VS) that has been widely studied in the past [69].

Last year, we established an analytical model linking the Interaural Time Difference (ITD) sound features and control input of the robot, defined and analyzed robotic homing tasks involving multiple sound sources, and validated the proposed approach by simulations and experiments with an actual robot [86]. This year, we consolidated these results and extended the range of applicative tasks [36] and obtained similar results (including theoretical and experimental) for the Interaural Level Difference (ILD), in combination with the absolute energy level [34]. Another set of experiments, presented during the IROS workshop [35] was successfully carried with a humanoid robot, notably without any measurement nor modeling of the robot's Head Relative Transfer Functions (HRTF). This work was mainly lead by Aly Magassouba, who defended his Ph.D. (co-supervised by Nancy Bertin and François Chaumette) in December 2016.

7.4.3. Emerging activities on Virtually-Supervised Sound Localization

Participants: Antoine Deleforge, Clément Gaultier, Saurabh Kataria.

Audio source localization consists in estimating the position of one or several sound sources given the signals received by a microphone array. It can be decomposed into two sub-tasks : (i) computing spatial auditory features from raw audio input and (ii) mapping these features to the desired spatial information.

Extracting spatial features from raw audio input: The most commonly used features in binaural (two microphones) sound source localization are frequency-dependent phase and level differences between the two microphones. To handle the presence of noise, several sources, or reverberation, most existing methods rely on some kind of aggregation of these features in the time-frequency plane, often in a heuristic way. In [25], we introduced the rectified binaural ratio as a new spatial feature. We showed that for Gaussian point-source signals corrupted by stationary Gaussian noise, this ratio follows a complex t-distribution with explicit parameters. This new formulation provides a principled, statistically sound and efficient method to aggregate these features in the presence of noise. Experiments notably showed the higher robustness of these features compared to traditional ones, in the task of localizing heavily corrupted speech signals.

Mapping features to spatial information: Existing methods to map auditory features to spatial properties divide into two categories. *Physics-driven* methods attempt to estimate an explicit mapping based on an approximate physical model of sound propagation in the considered system. *Data-driven* methods bypass the use of a physical model by learning the mapping from a training set, obtained by manually annotating features extracted from real data. We proposed a new paradigm that aims at making the best of physics-driven and data-driven approaches, referred to as *virtually-supervised acoustic space mapping* [26], [51]. 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 [51]. 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 [26].

7.5. Music Content Processing and Information Retrieval

Music structure, music language modeling, System & Contrast model

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 [24], in particular the System & Contrast model [1].

7.5.1. Tensor-based Representation of Sectional Units in Music

Participants: Corentin Guichaoua, Frédéric Bimbot.

Following Kolmogorov'ÂÂs 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 is currently developed and tested for the grouping of chord sequences into sectional units for pop music songs, with very encouraging segmentation results on pop songs.

7.5.2. Minimal Transport Graphs for the Modeling of Chord Progressions

Participants: Corentin Louboutin, Frédéric Bimbot.

In this work, we model relations between chords by minimal transport and we investigate different types of dependencies within chord sequences [33]. For this purpose we use the ÂÂSystem & Contrast (S&C) model [1], designed for the description of music sectional units, to infer non-sequential structures called chord progression graphs (CPG).

Minimal transport is defined as the shortest displacement of notes, in semitones, between a pair of chords. The paper [33] present three algorithms to find CPGs for chords sequences: one is sequential, and two others are based on the S& C model. The three methods are compared using the perplexity as an efficiency measure.

The experiments on a corpus of 45 segments taken from songs of multiple genres indicate that optimization processes based on the S&C model outperform the sequential model with a decrease in perplexity over 1.0.

7.5.3. Regularity Constraints for the Fusion of Music Structure Segmentation System

Participant: Frédéric Bimbot.

Main collaborations Gabriel Sargent (EPI LinkMedia, Rennes, France)

Music structure estimation has recently emerged as a central topic within the field of Music Information Retrieval. Indeed, as music is a highly structured information stream, knowledge of how a music piece is organized represents a key challenge to enhance the management and exploitation of large music collections.

Former work carried out in our group [9] has illustrated the benefits that can be expected from a regularity constraint on the structural segmentation of popular music pieces : a constraint which favors structural segments of comparable size provides a better conditioning of the boundary estimation process.

As a further investigation, we have explored the benefits of the regularity constraint as an efficient way for combining the outputs of a selection of systems presented at MIREX between 2010 and 2015. These experiments have yielded a level of performance which is competitive to that of the state-of-the-art on the ”MIREX10" dataset (100 J-Pop songs from the RWC database) [18].

SEMAGRAMME Project-Team

6. New Results

6.1. Syntax-semantics interface

Participants: Philippe de Groote, Sylvain Pogodalla.

6.1.1. Lambek categorial grammar as abstract categorial grammars

Abstract Categorial Grammars (ACG, for short) differ from classical categorial grammars in an essential way: the ACG type system is based on a commutative logic (namely, the implicative fragment of multiplicative linear logic). For this reason, it has been argued that the way of encoding wh-extraction in an ACG corresponds to an uncontrolled form of extraction, which results in syntactic overgeneration. In particular, an ACG could not accomodate left and right peripheral extractions like a Lambek categorial grammar (LG, for short) does. In order to challenge this claim, we have shown how LG may be encoded as ACG [14].

6.1.2. Lexical Semantics

The interpretation of natural language utterances relies on two complementary elements of natural language modeling. On the one hand, the description of the combinatorics of natural language expresses how elementary units, or *lexical units* (typically the word), combine in order to build more complex elements, such as sentences or discourses. On the other hand, the description of these elementary units specifies how they contribute to the meaning of the whole by their *lexical meaning*. This specification should also take into account how the different parts of the lexical meanings combine during the *composition* process and how they relate to their underlying meaning concepts. For instance, the verbs *buy* and *sell* should refer to a common conceptual representation. However, their syntactic arguments (e.g., the subject) play a different (semantic) role with respect to the *transaction* concept that they share.

The modeling of these concepts and how they relate to each other gave rise to Frames Semantics as a representation format of conceptual and lexical knowledge [37], [30], [25], [45]. Frames consists of directed graphs where nodes correspond to entities (individuals, events, ...) and edges correspond to (functional or non-functional) relations between these entities. Providing a fine-grained representation of the internal concept structure allows both for a *decomposition* of the lexical meaning and for a precise description of the sub-structural interactions in the semantic composition process [44].

Following up on our previous work based on Hybrid Logic (HL) [26], [24] on linking Frames and truth-logical semantics, with a specific focus on explicit quantification over entities or events that are lexically triggered, we extended our model to the interaction between bounded events and *for*-adverbials. This interaction turns bounded events (*John biked to the office*) to iterated events (*John biked to the office for three months*), when the bounded events themselves result from coercing a progression (*John biked*) by addition of a prepositional phrase (*to the office*). We also proposed a modeling taking into account the respective scopes of the quantifiers induced by *for*-adverbials (over events) and quantification introduced by indefinites (over entities) [17]. Finally, we used the flexibility of the approach to model semantic coercion as induced by verbs such as *read* that can syntactically have an entity as argument (*John began a book*) while it semantically relates to an event (e.g., *reading*, *writing*, etc.) [21].

6.2. Discourse dynamics

Participants: Philippe de Groote, Sylvain Pogodalla, Maxime Amblard, Jirka Maršík, Aleksandre Maskharashvili.

6.2.1. Effects and Handlers in Natural Language

In formal semantics, logical meanings are assigned to natural language utterances. This process is guided by the principle of compositionality: the meaning of an expression is a function of the meanings of its parts. These functions are often formalized using the λ-calculus. However, there are areas of language which challenge the notion of compositionality, e.g. anaphoric pronouns or presupposition triggers. These force one to either abandon compositionality or adjust the structure of meanings. In the first case, meanings are derived by processes that no longer correspond to pure mathematical functions but rather to context-sensitive procedures, much like the functions of a programming language that manipulate their context with side effects. In the second case, when the structure of meanings is adjusted, the new meanings tend to be instances of the same mathematical structure, the monad. Monads themselves being widely used in functional programming to encode side effects, the common theme that emerges in both approaches is the introduction of side effects. Furthermore, different problems in semantics lead to different theories which are challenging to unite. We claim that by looking at these theories as theories of side effects, we can reuse results from programming language research to combine them.

Our work extends the λ-calculus with a monad of computations. The monad implements effects and handlers, a recent technique in the study of programming language side effects. We have proven some of the fundamental properties of our extended calculus: subject reduction, confluence and termination. We have then demonstrated how to use our calculus to implement treatments of several linguistic phenomena: deixis, quantification, conventional implicature, anaphora and presupposition.

6.2.2. Discourse Modeling with Abstract Categorial Grammars

We have studied several TAG-based grammatical formalisms for discourse analysis (D-LTAG [38], G-TAG [34], and D-STAG [33]), and we have proposed an ACG encodings of them. G-TAG is a formalism introduced for generating natural language texts out of conceptual (semantic) representation inputs. D-STAG is a synchronous formalism for modeling the syntax-semantics interface for discourse. It was introduced for discourse analysis (parsing). The ACG encodings of G-TAG and D-STAG shed light on the problem of clause-medial connectives that TAG-based formalisms do not account for. To deal with a discourse that contains clause-medial connectives, D-LTAG, G-TAG, and D-STAG, all make use of an extra grammatical step. In contrast, the ACG encodings of G-TAG and D-STAG offer a purely grammatical approach to discourse connectives occupying clause-medial positions. The method we propose is a generic one and can serve as a solution for encoding clause-medial connectives with the formalisms based on TAGs. The ACG encodings of G-TAG and D-STAG that we propose are second-order. Importantly, the class of second-order ACGs consists of intrinsically reversible grammars. Grammars of this class use the same polynomial algorithm to build parse structures both for strings and for logical formulas. Thus, second-order ACGs can be used both for parsing and generation. Therefore, the problems of parsing and generation with the ACG encodings of G-TAG and D-STAG are of polynomial complexity.

6.3. Common basic resources

Participants: Bruno Guillaume, Guy Perrier, Nicolas Lefebvre.

6.3.1. Crowdsourcing Complex Language Resources

This work [15] presents the results we obtained on a complex annotation task (that of dependency syntax) using a specifically designed Game with a Purpose, ZombiLingo. [0] The design of the game has to deal with the fact that the task is complex and does not directly rely on human intuition. We show that with suitable mechanisms (decomposition of the task, training of the players and regular control of the annotation quality during the game), it is possible to obtain annotations whose quality is significantly higher than that obtainable with a parser, provided that enough players participate. The source code of the game and the resulting annotated corpora (for French) are freely available.

[0] See: http://zombilingo.org/.

6.3.2. Universal Dependencies

We participated to development of new versions of the French part of the Universal Dependencies project (http://universaldependencies.org/).

The version 1.3 [52] was released in May. In this version, the lemmatization and the morphological annotation were added automatically when possible and with manual verification for ambiguous occurrences.

The version 1.4 [51] was released in November. This version contains a large number of annotation corrections. The Grew software was used to explore, to check consistency and to correct systematically the data. For instance, all copula annotations where checked manually.

CHROMA Team

7. New Results

7.1. Bayesian Perception

Participants: Christian Laugier, Lukas Rummelhard, Amaury Nègre [Gipsa Lab since June 2016], Jean-Alix David, Julia Chartre, Jerome Lussereau, Tiana Rakotovao, Nicolas Turro [SED], Jean-François Cuniberto [SED], Diego Puschini [CEA DACLE], Julien Mottin [CEA DACLE].

7.1.1. Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT)

Participants: Lukas Rummelhard, Amaury Nègre, Christian Laugier.

The research work on *Bayesian Perception* has been done as a continuation and an extension of some previous research results obtained in the scope of the former Inria team-project e-Motion and of the more recent developments done in 2015 in the scope of the Chroma team. This work exploits the *Bayesian Occupancy Filter (BOF)* paradigm [42], developed and patented by the team several years ago [0]. It also extends the more recent concept of *Hybrid Sampling BOF (HSBOF)* [76], whose purpose was to adapt the concept to highly dynamic scenes and to analyze the scene through a static-dynamic duality. In this new approach, the static part is represented using an occupancy grid structure, and the dynamic part (motion field) is modeled using moving particles. The *HSBOF* software has been implemented and tested on our experimental platforms (equipped Toyota Lexus and Renault Zoe) in 2014 and 2015; it has also been implemented in 2015 on the experimental autonomous car of Toyota Motor Europe in Brussels.

The objective of the research work performed in the period 2015-16 was to overcome some of the shortcomings of the initial *HSBOF* approach [0] , and to obtain a better understanding of the observed dynamic scenes through the introduction of an additional object level into the model. The new framework, whose development has been continued in 2016, is called *Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT)* [84]. The whole CMCDOT framework and its results are presented and explained on a video posted on Youtube [0]. This work has mainly been performed in the scope of the project *Perfect* of IRT Nanoelec [0] (financially supported by the French ANR agency [0]), and also used in the scope of our long-term collaboration with Toyota.

In 2016, most of the efforts have been focused on the optimization of the implementation of our grid-based Bayesian filtering CMCDOT framework. Since the beginning of the development of this framework, we have chosen to construct models and algorithms specially designed to attain real-time performances on embedded devices, through a massively parallelization of the involved processes. The whole system have been implemented and scrupulously optimized in Cuda, in order to fully benefit from the Nvidia GPUs and technologies. Starting from the use of the Titan X and GTX980 GPUs (the hardware used in our computers and experimental platforms), we have successfully adapted and transferred our whole real-time perception chain on Nvidia dedicated-to-automotive cards Jetson K1 and X1 [0]. A specific optimization has been performed in term of data access and processing, allowing us to obtain real-time results when processing the data from the 8 lidar layers generated by our IBEO sensors, by using a grid containing 1400x600 cells and 65536 dynamic particles (for motion estimation). The observation grid generation and fusion (representing the input of the CMCDOT) is made in 17ms on Jetson K1 and only in 0.7ms on Jetson X1; a CMCDOT filtering update is performed in 70ms on Jetson K1 and only in 17ms on Jetson X1.

[0]The *Bayesian programming formalism* developed in e-Motion, pioneered (together with the contemporary work of Thrun, Burgards and Fox [94]) a systematic effort to formalize robotics problems under Probability theory –an approach that is now pervasive in Robotics.

[0]In the current implementation of the HSBOF algorithm, many particles are still allocated to irrelevant areas, since no specific representation models are associated to dataless areas. Moreover, if the filtered low level representation can directly be used for various applications (for example mapping process, short-term collision risk assessment [47], [85], etc), the retrospective object level analysis by dynamic grid segmentation can be computationally expensive and subjected to some data association errors.

[0]https://www.youtube.com/watch?v=uwIrk1TLFiM

[0]Nanoelec Technological Research Institute (Institut de Recherche Technologique Nanoelec)

[0]National Research Agency (Agence Nationale de la recherche)

[0]These new Nvidia devices are more suited for embedded applications, in term of power consumption and dimensions.

Figure 5. Jetson X1 card, Nvidia device dedicated to automotive applications

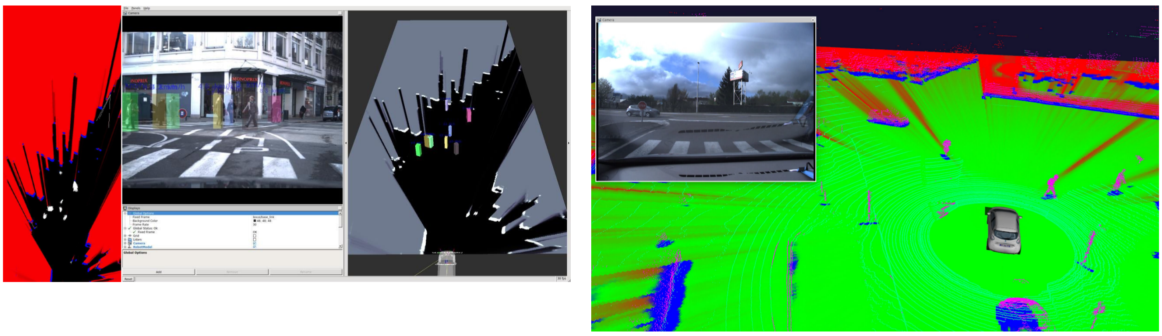


Figure 6. a) CMCDOT results : filtered occupancy grids, enhanced with motion estimations (vectors) and object detection (colored boxes) b) Example of an occupancy grid generated using the classified point cloud and the ground model

7.1.2. A new sensor model for 3D sensors, by Ground Estimation, Data segmentation and adapted Occupancy Grid construction

Participants: Lukas Rummelhard, Amaury Nègre, Anshul Paigwar, Christian Laugier.

As a starting point for the Bayesian perception framework embedded on the vehicles and on the perception boxes, the system generates instantaneous spatial occupancy grids, by interpreting the point clouds generated by the sensors (sensor model). With planar sensors, placed at the level of the wanted occupancy grid, such as the IBEO Lidar on the vehicles or the Hokuyo Lidar on the first developed perception box, a classic sensor model can be used: before the laser impact the space is considered as empty, occupied at the impact point and undefined after the impact. In our previous approach, the angular differences between the 4 laser layers of our IBEO Lidars was taken into account by introducing a *confidence factor* in the data, reducing in this way the effect of the impacts too close to the ground. In this approach the ground is assumed to be flat and the confidence factor is calculated geometrically. Then, given the orientation of these sensors and the environments traversed, such a model was quite satisfactory.

However, this traditional sensor model has to be adapted when using Velodyne or Quanergy sensors mounted on the top of the vehicle and providing dense 3D data with a high horizontal and vertical resolution. Indeed, in this case the laser layers are capable of depicting an obstacle from above, and consequently an impact at a given distance does not certify any more a free area until the impact. Moreover, many impact points are located on the ground and have to be appropriately modeled in order to systematically avoid deceptive obstacle detection. Then, the previous flat-ground assumption doesn't hold anymore, since the actual ground shape is integrated into the data and the correct segmentation of obstacle becomes critical in the process. This is why we have developed the new *Ground Estimator* approach.

The aim of the method is, upstream from the Bayesian filtering step of our current perception system (CMCDOT), to first dynamically *estimate the ground elevation*, to exploit this information for making a *relevant data classification* between actual obstacle impacts and ground impacts, and finally to generate the *relevant occupancy grid using this classified 3D point cloud* (sensor model). The developed method is based on a recursive spatial and temporal filtering of a Bayesian network of elevation nodes, constantly re-estimated and re-evaluated with respect to data and spatial continuity. The construction of the occupancy grid is based, on the one hand, on the location of the laser impacts, and on the other hand on the shape of the ground and the height at which the lasers pass through the different portions of the space.

The approach has been first successfully tested and validated with dense Lidar sensors (Velodyne and Quanergy). The use of the enhanced sensor model is also currently tested with sparser sensors, with the objective to increase their robustness. The obtained results show promising perspectives, offering a robust and efficient ground representation, data segmentation and relevant occupancy grid, and also offering quality inputs for the next steps of the perception framework. A journal paper and a patent are under preparation.

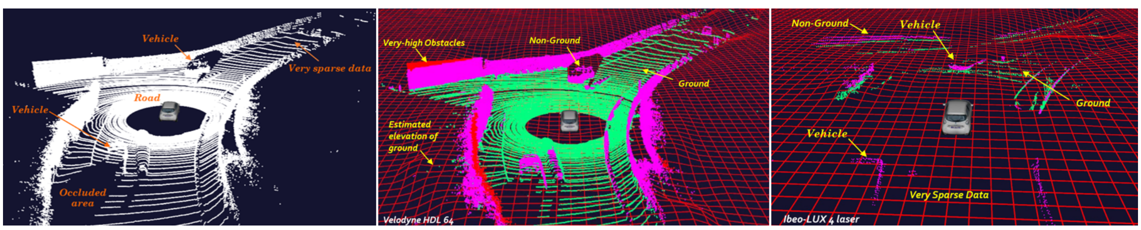


Figure 7. (a) Typical 3D point cloud generated by Velodyne LiDAR, (b) Point cloud segmentation between ground (green points) and non-ground (purple points), and estimated average elevation of the terrain (red grid) (c) Point Cloud Segmentation on 4-Ibeo Lux LiDAR data and estimated elevation of terrain.

7.1.3. Dense & Robust outdoor perception for autonomous vehicles

Participants: Victor Romero-Cano, Christian Laugier.

Robust perception plays a crucial role in the development of autonomous vehicles. While perception in normal and constant environmental conditions has reached a plateau, robust perception in changing and challenging environments has become an active research topic, particularly due to the safety concerns raised by the introduction of self-driving cars to public streets. In collaboration with Toyota Motors Europe and starting in April 2016 we have developed techniques that tackle the robust-perception problem by combining multiple complementary sensor modalities.

Our techniques, similar to those presented in [78], [91] explore the complementary relationships between passive and active sensors at the pixel level. Low-level sensor fusion allows for an effective use of raw data in the fusion process and encourages the development of recognition systems that work directly on multi-modal data rather than higher level estimates. During the last nine months we have developed low-level sensor fusion approaches that, differently from most of the related literature, do not have fixed requirements regarding coverage or density of the active sensors. This provides a competitive advantage due to the elevated costs of dense range sensors such as Velodyne LIDARs.

Our framework outputs a new image-like data representation where each pixel contains not only colour but also other low level features such as depth and regions of interest where generic objects are likely to be. Our approach is generic so it allows for the integration of data coming from any active sensor into the image space. Additionally, it does not aim at tackling the object detection problem directly but it proposes a multi-modal-data representation from which object detection methods may benefit. For evaluation purposes we have tackled the concrete problem of fusing color images and sparse lidar returns, however, as explained before, the framework is amenable for the inclusion of any other range-sensor modality. The framework creates *XDimages* by extrapolating range measurements across the image space in a two-stage procedure. The first stage considers locally homogeneous areas given by a super-pixel segmentation while the second one further expands depth values by performing self-supervised segmentation of areas seeded by the range sensor. The framework's pipeline is illustrated in Figure 8 .

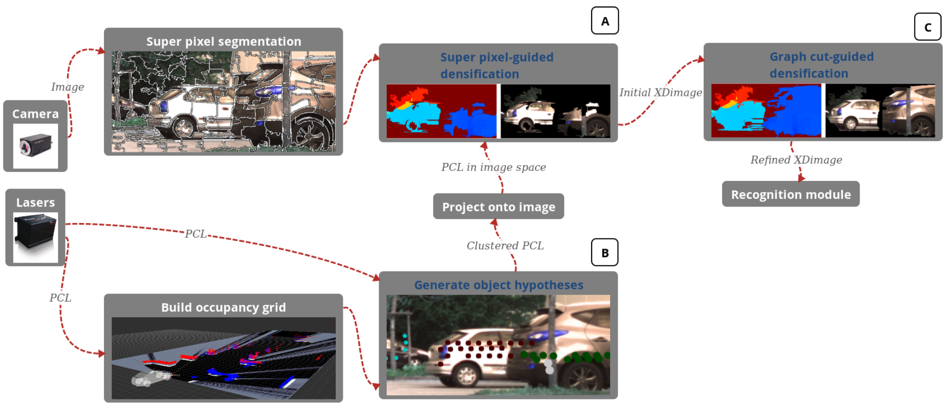


Figure 8. The XDvision framework.

We have named an instance of our data structure an *XDimage*. It corresponds to an augmented camera image where individual pixels contain both appearance and geometric information. The first and more challenging problem to be solved in order to build XDimages is that of densifying sparse point cloud data provided by active range sensors. In our approach we extrapolated depth information using a two-steps procedure as follows:

1. Extend depth values projected onto individual pixels to neighbouring pixels that have similar appearance.
2. Obtain geometry-based object hypothesis.
3. For each geometry-based object hypothesis, extrapolate range measurements in order to account for entire objects.

The results of this work have resulted in a patent application [82] and a paper submission to ICRA 2017 [83].

7.1.4. Integration of Bayesian Perception System on Embedded Platforms

Participants: Tiana Rakotovao, Christian Laugier, Diego Puschini [CEA DACLE], Julien Mottin [CEA DACLE].

Perception is a primary task for an autonomous car where safety is of utmost importance. A perception system builds a model of the driving environment by fusing measurements from multiple perceptual sensors including LIDARs, radars, vision sensors, etc. The fusion based on occupancy grids builds a probabilistic environment model by taking into account sensor uncertainties. Our objective is to integrate the computation of occupancy grids into embedded low-cost and low-power platforms. Occupancy Grids perform though intensive probability calculus that can be hardly processed in real-time on embedded hardware.

As a solution, we introduced a new sensor fusion framework called *Integer Occupancy Grid* [80]. Integer Occupancy Grids rely on a proven mathematical foundation that enables to process probabilistic fusion through simple addition of integers. The hardware/software integration of integer occupancy grids is safe and reliable. The involved numerical errors are bounded and parameterized by the user. Integer Occupancy Grids enable a real-time computation of multi-sensor fusion on embedded low-cost and low-power processing platforms dedicated for automotive applications. This research work has been conducted in the scope of the PhD thesis of Tiana Rakotovao, which will be defended in February 2017.

Figure 9. Fusion of three front LIDARs and one rear LIDAR on the ZOE platform

Experiences showed that Integer Occupancy Grids enable the real-time fusion of the four ibeo LUX LIDARs mounted on the ZOE experimental platform of IRT Nano-Elec [79]. The LIDARs produces about 80,000 points at 25Hz. These points are fused in real-time through a hardware/software integration of the Integer Occupancy Grid framework on an embedded CPU based on ARM A9@1GHz. The platform respects the low-cost and low-power constraints of processing hardware used for automotive. The fusion produces an occupancy grid at more than 25 Hz as illustrated on figure 9 .

7.1.5. Embedded and Distributed Perception

Participants: Christian Laugier, Julia Chartes, Amaury Nègre, Lukas Rummelhard, Jean-Alix David, Jerome Lussereau, Nicolas Turro [SED], Jean-Francçois Cuniberto [SED].

7.1.5.1. Embedded Perception in an Experimental Vehicle Renault ZOE

In the scope of the *Perfect* project of the IRT nanoelec, we have started to build an experimental platform using a Renault Zoe equipped with several types of sensors (see 2014 and 2015 annual activity reports). The platform includes multiple sensors, an embedded perception system based on the CMCDOT, and a collision risk component, figure 10 (a) illustrates.

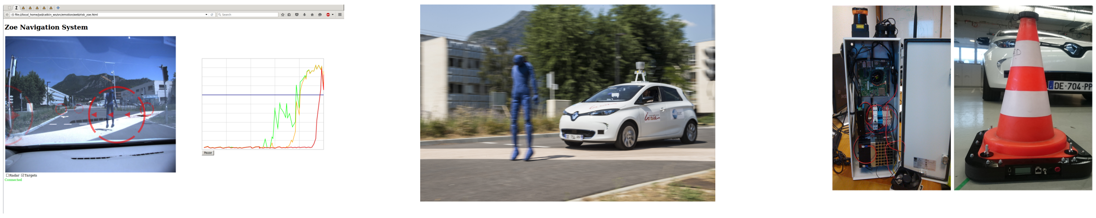


Figure 10. (a) Display of the HMI (b) Collision simulation with a mannequin (c) On left: picture of the smartbox, on right: picture of the cone.

In 2016, we have continued to develop and to improve the platform using the latest version of the CMCDOT, some optimized perception and localization components, and new V2X communication functions for distributed perception.

In particular, we have developed an improved the localization function using maps and V2X communication devices. We also developed a new embedded component for sharing data between the infrastructure perception boxes and the vehicle; this component is based on the use of V2X communication and GPS time synchronization. This is a first step towards a fully distributed perception system. The development of this system will be continued in 2017 (see next section).

During the year 2016, experiments have been pushed forward on testing the perception algorithms, the collision risk alert and the localization components using a fabric mannequin as shown on figure 10 (b). The mannequin has been motorized for easier and more realistic tests and demos. More details are given in the team publications at MCG 2016 [29] and at GTC Europe 2016. The work of the team is also explained on youtube videos "Irt Nanoelec Perfect Project" [55] and for the technical side "Bayesian Embedded Perception" [54].

New experiments have also been performed on some road intersections and highways, in order to collect new data on driver's behaviors. These experiments have been conducted on mountain roads with changing slopes and on highway (to study the lane changing behaviors). They have been performed in the scope of our cooperation with Renault and with Toyota. The way these experimental data have been used is described in the section “Situation Awareness”. More recently, we have also started to work on the development of the automatic driving controls on the Zoe vehicle. For that purpose, we have recently signed a cooperation agreement with Ecole Centrale de Nantes. The basic functions for automatic driving will be implemented on the Zoe at the beginning of 2017. For that purpose, a physical model of the Zoe is currently in construction under ROS Gazebo simulator. This should allow us to implement and to test the required control laws.

7.1.5.2. Distributed Perception

In 2015, we have developed a first *Connected Perception Box* including a GPS, a V2X communication device, a cheap Lidar sensor, and an Nvidia Tegra K1 board. The box was powered using a battery, and the objective was to reduce as far as possible the required energy consumption. Within the box, the perception process is performed using the CMCDOT algorithm. In 2016, we have continued to develop this concept of distributed perception. We have developed a second generation of the perception box, using a Quanergy M8 360° Lidar, a TX1 Nvidia Tegra board, an ITRI V2X communication device and the last version of the CMCDOT system.

This new box is more efficient and powerful than the previous one. It allows the real-time exchange of objects positions and velocities, through a V2X communication between the perception box and the connected vehicle. This leads to the extension of the vehicle perception area to some hidden areas, and to generate some alerts in case of a high collision risk, cf. fig. 11 . In this approach, time synchronization has been performed using GPS time and NTP protocol.

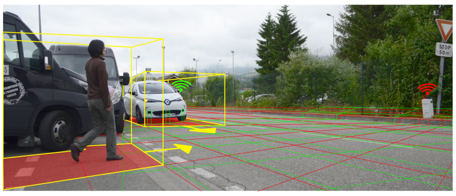


Figure 11. (a) Shared perception between car and perception box

7.1.5.3. Public demonstrations and Technological Transfer

2016 has been a year with many scientific events and public demos. Several public demonstrations of our experimental vehicle have performed, some of them in presence of local government officials during a GIANT show at CEA.

The collaboration with Nvidia on Embedded Perception for autonomous driving has been extended to 2017, and the "GPU research center" label has been renewed.

Toyota Motor Europe (TME) is strongly interested in the CMCDOT technology, and Inria is currently negotiating with them the conditions of a first licence for R&D purpose. A first implementation of the executable code of CMCDOT has successfully been implemented on the TME experimental vehicle in Brussels. We are currently discussing with TME an extension of the license to several other experimental vehicles located in some other places in the world.

At the end of 2016, we also started to transfer the CMCDOT technology to two industrial companies in the fields of industrial mobile robots and automatic shuttles. Confidential contracts for the joint development of proofs of concepts are under signing. The work will be performed at the beginning of 2017.

7.2. Situation Awareness

Participants: Christian Laugier, Olivier Simonin, Jilles Dibangoye, Alejandro Dizan Vasquez Govea [Apple since January 2016], Stephanie Lefevre [Mercedes-Benz North America], David Sierra-Gonzalez, Mathieu Barbier, Victor Romero-Cano.

7.2.1. Framework for Motion Prediction and Collision Risk Assessment

Participants: Christian Laugier, Alejandro Dizan Vasquez Govea [Apple since January 2016], Stephanie Lefevre [Mercedes-Benz North America], Lukas Rummelhard.

For several years, the challenging scientific problem of Motion Prediction, Risk Assessment and Decision-Making in open and dynamic environments has been one of our main research topics (see activity reports of the former e-Motion Inria team-project and Chroma team 2015 activity report).

Throughout 2016, we have continued this line of work by developing and experimentally testing new frameworks for Motion Prediction and Collision Risk Assessment in complex dynamic scenes involving multiple moving agents having various behaviors. This work has been conducted in the scope of three main scenarios: Short-term prediction in crowded urban environments (see approach and results in sections 7.1.1 and 7.1.5), Autonomous driving in highway environments (see section 7.2.2), and Safe Intersection crossing.

The main underlying concepts of the developed framework have recently been published in the second edition of the Handbook of Robotics [31]. They have also been presented into a Mooc course on “Autonomous Mobiles Robots and Vehicles”, dedicated to graduate and undergraduate students and to engineers in Robotics [57]. This Mooc has been published twice in 2015 and in 2016.

The recent results have been published at ICRA 2016 [27] and also presented by C. Laugier in several invited talks : Asprom2016 [16], ICIT2016 [14], CUHK2016 [15], GTC-Europe2016 [24] and ARSO2016 [17].

Another contribution relies in the implementation of some the proposed models on two experimental vehicles (Lexus and Zoe experimental platforms). As it has been mentioned in the section 7.1.5 , several experiments on short-term collision risk assessment have successfully been conducted with these platforms in 2015 and 2016 (c.f. [84], [67]).

This work will be continued in 2017, in the scope of our ongoing collaborative projects with Toyota, Renault and IRT nanoelec. It will also be used as a support for the planned technological transfers with two industrial companies in the fields of industrial mobile robots and automatic shuttles (see section 7.1.5).

7.2.2. Planning-based motion prediction for collision risk estimation in autonomous driving scenarios

Participants: David Sierra-Gonzalez, Christian Laugier, Jilles Dibangoye, Alejandro Dizan Vasquez Govea [Apple since January 2016].

The objective is to develop a collision risk estimation system capable of reliably finding the risk of collision associated to the different feasible trajectories of the ego-vehicle. This research work is done in the scope of the Inria-Toyota long-term cooperation and of the PhD thesis work of David Sierra-Gonzáles.

One key factor, and probably the biggest challenge in order to produce robust collision risk estimation in traffic scenes, is the motion prediction of the dynamic obstacles (i.e. the other drivers for highway scenarios). The difficulty stems from the fact that human behavior is determined by a complex set of interdependent factors, which are very hard to model (e.g. intentions, perception, emotions). As a consequence, most existing systems are based on simple short-term motion models assuming constant velocity or acceleration.

We opt here for a planning-based approach, which assumes that drivers instinctively act as to minimize a cost (or equivalently, maximize a reward). This cost function encodes the preferences of the driver to, for instance, keep a minimum distance with the vehicle in front, drive in the right lane in the highway, or respect the speed limits. By using Inverse Reinforcement Learning (IRL) algorithms, we can obtain such cost function directly from expert demonstrations (i.e. simply observing how people drive).

Two well-known IRL algorithms [35], [101] have been implemented and used to obtain driver models from human demonstrations. The algorithms have been adapted to work with simulated demonstrations obtained on a highway simulator, and with real-world data from the Lexus and Renault Zoe platforms. Figure 12 .a shows a slice of one such cost function in the context of a real highway scene.

A novel framework has been developed in order to exploit the predictive potential of these models for the task of highway scene prediction [26]. The ability of these dynamic models to capture the risk-aversive behavior of drivers leads to an interaction-aware prediction. In contrast to other state-of-the-art interaction-aware approaches [59], the complexity of our prediction framework does not grow exponentially in the number of vehicles in the scene, but only quadratically. Figure 12 .b shows the prediction produced by our framework in two prototypical simulated highway scenarios. The figure shows the result of summing up across the occupancy distributions over a discretization of the road for all the vehicles in the scene, at different timesteps (note that the result is no longer a probability distribution, but it is convenient to visualize the prediction).

This framework has been patented by Inria and Toyota Motor Europe in October 2016.

7.2.3. Functional space representation for automated road intersection crossing

Participants: Mathieu Barbier, Christian Laugier, Olivier Simonin.

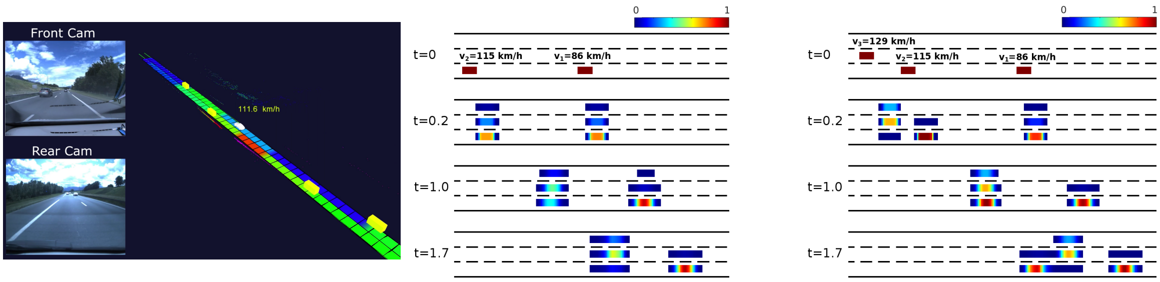


Figure 12. a) Slice of a cost function obtained from real demonstrated data superimposed on the road on a highway scene. Red indicates high cost, green intermediate, and blue low. b) and c) Prediction of two prototypical simulated traffic scenes with the framework from [26]. We show the predicted occupancy probabilities over a discretization of the road for different timesteps.

The objective is to develop a framework for modeling road intersections using relevant functional areas, which can be exploited by an autonomous vehicle for safely crossing the intersection. These functional areas try to capture various characteristics such as crossing, merging or approaching areas, the car dynamics when moving in such areas, or the related uncertainty. We made the assumption that such a functional space representation can be stored in a map and can be constructed using observed trajectories. This work is performed in the scope of the PhD thesis of Mathieu Barbier, which is supported by a CIFRE fellowship with Renault.

In a preliminary work done with map by Renault, it has been observed that the information stored in a map does not fully match with the real motions executed by cars within a given intersection. The differences between the stored and the real path might be important. This difference is not due to error during the map creation, but rather to other constraints related to the driving action (e.g. visibility, dynamics). Such a difference leads to serious difficulties at the level of the autonomous driving decision-making process.

Constructing a functional model, requires to first analyze the topological and dynamics structure of an intersection, and in a second step to imagine how it would be possible to extract such type of information from maps and observed trajectories. Two main structures seem to emerge from this study:

- Merging and Crossing points, areas where two cars are the most likely to collide.
- Approaching areas, areas where drivers are most likely to show constant driving behaviors.

In order to learn motions patterns of multiple cars, we have chosen to train Gaussian processes [81] [93] using simulated data set generated using the SCANER™system. The resulting distribution is segmented using different threshold, in order to find approaching areas and to combine pairs of such areas for constructing overlapping areas. The correlation between this discretization and both real and simulated velocity profiles has been shown by the experimental results, see Fig. 13 . The approach has been published at IEEE ITSC 2016 [18].

We recently started to make use of a Random Forest classifier to connect features of trajectories with an intended maneuver (stop, pass, yield) and to take advantage of the discretization. This research work will be continued in 2017.

7.3. Robust state estimation (Sensor fusion)

7.3.1. Visual-inertial structure from motion: observability properties and state estimation

Participant: Agostino Martinelli.

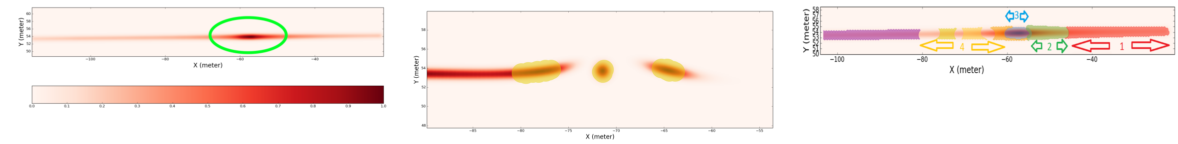


Figure 13. Different step of the framework to discretize the space : a) Map created with prediction from set of GPs, the highlighted area has a high mean probability b) Segmentation of crossing and merging areas, in red the probability of two cars being in the same position and in yellow where this probability is higher than the threshold c) Discretization of approaching area

This research is the follow up of our investigations carried out during the last three years. The main results obtained this year regard the following three topics:

1. Exploitation of the closed form solution introduced in [70] in the framework of Micro Aerial Vehicle (MAV) navigation;
2. Introduction of a new method for simultaneous localization and Gyroscope calibration;
3. Analytic solution of the Unknown Input Observability problem (UIO problem) in the nonlinear case.

Regarding the first two topics, we successfully implemented a new method for MAV localization and mapping, on the aerial vehicles available at the Vision and Perception lab at the university of Zurich [0]. This method is based on our previous closed form solution recently introduced in [70]. The practical advantage of this solution is that it is able to determine several physical quantities (e.g, speed, orientation, absolute scale) by only using the measurements provided by a monocular camera and an Inertial Measurement Unit (IMU) during a short interval of time (about 3 seconds). In other words, an initialization is not requested to determine the aforementioned physical quantities. This fact has a fundamental importance in robotics and it is novel with respect to all the state of the art approaches for visual-inertial sensor fusion, which use filter-based or optimization-based algorithms. Due to the nonlinearity of the system, a poor initialization can have a dramatic impact on the performance of these estimation methods.

Finally, by further studying the impact of noisy sensors on the performance of the closed-form solution introduced in [70], we found that this performance is very sensitive to the gyroscope bias. For, we developed a powerful and simple optimization approach to remove this bias. This method has been tested in collaboration with the vision and perception team in Zurich (in the framework of the ANR-VIMAD) and published on the IEEE Robotics and Automation Letters [12]. Additionally, these results have been presented at the International Conference on Robotics and Automation [21].

Regarding the third topic, we still considered the problem of deriving the observability properties of the visual-inertial structure from motion problem when the number of inertial sensors is reduced. This case corresponds to solve a problem that in control theory is known as the Unknown Input Observability (UIO). This problem was still unsolved in the nonlinear case. In [71] we introduced a new method able to provide sufficient conditions for the state observability. On the other hand, this method is based on a state augmentation. Specifically, the new extended state includes the original state together with the unknown inputs and their time-derivatives up to a given order. Then, the method introduced in [71] is based on the computation of a codistribution defined in the augmented space. This makes the computation necessary to derive the observability properties dependent on the dimension of the augmented state. Our effort to deal with this fundamental issue, was devoted to separate the information on the original state from the information on its extension. Last year, we fully solved this problem in the case of a single unknown input [73], [72]. This year we solved the problem for any number of unknown inputs. We presented this solution at the university of Pisa in June and at the university of Rome, Tor Vergata, in December.

[0] This is the partner of the ANR project VIMAD, in charge of the experiments

7.4. Motion-planning in human-populated environment

We explore motion planning algorithms to allow robots/vehicles to navigate in human populated environment, and to predict human motions.

We have proposed a novel planning-based motion prediction approach [27] which addresses the weaknesses of the previous state-of-the-art motion prediction technique [56], namely *High computational complexity* and *Limited ability to model the temporal evolution along the predicted path*. In 2016, this work has evolved in two new directions, which are prediction of pedestrian behaviors in urban environments and mapping of human flows. We also started to investigate the navigation of a telepresence robot in collaboration with the GIPSA Lab. These work are presented here after.

7.4.1. Urban Behavioral Modeling

Participants: Pavan Vasishta, Raphael Frisch, Anne Spalanzani.

The objective of modeling urban behavior is to predict the trajectories of pedestrians in towns and around car or platoons. We aim to integrate various entities of urban environments such as crosswalks, sidwalks, points of interest, but also characteristics of mobile obstacles (such cars and platoons) and proxemics in order to build a costmap that will show how pedestrians are driven around the ego-vehicle. This work is in the scope of the VALET project and the PhD of Pavan Vasishta (in collaboration with the Inria team Pervasive Interaction). It started in february 2016. Furthermore, a collaboration with the Laboratory of Psychology and NeuroCognition of Grenoble has been initiated to ground interaction and personal space models in experimental data from psychology.

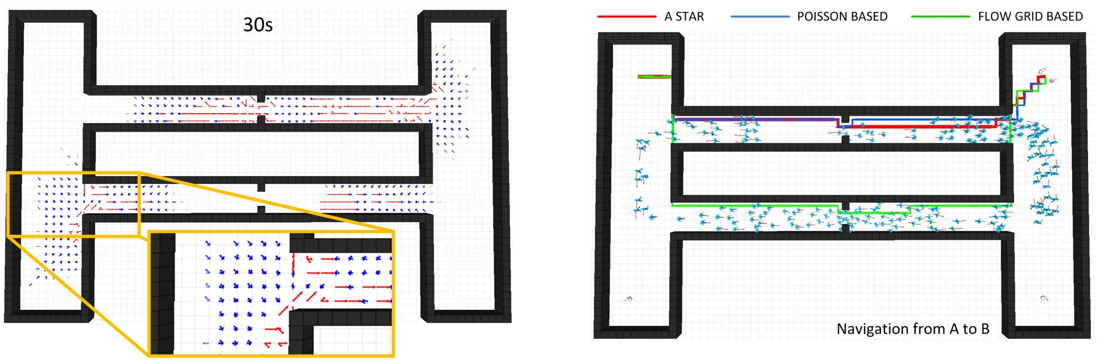


Figure 14. Illustration of (a) Flow-grid mapping in a two-corridor environment where 200 moving pedestrians turns (b) A path-planning computed with different cost functions in this populated environment*

7.4.2. Modeling human-flows from robot(s) perception

Participants: Olivier Simonin, Jacques Saraydaryan, Fabrice Jumel.

In order to deal with robot navigation in dense human populated environments, eg. in flows of humans, we started to investigate the problem of mapping these flows. The challenge is to build such a map from robot perceptions while robots move autonomously to perform some tasks. We developped a counting-based grid model which computes likelihoods of human presence and motion direction in each cell, see red vectors in Fig. 14 .a (this is a statistical learning of repetitive human motions). We extended the flow grid model with a human motion predictive model based on the Von Misses motion patern, allowing to "accelerate" the flow grid mapping, see blue vectors in Fig. 14 .a.

Then we explored how path-planning can benefit of such a flow grid, that is taking into account the risk for a robot to encounter humans in opposite direction. We first implement the Flow-Grid model in a simulator built upon PedSim and ROS tools, allowing to simulate mobile robots, crowd of pedestrians and sensors to detect their motion. Then, we compared three A*-based path-planning models using different levels of information about human presence: non-informed, a grid of human presence likelihood proposed by Tiplaldi [95] and our grid of human motion likelihood. Simulations involving 200 moving persons and 4 collaborative robots allowed to test simultaneously the mapping of human motions and the related path-planning. The different kind of paths obtained are illustrated in Fig. 14 .b, showing the ability of the flow-grid based A* to avoid to cross areas with a possible opposite human flow. These results have been presented at RSS workshop DEMUR [30].

We also started some experiments with our mobile indoor robots (incl. the Pepper) in human populated environments, see [30]. We plan to demonstrate the efficiency of the approach by participating to the new Pepper-league of the Robocup@Home competition, over the future period 2017-2020.

7.4.3. Navigation of telepresence robots

Participants: Olivier Simonin, Anne Spalanzani, Gerard Bailly [GIPSA, CNRS, Grenoble], Rémi Cambuzat.

In 2016 we obtained with the team of Gérard Bailly, from GIPSA/CNRS Grenoble, a regional support for the TENSIVE project. It funds the PhD thesis of Remi Cambuzat on immersive teleoperation of telepresence robots for verbal interaction and social navigation, started in October 2016. Chroma is focusing on the problem of social navigation, and more paricularly on the balance between human commands and autonomous navigation. Two issues are adressed : how to understand the expected direction given by the pilot to the teleprence robot, in order to ease the workload of the pilot ? how to assist the navigation, from embedded processes and sensors on the robot, while following the expected behavior given by the remote pilot ?

First months of the thesis concerned the building of a specific state-of-the-art, the formalization of some experimental scenarios, and the study of the Pepper robot capabilities in this scientific challenge.

7.5. Decision Making in Multi-Robot Systems

7.5.1. Multi-robot path-planning and patrolling

7.5.1.1. Patrolling under connectivity constraints

Participants: Olivier Simonin, Anne Spalanzani, Mihai Popescu, Fabrice Valois [Inria, Agora (ex Urbanet) team].

Patrolling is mainly used in situations where the need of repeatedly visiting certain places is critical. In this work, we consider a deployment of fixed targets, eg. wireless sensors, that several robots are in charge of patrolling while they have to maintain their (periodic) connectivity in order to collect and bring data up to a sink node. We have shown that this is fundamentally a problem of vertex coverage with bounded simple cycles (CBSC). We offered a formalization of the CBSC problem and proved it is NP-hard and at least as hard as the Traveling Salesman Problem (TSP). Then, we provided and analyzed heuristics relying on clusterings and geometric techniques. The proposed approach relies on two steps: the first one partitions the vertices, the second one computes hamiltonian cycles on each partition. We implemented two classic hamiltonian cycle heuristics, one is based on Minimum Spanning Trees computations and the other on Christofides algorithm. Comparisons on randomly-generated graphs showed that the Christofides algorithm computes shorter cycles. This work, started in the Master internship of Mihai-Ioan Popescu, now continuing as PhD student in Chroma, has been published in 2016 in [25]. Work is now focusing on the problem of synchronizing robots to meet at intersection nodes between the cycles.

Another important element of this work is the construction of a new collaboration with the team of Gabriela Czibula in Babes-Bolyai University at Cluj-Napoca (Romania). It will focus 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. We obtained, in the end of 2016, a french-romanian PHC [0] bilateral project, called DRONEM, funding students and researchers exchanges during two years.

[0] Hubert Curien Partnership

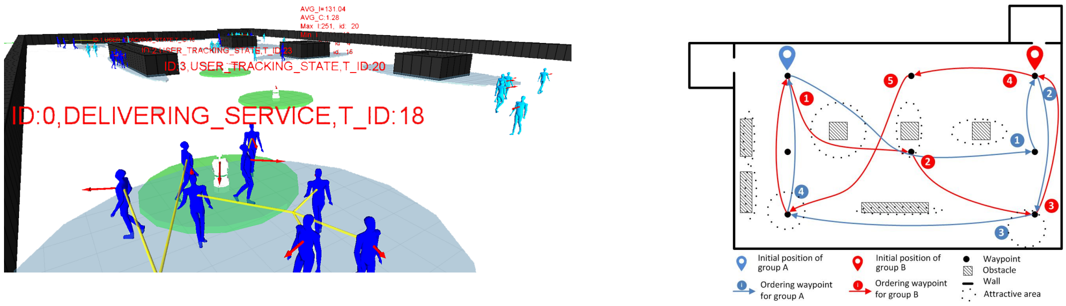


Figure 15. a) Simulator for dynamic patroling of people based on PedSim. b) Scenario of the 200 pedestrians moving along two predefined paths plus local attractors and randoms moves.

7.5.1.2. Patrolling moving people (dynamic patrolling)

Participants: Jacques Saraydaryan, Fabrice Jumel, Olivier Simonin.

In the context of service robotics, we address the problem of serving people by a set of collaborating robots, that is to delever regularly services to moving people. We showed that the problem can be formally expressed as a dynamic patrolling task. We call it the robot-waiters problem, where robots have to regularly visit all the moving persons (to deliver objects/information). In the publication [87], we proposed different criteria and metrics suitable to this problem, by considering not only the time to patrol all the people but also the equity of the delivery. We proposed and compared four algorithms, two are based on standard solutions to the static patrolling problem and two are defined according the specificity of patrolling moving entities, in particular greedy-based solutions on distance and idleness people information. In order to limit robot traveled distances, the last approach introduces a clustering heuristic to identify groups among people. To compare algorithms and to prepare real experiments we evaluated performances by using our simulator (combining PedSim and ROS). The simulator and the scenrio test - paths followed by humans - are illustrated in figure 15 .a and 15 .b. Experimental results show the efficiency of the specific new approaches over standard (static patrolling) approaches. We also analysed the influence of the number of robots on the performances, showing a convergence of performances when it grows. See [87] and extensions in 2016 [28].

We are currently developping new algorithms using the mapping and prediction of human flows based on the work presented in section 7.4.2 to allow robots to predict where human (groups) will move (under hypothesis of repetitive behaviors).

7.5.1.3. Global-local optimization in autonomous multi-vehicles systems

Participants: Olivier Simonin, Jilles Dibangoye, Laetitia Matignon, Florian Peyreron [VOLVO Group, Lyon], Guillaume Bono, Olivier Buffet [Inria Nancy Grand Est], Mohamed Tlig [IRT-Systemx, Paris].

We address transport and traffic management problems with driverless vehicles. We mainly study how local decisions can improve complexity of global (planning) solutions.

A previous work carried in the PhD of M. Tlig [96] concerned stop-free crossing roads with driverless vehicles. We explored distributed algorithms to optimize the global traffic in the road network (time and energy), based on Hill-Climbing techniques, so as to go from a synchronization within each intersection to the synchronization of a network. Experiments in simulation showed that proposed algorithms can efficiently optimize the initial decentralized solution, while keeping its properties (only the temporal phase for crossing in each intersection is modified). In 2016 we extended the experimental study, which was published in the RIA revue [13] and submited to an international journal.

In 2016, we started a new cooperation with the VOLVO Group, in the context of the INSA-VOLVO Chair. It funds the PhD thesis of G. Bono which deals with global-local optimization under uncertainty for goods distribution using a fleet of autonomous vehicles. First months of the thesis focused on building i) a state of the art about online pick-up and delivery solutions with a fleet of autonomous vehicles and ii) defining formally the scenario and hypothesis of the considered problem.

7.5.2. Anytime algorithms for multi-robot cooperation

7.5.2.1. Complex scenes observation

Participants: Olivier Simonin, Laetitia Matignon, Christian Wolf [LIRIS, INSA Lyon], Simon Bultmann [internship], Stefan Chitic.

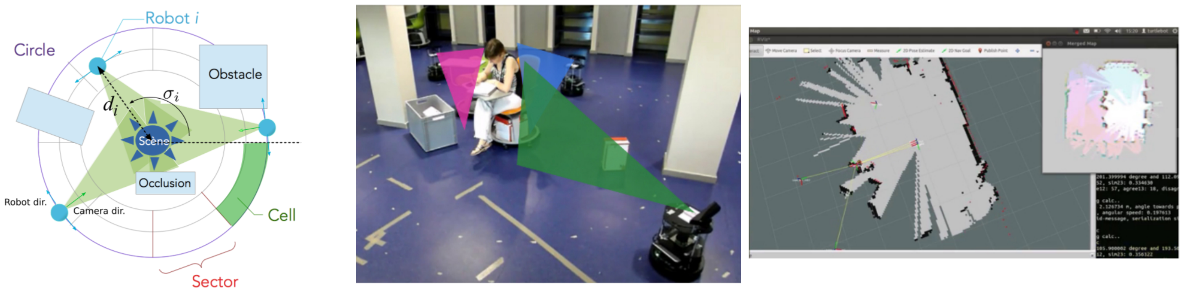


Figure 16. Illustrations (a) Concentric navigation model. (b) Experimental setup and multi-robot mapping with Turtlebot 2.

Solving complex tasks with a fleet of robots requires to develop general 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" [0] project [0] is 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 localization.

To attack the problem, in cooperation with colleagues from vision (C. Wolf, Liris), we proposed an original concentric navigation model allowing to keep easily each robot camera towards the scene (see fig. 16 .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. We developped a simulator that uses real data from real human pose captures to simulate dynamic scene and noise in sensor information. A video presenting the simulator interface and showing the incremental exploration and mapping can be found at . Results have been published in 2016 in [20] (ICTAI). It compares the variants of the approach and shows its features such as adaptation to the dynamic of the scene and robustness to the noise in the observations.

We have also developped an experimental framework for the circular navigation of several Turtlebot2 robots around a scene, presented in figure 16 .b. Especially, given that we assume in our work that robots have no map of the environment, we implemented a cooperative multi-robot mapping based on the merging of occupancy grid maps. Robots are individually building and communicating to other robots their local maps. Each robot tries to align these local maps to compute a joint, global representation of the environment. We carried out

[0] Coordination d'une flottille de robots mobiles pour l'analyse multi-vue de scènes complexes

[0] Funded by an INSA BQR in 2014-2015 (led by O. Simonin) and a LIRIS transversal project in 2016-2017 (led by L. Matignon)

the map-merging by adapting several methods known in literature [86] to our specific topology, i.e. the single hypothesis of a common center point (the scene) shared by robots. We compared the methods in real-world multi-robot scenarios (see Simon Bultmann's internship report).

7.5.2.2. Middleware for open multi-robot systems

Participants: Stefan Chitic, Julien Ponge [CITI, Dynamid], Olivier Simonin.

Multi-robots systems (MRS) require dedicated tools and models to face the complexity of their design and deployment (there is no or very limited tools/middleware for MRS). In the context of the PhD work of S. Chitic, we address the problem of neighbors and service discovery in an ad-hoc network formed by a fleet of robots. Robots needs a protocol that is able to constantly discover new robots in their coverage area. This led us to propose a robotic middleware, SDfR, that is able to provide service discovery. This protocol is an extension of the Simple Service Discovery Protocol (SSDP) used in Universal Plug and Play (UPnP) to dynamic networks generated by the mobility of the robots. Even if SDfR is platform independent, we proposed a ROS integration in order to facilitate the usage. We evaluated a series of overhead benchmarking across static and dynamic scenarios. Eventualy, we experimented some use-cases where our proposal was successfully tested with Turtlebot 2 robots. Results have been pubished in [19]. In 2016, the work continued by the definition of a timed automata based design and validation tool-set, that offers a framewok to formalize and implement multi-robot behaviors and to check some (temporal) properties.

7.5.3. Sequential decision-making under uncertainty

The holy grail of Artificial Intelligence (AI)—creating an agent (e.g., software, robot or machine) that comes close to mimicking and (possibly) exceeding human intelligence—remains far off. But past years have seen breakthroughs in agents that can gain abilities from experience with the environment, providing significant advances in the society and the industries including: health care, autonomous driving, recommender systems, etc. These advances are partly due to single-agent planning and (deep) reinforcement learning, that is, AI research subfields in which the agent can describe its world as a Markov decision process. Some stand-alone planning and reinforcement learning (RL) algorithms (e.g., Policy and Value Iteration, Q-learning) are guaranteed to converge to the optimal behavior, as long as the environment the agent is experiencing is Markovian. Although Markov decision processes provide a solid mathematical framework for single-agent planning and RL, they do not offer the same theoretical grounding in multi-agent systems, that is, groups of autonomous, interacting agents sharing a common environment, which they perceive through sensors and upon which they act with actuators. Multi-agent systems are finding applications everywhere today. At home, in cities, and almost everywhere, we are surrounded by a growing number of sensing and acting machines, sometimes visibly (e.g., robots, drones, cars, power generators) but often imperceptibly (e.g., smartphones, televisions, vacuum cleaners, wash- ing machines). Before long, through the emergence of a new generation of communication networks, most of these machines will be interacting with one another through the internet of things. In contrast to single-agent systems, when multiple agents interact with one another, how the environment evolves depends not only upon the action of one agent but also on the actions taken by the other agents, rendering the Markov property invalid since the environment is no longer stationary. In addition, a centralized (single-agent) control authority is often inadequate, because agents cannot (e.g., due to communication cost, latency or noise) or do not want (e.g., in competitive or strategic settings) to share all their information all the time. This raises a simple fundamental question: how to design a general algorithm for efficiently computing rational policies for a group of cooperating or competing agents in spite of stochasticity, limited information and computational resources? The remainder of this section points out some of the main results of the year to this question as well as ongoing projects.

7.5.3.1. Optimally solving cooperative games as continuous Markov decision processes

Participants: Jilles S. Dibangoye, Olivier Buffet [Inria Nancy], Christopher Amato [Univ. New Hampshire], François Charpillet [Inria Nancy, Larsen team].

Decentralized partially observable Markov decision processes (Dec-POMDPs) provide a general model for decision-making under uncertainty in decentralized settings, but are difficult to solve optimally (NEXP-Complete). As a new way of solving these problems, we introduce the idea of transforming a Dec-POMDP into a continuous-state deterministic MDP with a piecewise-linear and convex value function. This approach makes use of the fact that planning can be accomplished in a centralized offline manner, while execution can still be decentralized. This new Dec-POMDP formulation, which we call an occupancy MDP, allows powerful POMDP and continuous-state MDP methods to be used for the first time. To provide scalability, we refine this approach by combining heuristic search and compact representations that exploit the structure present in multi-agent domains, without losing the ability to converge to an optimal solution. In particular, we introduce a feature-based heuristic search value iteration (FB-HSVI) algorithm that relies on feature-based compact representations, point-based updates and efficient action selection. A theoretical analysis demonstrates that FB-HSVI terminates in finite time with an optimal solution. We include an extensive empirical analysis using well-known benchmarks, thereby demonstrating that our approach provides significant scalability improvements compared to the state of the art. This work has been published in JAIR journal [11].

7.5.3.2. Optimally solving two-person zero-sum partially observable stochastic games: a convex optimization approach

Participants: Jilles S. Dibangoye, Olivier Buffet [Inria Nancy], Mihai Indricean [INSA Lyon internship].

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 the result demonstrating 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 [39] 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. We also showed how our results apply to more restrictive settings, essentially leading to more concise information. Together these findings allow us to introduce a key algorithm avoiding exhaustive enumeration of doubly exponentially many pure strategies, as suggested so far. We further illustrate the use of our algorithm through numerical examples.

7.5.3.3. Decentralized Markov decision processes in open systems: models and first algorithms

Participants: Jilles S. Dibangoye, Abdel-Illah Mouaddib [Univ. Caen Basse-Normandie], Jonathan Cohen [Univ. Caen Basse-Normandie].

Many real-world multiagent applications, e.g., rescue operations, require to dynamically assemble or disassemble teams needed to provide a service based on agents entering or quitting the system. While Decentralized Partially Observable Markov Decision Processes (Dec-POMDPs) formalize decision-making by multiple agents, they fail to exploit the team flexibility. Queueing models can formalize birth-death processes by which agents enter or exit a team, but they fail to capture multiagent planning under uncertainty. This work, in the context of the PhD work of J. Cohen, introduces a new model synthesized from Dec-POMDPs and birth-death processes, called open Dec-POMDPs. The primary result is the proof that the latter is NEXP-Complete. Exploiting the team flexibility, enables us to present a best-response dynamics' algorithm, which can dynamically adapt to agents entering or quitting a team and computes local optimum solutions.

7.5.3.4. Does randomization makes cooperative multi-agent planning easier?

Participant: Jilles S. Dibangoye.

These recent years have seen significant progress in multi-agent planning problems formulated as decentralized partially observable Markov decision processes (Dec-POMDPs). In state-of-the-art algorithms, agents use policies that do not employ random devices, i.e., deterministic policies, which are simple to handle and to implement, and yet are good candidates to be optimal. Integer linear programming (ILP) or constraint optimization programming (COP) can formalize the search for deterministic policies, unfortunately their worst case complexity (NP-Complete) suggest to be little hope for optimally solving real-world instances. In this paper, we show—for the first time—that the randomization allows us to use linear programming (LP) instead of ILP while preserving optimality, which drops the worst-case complexity from NP to P. Specifically, we introduce the first linear programs for incrementally approaching the optimal value function starting from upper- and lower-bound functions. We further extends state-of-the-art algorithm for Dec-POMDPs to use randomized policies. Finally, empirical results demonstrate significant improvements in time needed to find an ε-optimal solution on all tested benchmarks.

7.5.3.5. Reinforcement learning approach for active perception using multiple robots

Participants: Jilles S. Dibangoye, Jacques Saradaryan, Laëtitia Matignon, Trad Ahmed Yahia [Master Internship], Lorcan Charonnat [Internship INSA], Yuting Zhang [Internship INSA], Yifan Xiong [Internship INSA].

We consider cooperative, decentralized stochastic control problems represented as a decentralized partially observable Markov decision process. A critical issue that limits the applicability of that setting to realistic domains is how agents can learn to act optimally by interacting with the environment and with one another, given only an incomplete knowledge about the model. Reinforcement learning has previously been applied to decentralized decision making with a focus on distributed methods, which often results in suboptimal solutions. On the contrary, we build upon the idea that plans that are to be executed in a decentralized fashion can nonetheless be formulated in a centralized manner using a generative model of the environment. Following this line of thought, we propose the first (centralized) reinforcement learning algorithm for computing the optimal Q-value functions for cooperative, decentralized stochastic control problems. Experiments show our approach can learn to act optimally in many domains from the literature. We currently investigate robotic applications of this approach, including unknown scene reconstruction by a fleet of mobile robots.

DEFROST Team

7. New Results

7.1. Cochlear Implants

Publication at MICCAI 2016 (Medical Image Computing and Computer Assisted Intervention conference): **Numerical Simulation of Cochlear-Implant Surgery: Towards Patient-Specific Planning**, *Olivier Goury, Yann Nguyen, Renato Torres, Jeremie Dequidt, Christian Duriez*. **Abstract.** During Cochlear Implant Surgery, the right placement of the implant and the minimization of the surgical trauma to the inner ear are an important issue with recurrent fails. In this study, we reproduced, using simulation, the mechanical insertion of the implant during the surgery. This simulation allows to have a better understanding of the failing cases: excessive contact force, buckling of the implant inside and outside the cochlea. Moreover, using a patient-specific geometric model of the cochlea in the simulation, we show that the insertion angle is a clinical parameter that has an influence on the forces endured by both the cochlea walls and the basilar membrane, and hence to post-operative trauma. The paper presents the mechanical models used for the implant, for the basilar membrane and the boundary conditions (contact, friction, insertion etc...) and discuss the obtained results in the perspective of using the simulation for planning and robotization of the implant insertion.

https://hal.archives-ouvertes.fr/hal-01370185

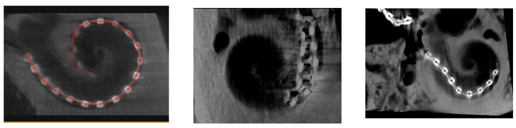


Figure 3. Three outcomes of implant insertion (from left to right): successful insertion; failed insertion (Folding tip); incomplete insertion

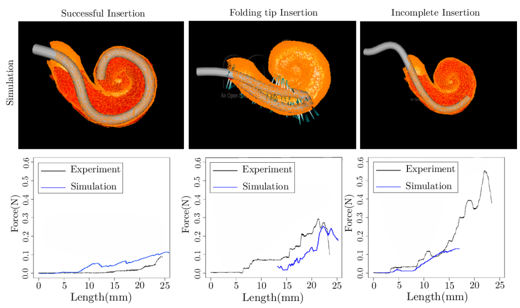


Figure 4. Reproduction of real insertion cases with the simulation

7.2. Physics based model of soft-robots

Book chapter in Soft Robotics: Trends, Applications and Challenges, Springer, 2016 **Soft Robot Modeling, Simulation and Control in Real-Time**, *Christian Duriez and Thor Bieze*. https://hal.inria.fr/hal-01410293. We were asked to write a chapter in this book on Soft Robotics. Our chapter presents new real-time and physics-based modeling methods dedicated to deformable soft robots. In our approach, continuum mechanics provides the partial derivative equations that govern the deformations, and Finite Element Method (FEM) is used to compute numerical solutions adapted to the robot. A formulation based on Lagrange Multipliers is presented to model the behavior of the actuators as well as the contact with the environment. Direct and inverse kinematic models are also obtained for real-time control. Some experiments and numerical results are presented.

7.3. Kinematic Modeling and control of soft robots

Publication at IROS 2016 : **Kinematic Modeling and Observer Based Control of Soft Robot using Real-Time Finite Element Method**, *Zhongkai Zhang, Jeremie Dequidt, Alexandre Kruszewski, Frederick Largilliere, Christian Duriez.*. **Abstract.**This paper aims at providing a novel approach to modeling and controlling soft robots. Based on real-time Finite Element Method (FEM), we obtain a globally defined discrete-time kinematic model in the workspace of soft robots. From the kinematic equations, we deduce the soft-robot Jacobian matrix and discuss the conditions to avoid singular configurations. Then, we propose a novel observer based control methodology where the observer is built by Finite Element Model in this paper to deal with the control problem of soft robots. A closed-loop controller for position control of soft robot is designed based on the discrete-time model with feedback signal being extracted by means of visual servoing. Finally, experimental results on a parallel soft robot show the efficiency and performance of our proposed controller. https://hal.inria.fr/hal-01370347

7.4. Stiffness rendering

Publication at IROS 2016 : **Stiffness rendering on soft tangible devices controlled through inverse FEM simulation**, *Frederick Largilliere, Eulalie Coevoet, Mario Sanz-Lopez, Laurent Grisoni, Christian Duriez*. **Abstract.**Haptic rendering of soft bodies is essential in medical simulations of procedures such as surgery or palpation. The most commonly used approach is to recreate the sense of touch using a specific design and control of a robotic arm. In this paper, we propose a new approach, based on soft-robotics technology. We create a tangible deformable device that allows users to " touch " soft tissues and perceive mechanical material properties, in a realistic manner. The device is able to dynamically provide user touch with different stiffness perceptions, thanks to actuators placed at the boundaries. We introduce a control algorithm, based on inverse Finite Element Analysis, which controls the actuators in order to recreate a desired stiffness that corresponds to the contact with soft tissues in the virtual environment. The approach uses antagonistic actuation principle to create a wide range of stiffness. We validate our algorithm and demonstrate the method using prototypes based on simple mechanisms. https://hal.inria.fr/hal-01386787

7.5. Framework for soft robot simulation

Publication at SIMPAR 2016 : **Framework for online simulation of soft robots with optimization-based inverse model**, *C. Duriez, E. Coevoet, F. Largilliere, T. Morales-Bieze, Z. Zhang, M. Sanz-Lopez, B. Carrez, D. Marchal, O. Goury, J. Dequidt*. **Abstract.**Soft robotics is an emerging field of robotics which requires computer-aided tools to simulate soft robots and provide models for their control. Until now, no unified software framework covering the different aspects exists. In this paper, we present such a framework from its theoretical foundations up to its implementation on top of Sofa, an open-source framework for deformable online simulation. The framework relies on continuum mechanics for modeling the robotic parts and boundary conditions like actuators or contacts using a unified representation based on Lagrange multipliers. It enables the digital robot to be simulated in its environment using a direct model. The model can also be inverted online using an optimization-based method which allows to control the physical robots in the task space. To demonstrate the effectiveness of the approach, we present various soft robots scenarios including ones where the robot is interacting with its environment. https://hal.inria.fr/hal-01425349

7.6. Closed-loop control

Closed-loop control based on dynamic models of soft robots. Model-order reduction provides a system of achievable size to apply traditional control science techniques. During the internship of Maxime Thieffry, we obtain the first results in that direction that will be extended during a PhD thesis.

FLOWERS Project-Team

7. New Results

7.1. Robotic And Computational Models Of Human Development and Cognition

7.1.1. Computational Models Of Information-Seeking, Curiosity And Attention in Humans and Animals

Participants: Manuel Lopes, Pierre-Yves Oudeyer [correspondant], Jacqueline Gottlieb, Celeste Kidd, Alvaro Ovalle, William Schueller, Sebastien Forestier, Nabil Daddaouda, Nicholas Foley.

This project involves a collaboration between the Flowers team, the Cognitive Neuroscience Lab of J. Gottlieb at Columbia Univ. (NY, US), and the developmental psychology lab of Celeste Kidd at Univ. Rochester, US, on the understanding and modeling of mechanisms of curiosity, attention and active intrinsically motivated exploration that until now have been little explored in neuroscience, machine learning and cognitive robotics.

It is organized around the study of the hypothesis that information gain (or control gain) 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 (Gottlieb's lab) and children (Kidd's lab) together with computational modeling of curiosity/intrinsic motivation and learning in the Flowers team.

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 [29], [155], [162]. Therefore we hypothesize that, rather than using a single optimization process as it has been the case in most previous theoretical work [131], 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 [153] [110], [149]. 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. We will quantitatively characterize this dynamic, multi-dimensional system in the framework of Bayesian Reinforcement Learning, as described below.

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 [143], [161], and rats show individual variation in learning styles and novelty seeking behaviors [128], 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 [159], [171], [106], [175]. 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 model based on Bayesian reinforcement learning.

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 [143], [141] 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 Bayesian Reinforcement 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

In particular, new results in 2015 include:

7.1.1.4. Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates

Using a novel oculomotor paradigm, combined with reinforcement learning (RL) simulations, we show that monkeys are intrinsically motivated to search for and look at reward-predictive cues, and that their intrinsic motivation is shaped by a desire to reduce uncertainty, a desire to obtain conditioned reinforcement from positive cues, and individual variations in decision strategy and the cognitive costs of acquiring information. The results suggest that free-viewing oculomotor behavior reveals cognitive and emotional factors underlying the curiosity driven sampling of information. These results were published in [66].

7.1.1.5. Experiments in Active Categorization

An ongoing effort to characterize curiosity and exploration in an experimental setting consists in evaluating the manner in which diverse tasks or goals are selected. This would include monitoring what does a test subject decide to learn, in what order and how is it done. This has been referred to as strategic learning [31]. Accordingly, it is of particular interest for the project to observe the type of learning dynamics in relation to their learning progress [153]. This principle tries to establish links between the selection and ordering of tasks and the speed or the rate of improvement a subject may achieve. This implies that during free exploration the subject would focus on tasks that are considered of certain complexity and where it makes consistent progress. At the same time the subject would avoid: (1) trivial tasks that do not offer much learning due to their simplicity or (2) very complicated tasks where little or no progress is achieved.

We have been working on prototyping an experiment where the subject is presented with different stimuli classification tasks of varying difficulty. The goal for each of the tasks is to correctly predict and differentiate between different classes of stimuli. Two main aspects of the task are under the control of the subject: (1) the task that he/she wants to learn and (2) once selected a task, what elements to explore in order to subsequently being able to predict future stimuli. Essentially the subject autonomously organizes which tasks to focus on and in what order. Therefore one of the objectives of this investigation is to analyze if the learning dynamics are guided by the amount of progress the subject achieves in the tasks.

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], Clement Moulin-Frier, Sébastien Forestier, Linda Smith.

7.1.2.1. Modeling Cognitive Development and Tool Use 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 [131]. 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 [157].

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 [3]). On the other hand, other algorithmic models have shown how several social learning mechanisms could allow robots to acquire elements of speech and language [118], 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 [49].

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 [37]. 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 [152].

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 [118]. 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 [152],[3], [43], [37]. 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.

A first question that we study in this work is the type of mechanisms that could be used for hierarchical skill learning allowing to manage new task spaces and new action spaces, where the action and task spaces initially given to the robot are continuous and high-dimensional and can be encapsulated as primitive actions to affect newly learnt task spaces.

We presented firsts results on that question at the 38th Annual Meeting of the Cognitive Science Society, Philadelphia, Pennsylvania, USA, August 10-13th [80]. In this work, we presented the HACOB (Hierarchical Active Curiosity-driven mOdel Babbling) architecture of algorithms that actively chooses which sensorimotor model to train in a hierarchy of models representing the environmental structure. We studied this architecture using a simulated robotic arm interacting with objects in a 2D environment (See Fig. 8). Studies of child development of tool use precursors showed successive but overlapping phases of qualitatively different types of behaviours [167]. We hypothesized that two mechanisms in particular play a role in the structuring of these phases: the intrinsic motivation to explore and the representation used to encode sensorimotor experience.

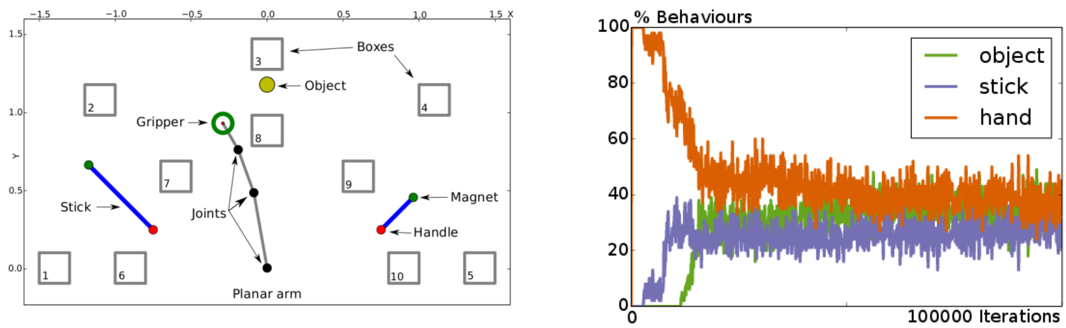


Figure 8. Left: simulated robotic environment with a 4 DOF robotic arm, 2 tools and a toy. Right: Observed behaviours of an agent: it first explores its arm to move its hand, then also explore to move the stick and the toy.

We showed that using a hierarchical structure of sensorimotor models and active model babbling as an intrinsic motivation to explore sensorimotor models that have a high learning progress, then overlapping phases of behaviours are autonomously emerging in the developmental trajectories of agents. To our knowledge, this is the first model of curiosity-driven development of simple tool use and of the self-organization of overlapping phases of behaviours. In particular, our model explains why and how intrinsically motivated exploration of non-optimal methods to solve certain sensorimotor problems can be useful to discover how to solve other sensorimotor problems, in accordance with Siegler's overlapping waves theory, by scaffolding the learning of increasingly complex affordances in the environment.

In computational models of strategy selection for the problem of integer addition, Shrager and Siegler proposed a mechanism that maintains the concurrent exploration of alternative strategies with use frequencies that are proportional to their performance for solving a particular problem. This mechanism was also used by Chen and Siegler to interpret an experiment with 1.5- and 2.5-year-olds that had to retrieve an out-of-reach toy, and where they could use one of several available strategies that included leaning forward to grasp a toy with the hand or using a tool to retrieve the toy.

In a paper that we presented at the The Sixth Joint IEEE International Conference on Developmental Learning and Epigenetic Robotics, Cergy-Pontoise, France, September 19-22nd [82], we studied tool use discovery and considered other mechanisms of strategy selection and evaluation. In particular, we presented models of curiosity-driven exploration where strategies are explored according to the learning progress/information gain they provide (as opposed to their current efficiency to actually solve the problem). In these models, we defined a curiosity-driven agent learning a hierarchy of different sensorimotor models in a simple 2D setup with a robotic arm, a stick and a toy. In a first phase, the agent learns from scratch how to use its robotic arm to control the tool and to catch the toy, and in a second phase with the same learning mechanisms, the agent has to solve three problems where the toy can only be reached with the tool (See Fig. 9). We showed that agents choosing strategies based on learning progress also display overlapping waves of behavior compatible with the one observed in infants, and we suggested that curiosity-driven exploration could be at play in Chen and Siegler's experiment, and more generally in tool use discovery.

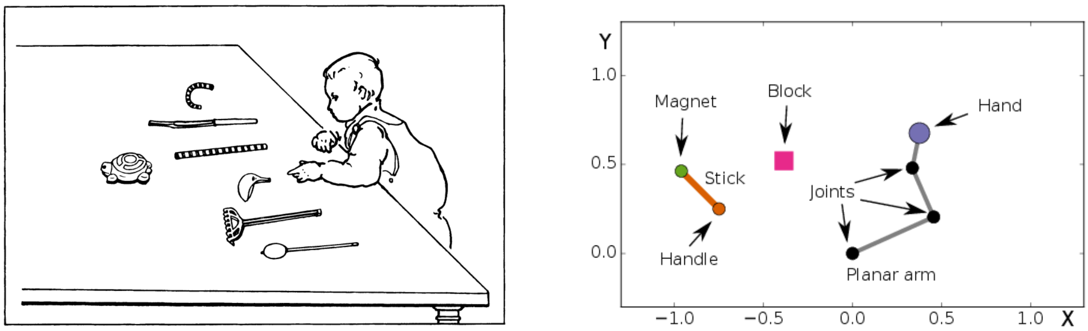


Figure 9. Left: Chen and Siegler's experimental setup with 1.5 and 2.5 years old babies who have to pick the good toy to retrieve an interesting toy. Right: Simulated robotic setup with a 3 DOF robotic arm that has 2 strategies to retrieve a toy: either grasp it with the hand, or use the stick to pull the toy.

7.1.2.2. Curiosity-driven developmental processes and their role in development and evolution of language

Infants' own activities create and actively select their learning experiences. In a collaboration with Linda Smith, we have analyzed recent models of embodied information seeking and curiosity-driven learning and have showed that these mechanisms have deep implications for development and evolution. In [69], we have discussed how these mechanisms yield self-organized epigenesis with emergent ordered behavioral and cognitive developmental stages. We described a robotic experiment that explored the hypothesis that progress in learning, in and for itself, generates intrinsic rewards: the robot learners probabilistically selected experiences according to their potential for reducing uncertainty. In these experiments, curiosity-driven learning led the robot learner to successively discover object affordances and vocal interaction with its peers. We explain how a learning curriculum adapted to the current constraints of the learning system automatically formed, constraining learning and shaping the developmental trajectory. The observed trajectories in the robot experiment share many properties with those in infant development, including a mixture of regularities and diversities in the developmental patterns. Finally, we argued that such emergent developmental structures can guide and constrain evolution, in particular with regards to the origins of language.

7.1.3. Computational Models Of Developmental Exploration Mechanisms in Vocal Babbling and Arm Reaching in Infants

Participants: Pierre-Yves Oudeyer [correspondant], Clement Moulin-Frier, Freek Stulp, Jules Borchard.

7.1.3.1. Proximodistal Freeing of DOFs in Motor Learning as an Emergent Property of Stochastic Optimization Principles

To harness the complexity of their high-dimensional bodies during sensorimotor development, infants are guided by patterns of freezing and freeing of degrees of freedom. We have formulated and studied computationally the hypothesis that such patterns, such as the proximodistal freeing of degrees of freedom when learning to reach, can emerge spontaneously as the result of a family of stochastic optimization processes, without an innate encoding of a mat- urational schedule. In particular, we present simulated experiments with a 6-DOF arm where a computational learner progressively acquires reaching skills through adaptive exploration, and we show that a proximodistal organization appears spontaneously, which we denote PDFF (ProximoDistal Freezing and Freeing of degrees of freedom). We also compare the emergent structuration as different arm structures are used – from human-like to quite unnatural ones – to study the effect of different kinematic structures on the emergence of PDFF.

7.1.3.2. Emergent Jaw Predominance in Vocal Development through Stochastic Optimization

Infant vocal babbling is strongly relying on jaw oscillations, especially at the stage of canonical babbling, which underlies the syllabic structure of world languages. We have proposed, modelled and analyzed an hypothesis to explain this predominance of the jaw in early babbling. This hypothesis states that general stochastic optimization principles, when applied to learning sensorimotor control, automatically generate ordered babbling stages with a predominant exploration of jaw movements in early stages, just like they generate proximo-distal organization of exploration in arm reaching as described in the paragraph above. In particular, such stochastic optimization principles predominantly explore jaw movement at the beginning of vocal learning, and when close to the rest position of the vocal tract, as it impacts the auditory effects more than other articulators.

7.1.4. Learning and Teaching in Adult-Child and Human-Robot Interaction

Participants: Anna-Lisa Vollmer [correspondant], Pierre-Yves Oudeyer.

7.1.4.1. Pragmatic Frames

One of the big challenges in robotics today is to learn from human users that are inexperienced in interacting with robots but yet are often used to teach skills flexibly to other humans and to children in particular. A potential route toward natural and efficient learning and teaching in Human-Robot Interaction (HRI) is to leverage the social competences of humans and the underlying interactional mechanisms. In this perspective, we propose 'pragmatic frames' as flexible interaction protocols that provide important contextual cues to enable learners to infer new action or language skills and teachers to convey these cues. Following the concept developed in the field of developmental linguistics [117], we define a pragmatic frame to be an interaction protocol negotiated over time between interaction partners. We further specify a Pragmatic Frame to especially involve an observable **coordinated sequence of behaviors** and also relevant **cognitive operations**.
Figure 10 depicts the book reading frame Bruner observed in his studies on word learning.

At home, a mother is sitting on the sofa with her child on her lap and she is holding a picture book in front of them. The mother points to the book and says "look!" to direct the child's attention. The child then gazes to the image. And the mother asks "What‘s that?", prompting the child's performance. The child answers with babble strings and smiles, maybe "auo". "Yes, a pineapple!" The mother gives positive feedback and the correct label. "Anappo", again babble strings and smiles. And the mother gives positive feedback. This stable sequence that the child is familiar with helps the child to participate and to pick up the only variable information he or she is supposed to learn. We argue that this frame also triggers the relevant cognitive functions to process the information.

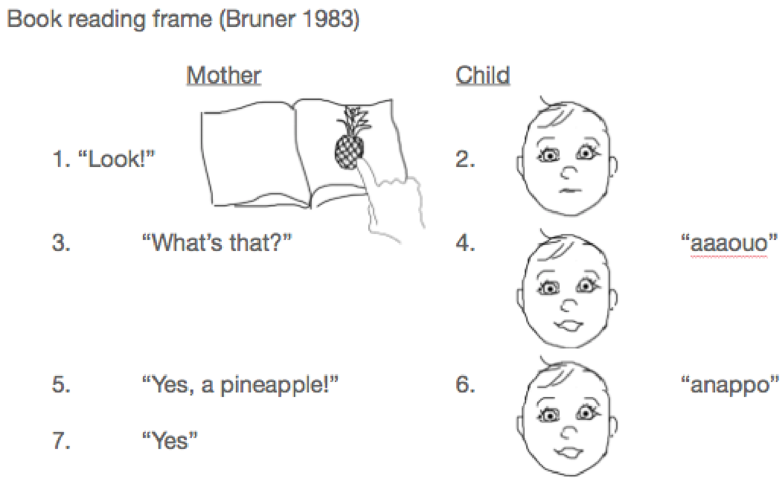


Figure 10. Example of a learning/teaching pragmatic frame.

Our results in 2016 have been twofold. First, in a paper published in Frontiers in Psychology [70], we have given a theoretical account of pragmatic frames as an alternative to the mapping metaphor which posits that children learn a word by mapping it onto a concept of an object or event. However, we believe that a mapping metaphor cannot account for word learning, because even though children focus attention on objects, they do not necessarily remember the connection between the word and the referent unless it is framed pragmatically, that is, within a task. Word learning with pragmatic frames occurs as children accomplish a goal in cooperation with a partner. We elaborate on pragmatic frames, offer some initial parametrizations of the concept, and embed it in current language learning approaches.

Second, aiming at leveraging the concept of pragmatic frames for Human-Robot Interaction, we published an article in Frontiers in Neurorobotics [71] in which we study a selection of HRI work in the literature which has focused on learning–teaching interaction and analyze the interactional and learning mechanisms that were used in the light of pragmatic frames. This allows us to show that many of the works have already used in practice, but not always explicitly, basic elements of the pragmatic frames machinery. However, we also show that pragmatic frames have so far been used in a very restricted way as compared to how they are used in human–human interaction and argue that this has been an obstacle preventing robust natural multi-task learning and teaching in HRI. In particular, we explain that two central features of human pragmatic frames, mostly absent of existing HRI studies, are that (1) social peers use rich repertoires of frames, potentially combined together, to convey and infer multiple kinds of cues; (2) new frames can be learnt continually, building on existing ones, and guiding the interaction toward higher levels of complexity and expressivity. To conclude, we give an outlook on the future research direction describing the relevant key challenges that need to be solved for leveraging pragmatic frames for robot learning and teaching.

7.1.5. 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 underlie 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 [168]. 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 [30], [131], [140]. This approach has been mostly studied for single robotic agents learning sensorimotor affordances [153], [38]. However active learning might represent an evolutionary advantage for language formation at the population level as well [49], [170].

Naming Games are a computational framework, elaborated to simulate the self-organization of lexical conventions in the form of a multi-agent model [169]. 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 [53], [46], [122] or the hearer [90] actively controlling vocabulary growth.

We study here how the convergence time and the maximum level of complexity scale with population size, for three different strategies (one with random topic choice and two with active topic choice) detailed in table 11 .

Naive (random)	Success Threshold	Minimal counts
$m \leftarrow \text{random}(\mathcal{M})$	**if** mean$\left(\frac{succ(i)}{succ(i)+fail(i)}\right)_{i\in\mathcal{LM}} \geq \boldsymbol{\alpha}$: $m \leftarrow \text{random}(\mathcal{UM})$ **else**: $m \leftarrow \text{argmin}_{i\in\mathcal{LM}}\left(\frac{succ(i)}{succ(i)+fail(i)}\right)$	**if** $\forall i \in \mathcal{LM}\ succ(i) > \boldsymbol{n}$: $m \leftarrow \text{random}(\mathcal{UM})$ **else**: $m \leftarrow \text{argmin}_{i\in\mathcal{LM}}(succ(i))$
$\mathcal{M}$: all meanings, $\mathcal{LM}$: labeled meanings, $\mathcal{UM}$: unlabeled meanings, μ: vocabulary size (# word-meaning associations) $succ$: # successful interactions per meaning, $fail$: # failed interactions per meaning		

Figure 11. Strategies: Choice of meaning m. Both active strategies use a parameter (α and n), which is each time chosen optimal in our simulations.

As for the version of the Naming Game used in our work, the scenario of the interaction is described in [90]. Vocabulary is updated as described in the Minimal Naming Game, detailed in [177]. In our simulations, we choose to set $N = M = W$, where N is the population size, M the number of meanings, and W the number of possible words. The computed theoretical success ratio of communication is used to represent the degree of convergence toward a shared lexicon for the whole population. A value of 1 means that the population reached full convergence. Complexity level of an individual lexicon is measured as the total number of distinct associations between meanings and words in the lexicon, or in other words: memory usage.

We show here (see figures 12 ,13) that convergence time and maximum complexity are reduced with active topic choice, a behavior that is amplified as larger populations are considered. The minimal counts strategy yields a strictly minimum complexity (equal to the complexity of a completed lexicon), while converging as fast as the success threshold strategy. Further work will deal with other variants of the Naming Game (with different vocabulary update, population replacement, and different ratio for N, M and W). For the moment only the hearer's choice scenario is studied, because of its high robustness to changes in parameter values for the different strategies [90].

7.2. Lifelong Robot Learning And Development Of Motor And Social Skills

7.2.1. Intrinsically Motivated Multitask Reinforcement Learning

Participants: Sébastien Forestier [correspondant], Pierre-Yves Oudeyer, Fabien Benureau.

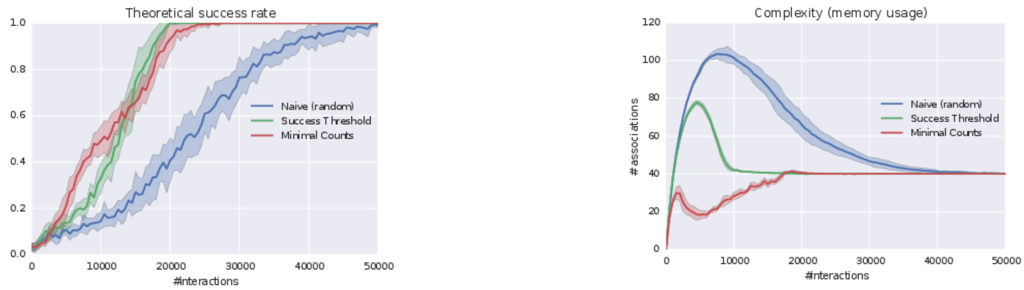


Figure 12. Strategy comparisons, in terms of convergence time (theoretical success ratio) and complexity level (memory usage). In this case, the hearer is the one choosing the topic. M=W=N=40, averaged over 8 trials

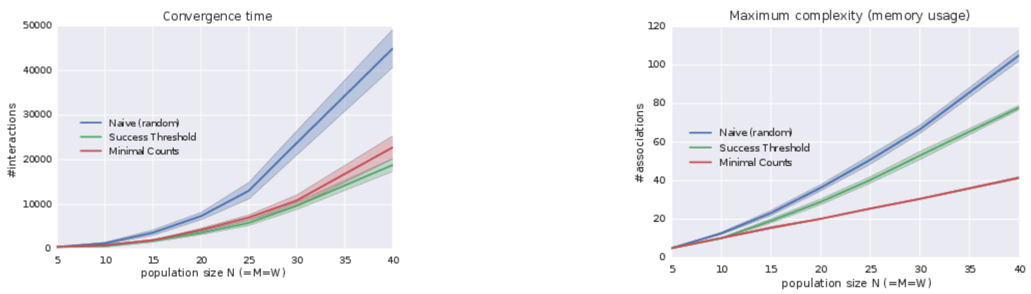


Figure 13. Scaling of maximum memory usage and convergence time for the different strategy, in function of population size. In this case, the hearer is the one choosing the topic. M=W=N, averaged over 8 trials.

7.2.1.1. *Intrinsically Motivated Exploration of Spaces of Parameterized Skills/Tasks and Application to Robot Tool Learning*

A major challenge in robotics is to learn 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 problems 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 [109].

In this work we study algorithms used by a learner to explore high-dimensional structured sensorimotor spaces such as in tool use discovery. We consider goal babbling architectures that were designed to explore and learn solutions to fields of sensorimotor problems, i.e. to acquire inverse models mapping a space of parameterized sensorimotor problems/effects to a corresponding space of parameterized motor primitives. However, so far these architectures have not been used in high-dimensional spaces of effects. Here, we show the limits of existing goal babbling architectures for efficient exploration in such spaces, and introduce a novel exploration architecture called Model Babbling (MB). MB exploits efficiently a modular representation of the space of parameterized problems/effects. We also study an active version of Model Babbling (the MACOB architecture). We compared those architectures in a simulated experimental setup with an arm that can discover and learn how to move objects using two tools with different properties, embedding structured high-dimensional continuous motor and sensory spaces (See Fig. 14).

7.2.1.2. *Transfer Learning through Measures of Behavioral Diversity Generation in Autonomous Exploration*

The production of behavioral diversity – producing a diversity of effects – is an essential strategy for robots exploring the world when facing situations where interaction possibilities are unknown or non-obvious. It allows to discover new aspects of the environment that cannot be inferred or deduced from available knowledge. However, creating behavioral diversity in situations where it is most crucial – new and unknown ones – is far from trivial. In particular in large and redundant sensorimotor spaces, only small areas are interesting to explore for any practical purpose. When the environment does not provide clues or gradient toward those areas, trying to discover those areas relies on chance. To address this problem, we introduce a method to create behavioral diversity in a new sensorimotor task by re-enacting actions that allowed to produce behavioral diversity in a previous task, along with a measure that quantifies this diversity. We have shownd that our method can learn how to interact with an object by reusing experience from another, that it adapts to instances of morphological changes and of dissimilarity between tasks, and how scaffolding behaviors can emerge by simply switching the attention of the robot to different parts of the environment. Finally, we show that the method can robustly use simulated experiences and crude cognitive models to generate behavioral diversity in real robots. This work was published in [62].

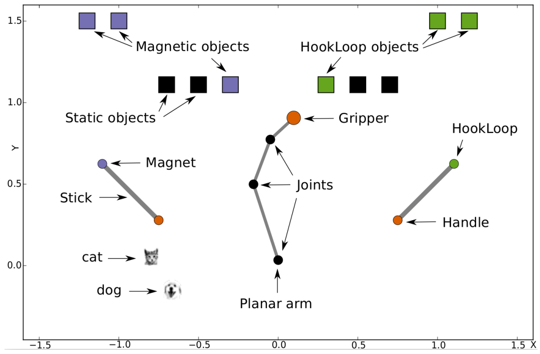
We presented the results at the IEEE/RSJ International Conference on Intelligent Robots and Systems, Daejeon, Korea, October 9-14th [81].

7.2.2. *Social Learning of Interactive Skills*

Participants: Manuel Lopes [correspondant], Thibaut Munzer, Marc Toussaint, Li Wang Wu, Yoan Mollard, Baptiste Busch, Jonathan Grizou, Marie Demangeat, Freek Stulp.

7.2.2.1. *Relational Activity Processes for Modeling Concurrent Cooperation*

In human-robot collaboration, multi-agent domains, or single-robot manipulation with multiple end-effectors, the activities of the involved parties are naturally concurrent. Such domains are also naturally relational as they involve objects, multiple agents, and models should generalize over objects and agents. We propose a novel formalization of relational concurrent activity processes that allows us to transfer methods from standard relational MDPs, such as MonteCarlo planning and learning from demonstration, to concurrent cooperation domains. We formally compare the formulation to previous propositional models of concurrent



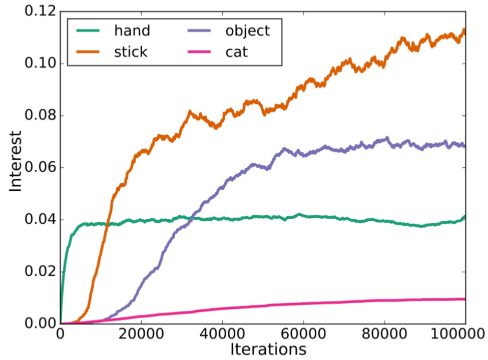


Figure 14. Left: Simulated robotic setup with a robotic arm that can grab tools to retrieve some interesting objects among a set of controllable and non-controllable (cat and dog) objects. Right: Evolution of the self-measured learning progress to move objects, with the MACOB active exploration architecture. The learning progress to explore objects is increasing for the tool and toy objects and stays low for the uncontrollable animals.

decision making and demonstrate planning and learning from demonstration methods on a real-world human-robot assembly task. A paper summarizing this research has been publish to the *International Conference on Robotics and Automation* (ICRA) 2016 [84].

7.2.2.2. Interactive Behavior Learning for Cooperative Tasks

This work goal is to propose a method to learn cooperative behavior to solve a task while performing the task with the user. The proposed approach reuses previous work on learning policy for RAP. The main differences are: i) formulate the problem as a cooperative process. In MDP and RAP, it is assumed that there is one central decision maker. However, in a cooperative both the robot and the operator are taking decisions. ii) estimating the confidence. A Query by Bagging approach has been used where many policies are learned from a subset of the data. Their potential disagreement allows quantifying the confidence. iii) Using the confidence for autonomous acting and for query making. Based on the confidence, the robot either act before acting or ask confirmation before acting.

Results show that using an interactive approach require less instruction from the user while producing a policy that makes fewer mistakes. We developed a robotic implementation 15 using a Baxter robot. A first article resulting from this work focusing on interactive preferences learning have been submitted to the *International Conference on Robotics and Automation* (ICRA) 2017 and a video demonstration can be view at : https://vimeo.com/182913540. A broader journal article is in preparation. We also conducted a user study to evaluate the impact of interactive learning on naïve users acceptation and performances.

7.2.2.3. Legible Motion

In a human-robot collaboration context, understanding and anticipating the robot intentions ease the completion of a joint-task. Whereas previous work has sought to explicitly optimize the legibility of behavior, we investigate legibility as a property that arises automatically from general requirements on the efficiency and robustness of joint human-robot task completion.

Following our previous work on legibility of robot motions [56], we have conducted several user experiments to analyze the effects of the policy representation on the universality of the legibility.

This work lead to a submission of a journal article to the International Journal of Social Robotics (IJSR) under the special issue: Towards a Framework for Joint Action. The article has been accepted with minor revisions and is currently in the final stage of the review process.

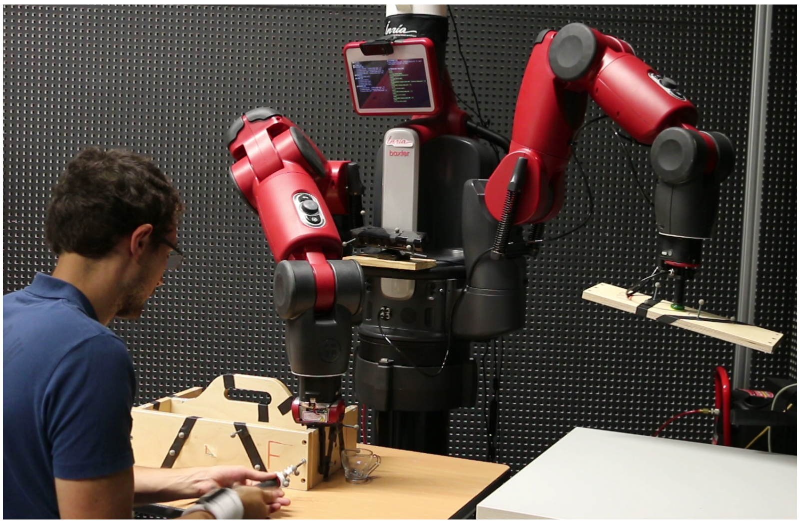


Figure 15. Interactive cooperative task learning.

7.2.2.4. Postural optimization for a safe and comfortable human-robot interaction

When we, humans, accomplish a task our body posture is (partially) constrained. For example, acting on an object constrains the pose of the hand relatively to the object, and the head faces the object we are acting upon. But due to the large number of degrees of freedom (DOF) of the human body, other body parts are unconstrained and several body postures are viable with respect to the task. However, not all of them are viable in terms of ergonomics. Using a personalized human model, observational postural assessment techniques can be automatized. Optimizing the model body posture is then the logical next step to find an ergonomically correct posture for the worker to accomplish a specific task.

To optimize the subject's model to achieve a specific task, we define an objective function that minimizes the efforts of the whole body posture, based on the Rapid Entire Body Assessment (REBA) technique [135]. The objective function also account for visibility of the target object and worker's laterality. We have also implemented an automatic assessment of the worker's body posture based on the REBA method.

Using a spherical object, carried by a Baxter humanoid robot as illustrated in Fig. 16 , we mimic an industrial scenario where the robot helps the worker by positioning and orienting an object in which the worker has to insert specific shapes. In a user-study with forty participants, we compare three different robot's behaviors, one of them being the result of the postural optimization of the subject's personalized model. By the mean of a survey session, and the online assessment of the subject's posture during the interaction, we prove that our method leads to a safer posture, and is perceived as more comfortable.

This work has been submitted to the IEEE Robotics and Automation Letters (RA-L) with the ICRA conference option and is currently under review.

7.3. Representation Learning

Participants: David Filliat [correspondant], Celine Craye, Yuxin Chen, Clement Masson, Adrien Matricon, Freek Stulp.

7.3.1. Incremental Learning of Object-Based Visual Saliency

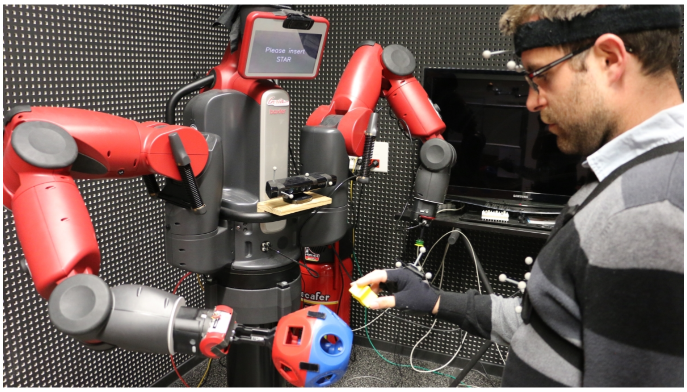


Figure 16. Representation of the setup considered in the user study. The robot presents to the user a spherical ball in which multiple shapes can be inserted. Final pose of the object is calculated from the user posture at his current location. Body motions during the insertion are recorded using a suit made from OptiTrack markers.

Searching for objects in an indoor environment can be drastically improved if a task-specific visual saliency is available. We describe a method to learn such an object-based visual saliency in an intrinsically motivated way using an environment exploration mechanism. We first define saliency in a geometrical manner and use this definition to discover salient elements given an attentive but costly observation of the environment. These elements are used to train a fast classifier that predicts salient objects given large-scale visual features. In order to get a better and faster learning, we use intrinsic motivation to drive our observation selection, based on uncertainty and novelty detection. Our approach has been tested on RGB-D images, is real-time, and outperforms several state-of-the-art methods in the case of indoor object detection. We published these results in two conferences [78],[77].

7.3.2. Cross-situational noun and adjective learning in an interactive scenario

Learning word meanings during natural interaction with a human faces noise and ambiguity that can be solved by analysing regularities across different situations. We propose a model of this cross-situational learning capacity and apply it to learning nouns and adjectives from noisy and ambiguous speeches and continuous visual input. We compared two different topic models for this task: Non Negative Matrix Factorization and Latent Dirichlet Association. We present experiments on learning object names and color names showing the performance of these model on realistic data and show how active learning can be used to speed-up learning by letting the learner choose the objects to be described. We published these results in a conference paper [75]

7.3.3. Learning representation with gated auto-encoders

We investigated algorithms that would be able to learn relevant visual or multi-modal features from data recorded while the robot performed some task. Representation learning is a currently very active research field, mainly focusing on deep-learning, which investigates how to compute more meaningful features from the raw high dimensional input data, providing a more abstract representation from which it should be easier to make decision or deduction (e.g classification, prediction, control, reinforcement learning). In the context of robotics, it is notably interesting to apply representation learning in a temporal and multi-modal approach exploiting vision and proprioception so as to be able to find feature that are relevant for building models of the robot itself and of its actions and their effect on the environment. Among the many existing approaches, we decided to explore the use of gated auto-encoders [104], a particular kind of neural networks including multiplicative connections, as they seem well adapted to this problem. Preliminary experimentations have been carried out with gated auto-encoders to learn transformations between two images. We observed that

Gated Auto-Encoders (GAE) can successfully find compact representations of simple transformations such as translations, rotation or scaling between two small images. This is however not directly scalable to realistic images such as ones acquired by a robot's camera because of the number of parameters, memory size and compational power it would require (unless drastically downsampling the image which induces sensible loss of information). In addition, the transformation taking an image to the next one can be the combination of transformations due to the movement of several object in the field of view, composed with the global movement of the camera. This induces the existence of an exponential number of possible transformations to model, for which the basic GAE architecture is not suited.

7.3.4. Incremental Learning in high dimensions

Participants: Alexander Gepperth [correspondant], Cem Karaoguz.

7.3.4.1. Incremental learning in data spaces of high dimensionality

Currently existing incremental learning algorithms in robotics have achieved a relatively high degree of usability due to the reduction of free model parameters in such approaches LWPR. Indeed, such algorithms are usually applied to low-dimensional tasks such as graspin with very good success, as the incremental learning paradigm is very appropriate to the robotics domain in general, especially in interactive scenarios. On the other hand, the partitioning of input space that is performed by LWPR and related approaches fails to be applicable if data dimension exceeds 50 elements since the used covariance matrices grow quadratically in size w.r.t. data dimensionality. Therefore, especially the incremental treatment of visual information is difficult, particularly for recognition and classification of objects or obstacles in general. To remedy this, we developed the incremental learning algorithm PROPRE [130] of fixed model complexity that can easily deal with data dimensionalities of 10.000 and beyond, where the only assumption is the same that is explicitly made for LWPR: that the data has structure, i.e., lies on a low-dimensional sub-manifold. We demonstrated the feasibility of the algorithm on several realistic datasets, on the one hand MNIST and on the other hand a much more challenging visual pedestrian pose recognition task from the intelligent vehicle domain[65].

7.3.4.2. Incremental learning with two memory systems

In order to increase PROPRE's ability to react quickly to changes in data statistics (e.g., a newly added visual class) while at the same time avoiding fast forgetting, a second, short-term memory system was proposed for PROPRE in [65]. This short-term memory is filled when task failures occur and is used to re-train the incremental long-term memory at a later time and on a slower time scale. In this way, abrupt changes in data statistics maybe immediately reacted upon, whereas the long-term memory can retain its stability that ensures that any forgetting happens gradually, on a determined time scale.

7.3.4.3. Steps towards incremental deep learnig

Since PROPRE is a neural architecture with just one hidden layer, its capacity is limited. This is why steps were taken to create deeper hierarchies with PROPRE in afashion totally analogous to current deep learning approaches. First of all, it was shown that a deep PROPRE architecture can achieve the same classification accuracy on MNIST as a shallow one but at a significantly lower computational cost [86]. Furthermore, it was shown that a deep PROPRE architecture is capable of change detection at multiple levels, a prerequisite for incremental learning [87]. Next steps will consist of creating regular deep PROPRE architectures and testing them on curently accepted machine learning benchmark tasks.

7.3.4.4. Real-world application of incremental learning

In [88], the incremental PROPRE algorithm was applied to object recognition and detection problems in the domain of intelligent vehicles. I was shown that, by re-casting pedestrian detection as an incremental learning problem where the background class is added only after learnig the pedestrian class, the number of required model resources for representing the background is reduced, and better accuracy can be obtained.

7.3.5. Measuring Uncertainty in Deep Learning Networks

Participants: Florian Golemo [correspondant], Manuel Lopes.

As precursor to the main objective of the IGLU project, we investigated methods that would enable deep neural networks to judge their knowledge about a domain.

Neural networks, especially deep ones, have been shown to be able to model arbitrarily complex problems, and thus offer powerful tools for machine learning. Yet they come with a significant flaw of not being inherently able to represent certainty of their predictions. By adding a measure of uncertainty to neural networks, this technology could be applied to autonomous exploration and open-ended learning tasks.

Thus the goal of this project was to find a method to measure how much knowledge a neural network has about about an unlabeled data item (measure of uncertainty), and to apply this new measure in an active learning context. The objective of the latter was to demonstrate the efficiency in handpicking interesting data, to optimally extend the system's own capabilities.

We were successful in finding a measure of uncertainty that would reliably distinguish data that the network has seen before, from data that was generally unfamiliar to the network. This measure was created by measuring the entropy of the network's last layer across a batch of stochastic samples generated by adding Poisson noise to the inputs.

The measure failed however to outperform random sampling in several active learning scenarios. Yarin Gal published related work as part of his dissertation [129] after this project was concluded. He elaborated that deep neural networks are very effective in canceling out input noise. The author suggested to use existing "Dropout" layers instead for stochastic sampling, but he reaches the same conclusion of using the last layer entropy as measure of uncertainty.

7.3.6. Learning models by minimizing complexity

We introduce COCOTTE (COnstrained Complexity Optimization Through iTerative merging of Experts), an iterative algorithm for discovering discrete, meaningful parameterized skills and learning explicit models of them from a set of behaviour examples. We show that forward-parameterized skills can be seen as smooth components of a locally smooth function and, framing the problem as the constrained minimization of a complexity measure, we propose an iterative algorithm to discover them. This algorithm fits well in the developmental robotics framework, as it does not require any external definition of a parameterized task, but discovers skills parameterized by the action from data. An application of our method to a simulated setup featuring a robotic arm interacting with an object is shown. This work was published in a conference paper [83]

7.4. Applications for Robotic myoelectric prostheses: co-adaptation algorithms and design of a 3D printed robotic arm prosthesis

Participants: Pierre-Yves Oudeyer [correspondant], Manuel Lopes, Mathilde Couraud, Sebastien Mick, Aymar de Rugy, Daniel Cattaert, Florent Paclet.

Together with the Hybrid team at INCIA, CNRS, the Flowers team continued to work on establishing the foundations of a long-term 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. In particular, in 2016 two lines of work were achieved, concerning two important scientific challenges, and in the context of one PEPS CNRS projects:

First, a new version of the experimental setup was designed to allow fast prototyping of 3D printed robotic prostheses. This work was based on the use of the Poppy open-source modular platform, and resulted in a functional prototype. A video demonstrations is available at: https://github.com/s-mick

Second, we have designed various control models allowing to transform signals coming from the human arm (either measured through EMGs or direct force sensors) and we have studied the influence of control modes on usability in the operation of a robotic arm prosthesis. In this context, we designed an experimental framework centered on a target-reaching task, and carried out tests with healthy subjects. The usability assessment relies on performance metrics on one hand, and a post-experiment questionnaire on another hand, in order to explore the multiple dimensions of the system's usability rather than focus only on measurements evaluating skills and performances. The code associated to this experimental setup is open-source and available at https://github.com/s-mick.

7.5. Applications for Educational Technologies

7.5.1. Multi-Armed Bandits for Adaptive Personalization in Intelligent Tutoring Systems

Participants: Manuel Lopes [correspondant], Pierre-Yves Oudeyer, Didier Roy, Alexandra Delmas, Benjamin Clement.

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 present 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 introduce 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 is evaluated in a scenario where 7-8 year old schoolchildren learn how to decompose numbers while manipulating money. Systematic experiments are presented with simulated students, followed by results of a user study across a population of 400 school children. [14]

7.5.1.2. A Comparison of Automatic Teaching Strategies for Heterogeneous Student Populations

Online planning of good teaching sequences has the potential to provide a truly personalized teaching experience with a huge impact on the motivation and learning of students. In this work we compare two main approaches to achieve such a goal, POMDPs that can find an optimal long-term path, and Multi-armed bandits that optimize policies locally and greedily but that are computationally more efficient while requiring a simpler learner model. Even with the availability of data from several tutoring systems, it is never possible to have a highly accurate student model or one that is tuned for each particular student. We study what is the impact of the quality of the student model on the final results obtained with the two algorithms. Our hypothesis is that the higher flexibility of multi-armed bandits in terms of the complexity and precision of the student model will compensate for the lack of longer term planning featured in POMDPs. We present several simulated results showing the limits and robustness of each approach and a comparison of heterogeneous populations of students.

This work has been publish and presented at Educational Data Mining 2016 conference in Raleigh, USA [76].

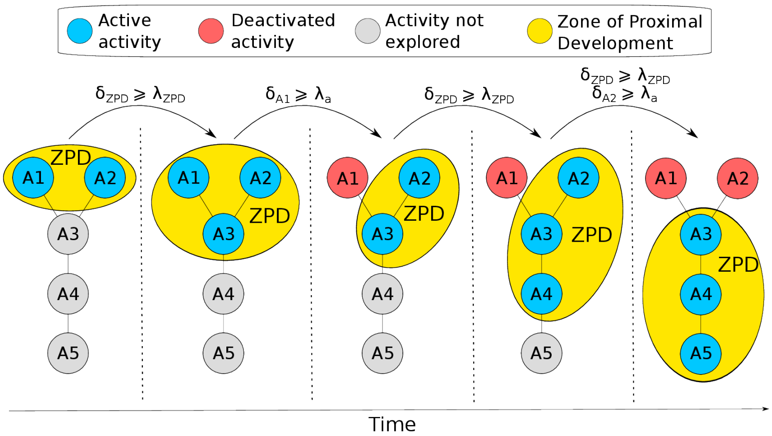


Figure 17. ZPDES exploration of an activity graph, with δ_{ZDP} the success rate over all active activities, λ_{ZPD} the threshold to expand the ZPD, δ_{Ax} the success rate for the activity Ax, and λ_a the threshold to reach to deactivate an activity.

Github link of the experiments paper code : https://github.com/flowersteam/kidlearn/tree/edm2016

7.5.1.3. The KidBreath project

To create learning contents linked to asthma to personalize it like mathematics activities in Kidlearn project [14] we used recommandation criterias in Therapeutic Education Program for asthma kids made by Health High Autority. Following an approach of participatory design [114], contents were validated by medical experts like health educators, pulmonoligists and pediatrics. Then, we conducted a workshop with forty kids aged 8 in order to iterate over the application interfaces and evaluate enjoy about it with observations. Finally, we realized a focus group with 5 asthma kids to validate the global comprehension of a part of the content. It revealed that children wanted more contents about the crisis treatment and how the asthma works in the human system (verbatims).

In a preliminary study, we experimented two conditions in 20 control children (with 3 asthma kids), one giving the possibility of choosing activities like the child wants, and one no giving this choice (activities displayed in random). No significative difference between the two groups, but results showed KidBreath was easy to use with scores > 75 using System Usability Scale [115]. Based on Cordova and Lepper works to evalueate motivation and knowledge with similar system and population [121], children had their disease knowledge increased with just one week use and were motivated using it. Finally, asthma kids showed they were more engaged than healthy kids and used KidBreath more seriously (stayed in breaks). These results was presented in the 5th edition of Serious Games in Medicine Conference in Nice.

We presented Thesis project in some events this year, with one publication surbmitted and validated:

- 2nd Meeting for Aquitaine and Euskadi companies in Biology and Health Between, February 11th 2016 in San sebastian (poster),
- Hackathon of innovation in pulmonary diseases, Respirhacktion, September 16th to 18th 2016 in Paris (project development in hackathon),
- 5th Conference in Health Ergonomics and Patient Safety, October 5th to 7th 2016 (poster) [102],
- Learning Lab day, November 16th in Inria Paris (oral presentation of project),
- 5th edition of Serious Games in Medicine Conference, December 2nd to 3rd 2016 in Nice Sofia-Antipolis University (oral presentation).

7.5.2. Poppy Education: Designing and Evaluating Educational Robotics Kits

Participants: Pierre-Yves Oudeyer [correspondant], Didier Roy, Théo Segonds, Stéphanie Noirpoudre, Marie Demangeat, Thibault Desprez, Matthieu Lapeyre, Pierre Rouanet, Nicolas Rabault.

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.

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, engineer schools, ...). For secondary education and higher education, scientific literacy centers, Fablabs.

Poppy Education is based on the robotic platform poppy (open-source platform for the creation, use and sharing of interactive 3D printed robots), including:

- Poppy Humanoid, a robust and complete robotics platform designed for genuine experiments in the real world and 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. 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, which is a variant of Scratch. Its features allow a thorough introduction of information technology.
- 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. It is possible in the classroom to work on the simulated model and then allow students to run their program on the physical robot.

7.5.2.1. Pedagogical experimentations : Design and experiment robots and the pedagogical activities in classroom

This project is user centred design. The pedagogical tools of the project (real and virtual robots, pedagogical activities, etc.) are being created directly with the users and evaluated in real life by experiments. For our experimentations in the classroom we are mainly using the robot Poppy Ergo Jr (real and virutal) and Snap! Our purpose is to improve this pedagogical tools and to create pedagogical activities and resources for teachers.

- A pedagogical working group:

  At the beginning of the project, we established a pedagogical working group of 12 volunteers, teachers from different level (mainly high school teachers of the Aquitaine region) to help to design educational activities in line with the needs of the school curriculum and to test them in the classroom.

  At the beginning of the second year of the project we added 7 other teachers from different background (middle-school and high school teachers) into the group to add more diversity.

  We organised some training to help them to discover and learn how to use the robotics platform, then we met monthly to exchange about the project and to get some feedbacks from them.

  You can see the videos of pedagogical robotics activities here:
  https://www.youtube.com/playlist?list=PLdX8RO6QsgB7hM_7SQNLvyp2QjDAkkzLn

- Experiment and Evaluate the pedagogical kits:

  Some engineer of the Poppy Education team went to visit the teachers in their school to see and to evaluate the pedagogical tools (robot and activities) in real contexts of use.

Figure 18. Experiment robots and pedagogical activities in classroom

In addition to the observations in classroom, two trainee students of Master 2 in cognitive sciences (M. Demangeat, D. Thibaut) have established an experimental protocol to evaluate the utility and the integration of the pedagogical kits in class. They created and filled out questionnaires by teachers and students. The analyzes of the results are presented in their paper thesis.

This experimentations are helping us to understand the educational needs, to create and improve the pedagogical tools.

7.5.2.2. 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.

  Last year student of "bachelor degree" (ENSAM-Talence) have designed, manufactured, assembled and programmed 4 different solutions to replace the fixed hand of Poppy by a gripper device: https://www.youtube.com/watch?v=DZjGaJk2fQk

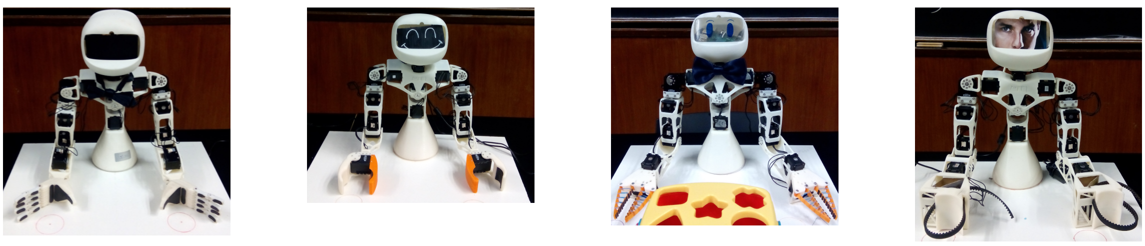


Figure 19. 4 grippers hands designed by students

- Audiovisual Students project

  Students from the BTS audiovisual of Saint-Genes La Salle have created a complete video report on the Poppy project to highlight the use in education and art:
  https://www.youtube.com/watch?v=NMwwH7AWO2Q

- Poppy entre dans la danse (Poppy enters the dance)

  The project "Poppy enters the dance" (Canope 33) uses the humanoid robot Poppy, 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 an overview of the project with kindergarten students :
  https://www.youtube.com/watch?v=XB9IXwcfJo0

7.5.2.3. Created pedagogical documents and resources

- Rebuilt the documentation of Poppy-project

  It was necessary for us to have an accessible and clear documentation to help teachers to use and create projects with the robots in the classroom so we rebuilt the existing documentation of the robotics platform Poppy. We added and improve the contents and we used the platform gitbook :
  https://docs.poppy-project.org/en/

- Pedagogical booklet

  The pedagogical booklet [96] brings together all the pedagogical activities and project testing in the classroom. It provides guided activities, small challenges and projects to become familiar with the Poppy Ergo Jr robot and the Programming language Snap!
  https://drive.google.com/file/d/0B2jV8VX-lQHwTUxXZjF3OGxHVGM/view

Figure 20. Pedagogical booklet: learn to program the robot Poppy Ergo Jr in Snap!

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

- Guide on the pedagogical use of the kit Poppy Ergo Jr in classroom

  We wrote an article [95] to explain how to use the Robot Ergo Jr in a classroom. It includes a summary of the characteristics of the robots, activities example and give all the necessary resources: https://pixees.fr/dans-la-famille-poppy-je-voudrais-le-robot-ergo-jr
- Demonstration guide to introduce the project

  This document is for people who already have a little experience with the Poppy Ergo Jr robot and snap! and wishing to present the project (i.e: to a colleague/acquaintance, on a exhibition stand, during a conference).

  The purpose of this document is to provide the necessary elements to enable the Poppy Education project to be presented through the use of Poppy Ergo Jr. robot. The key points of the Poppy Education project and the features of Poppy Ergo Jr kit are presented as well as examples of demonstrations of educational activities (videos and snap! projects) and educational projects (videos). An example of structuring a demo is provided at the end of the document.
  https://forum.poppy-project.org/t/guide-de-demo-du-kit-pedagogique-poppy-ergo-jr-version-beta/2698
- Model of pedagogical activities sheet

  We designed a model of pedagogical activity sheet. It helps us to get back the various activities and allows to have a homogeneous presentation. It is simpler to share and get back the creations of each.
  https://forum.poppy-project.org/t/modele-de-fiche-pedagogique-telechargeable-pour-les-activites-robotiques/2706

7.5.2.4. Scientific mediation

To promote educational uses of the platform, we participated in events (conference, seminar etc.).

We participated as well at some workshops to introduce students to robotics and programming.

7.5.2.5. Symposium robotics

We organized a symposium robotics (http://dm1r.fr/colloque-robotique-education/) that present research results and feedback on the use of Poppy and Thymio robots in education (other robots have been discussed, such as BeeBot and Metabot), from kindergarten to higher education, The Centers for Scientific and Technical Culture.

It was a 2 day event : 200 participants, 40 speakers (conferences and workshops).

Poppy Education team and the working group teachers helped with the organisation of the event and during the event (talk and workshops).

All conference videos are available on the web :
https://www.youtube.com/watch?v=prFmC-BpdY8&index=1&list=PL9T8000j7sJBC_H3L_hS-i4Ltlh1Fz2FY

7.5.3. IniRobot: Educational Robotics in Primary Schools

Participants: Didier Roy [correspondant], Pierre-Yves Oudeyer.

IniRobot (a project done in collaboration with EPFL/Mobsya) aims to create, evaluate and disseminate a pedagogical kit which uses Thymio robot, open-source and low cost, for teaching computer science and robotics.

IniRobot Project consists to produce and diffuse a pedagogical kit for teachers and animators, to help 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 enquiry-based educational approach, where kids are led to construct their understanding through practicing an active investigation methodology within teams. It is based on the use of the Thymio II robotic platform. More details about this projects were published in RIE 2015 [50] , which presents the detailed pedagogical objectives and a first measure of results showing that children acquired several robotics-related concepts. See also http://www.inirobot.fr.

Deployment: After 24 months of activity, IniRobot is used by about 1400 adults and 16 000 children in 54 cities of France. Example of action in university: MEEF teacher training for the hope of Aquitaine. Example of action in school: training of all Gironde Pedagogical ICT Advisors, covering nearly 1000 schools. Example of action in the extracurricular time: training 82 facilitators TAP cities of Talence, Pessac, Lille, ..., CDC Gates of inter-seas. Example of national action: Training of the digital mediators of the 8 Inria centers.

7.5.3.1. Partnership

The project is carried out in main collaboration with the LSRO Laboratory from EPFL (Lausanne) and others collaborations with French National Education/Rectorat d'Aquitaine, with Canopé Educational Network, with ESPE (teacher's school) Aquitaine, ESPE Martinique, ESPE Poitiers, National Directorate of Digital Education

7.5.3.2. Created pedagogical documents and resources

- IniRobot pedagogical kit [94]: This pedagogical booklet provides activities scenarized as missions to do. A second pedagogical booklet has been also created by three pedagogical advisers for primary school, with pedagogical instructions and aims, under ou supervision. http://tice33.ac-bordeaux.fr/Ecolien/ASTEP/tabid/5953/language/fr-FR/Default.aspx A new pedagogical kit is in progress, Inirobot Scratch, which will propose activities with Scratch and Snap! and Thymio robot.
- Inirobot website and forum http://www.inirobot.fr With this website, teachers, animators and general public can download documents, exchange about their use of inirobot' kit.
- Publication about Inirobot and Poppy Education A poster and talk were produced in Didapro-Didastic 6 Conference in Namur (Belgium) on 2016 January. [99]

7.5.3.3. Scientific mediation

Inirobot is very popular and often presented in events (conferences, workshops, ...) by us and by others.

7.5.3.4. Symposium robotics

With Poppy Education, Inirobot is a main line in our colloquium "Robotics and Education" (http://dm1r.fr/colloque-robotique-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[12]. For control it is essential to determine the current pose of the robot for given cable lengths (forward kinematics, FK) and to be able to calculate the cable lengths for a given pose of the platform (inverse kinematics, IK). If the cables are supposed to be non-deformable the IK problem is trivial and has a single solution but the FK is complex, admits several solutions and raises several issues. We have shown in the past that to get all FK solutions for a CDPR with m cables we have to consider not only the case where all cables are under tension but also have to solve the FK for all combinations of cables under tension with 1 to m cables. Surprisingly the FK is more difficult if the CDPR has less than 6 cables under tension. Our team, in collaboration with M. Carricato of Bologna University, is the first to have designed a solving algorithm that allow to compute in a guaranteed manner all FK solutions while a theoretical approach has allowed us to provide a bound for the maximal number of solutions according to the number of cables under tension (respectively 24, 156, 216, 140 and 40 for 2, 3, 4, 5, 6 cables).

Even more complex kinematic problems are involved if we assume that the cable are catenary-like, which is valid for large dimension robot, and involves inverse hyperbolic functions and square root, prohibiting to use algebraic geometry tools for estimating the maximal number of solutions and for the solving. In that case both the IK and FK may have multiple solutions and we have exhibited last year interval analysis-based solving algorithms for the IK and FK based on our interval analysis library `ALIAS`, that is the only existing algorithm for managing such complex cables. However such algorithm has the drawback, beside being computer intensive, to provide only solution(s) within a given search space for the unknowns. In our IK and FK problems two unknowns for each cable are the horizontal and vertical components F_x, F_z of the force exerted by the cables on the platform. In our case we have only the constraint $F_x > 0$ and F_z lower than half the mass of the cable but have no upper bound for F_x and lower bound for F_z. We may choose arbitrary large values for these bounds at the expense of an exponentially increasing computation time. As for the IK, beside F_x, F_z, the length of the cable at rest L_0 is an unknown with $L_0 > 0$ but no known upper bound. This year we have both improved the interval analysis algorithms but have also explored an original continuation scheme that be used both for the IK and FK whatever is the cable model. The idea is to assume that the cable model includes a set of physical parameters $\mathcal{P}$ which describe the elastic and deformation behavior of the cable material. We assume that their are limit values $\mathcal{P}_r$ for these parameters such that the cable behave like a non-deformable, non-elastic cable while the real cable parameter is $\mathcal{P}_d$. For example for catenary cables elasticity is defined by the Young modulus E of the cable material while the cable deformations is conditioned by its linear density μ. If E goes to infinity and μ to 0, then the cable is non-deformable, non-elastic. Now let us assume that we have a robot state for which the IK or FK are satisfied with the parameters $\mathcal{P}$. Assume that we modify $\mathcal{P}_d$ by a sufficient small amount ϵ toward $\mathcal{P}_r$ so that the Newton scheme allow us to determine the new robot state for $\mathcal{P} = \mathcal{P}_d + \epsilon$. Proceeding iteratively along this way will lead us to a robot state that must be very close to one obtained for non deformable, non elastic cables. Now we may revert the process: starting from all the IK or FK solutions obtained for non deformable cables (corresponding to $\mathcal{P} = \mathcal{P}_r$) we use Newton to compute a new robot state with $\mathcal{P}$ closer to $\mathcal{P}_d$ and doing that iteratively will lead to the solution(s) for $\mathcal{P} = \mathcal{P}_d$). We have also shown that a safe value of ϵ (ie. one that guarantee to obtain continuous solution without jump) may be calculated at each step by using the Kantorovitch theorem. We have implemented this principle for both the IK and FK problems (for 6 cables for the IK) and have found new IK and FK solutions which not been found previously because they were outside the search space of the interval analysis algorithms. A side benefit of this principle is that it has allowed us to be the first to provide an upper bound on the maximal number of solutions (63 for the IK of a robot with 6 cables, 33383 for the FK of a robot with 8 cables) whatever is the cable model. These new algorithms are much faster than the previous one (around one minute for the IK and 10 mn for the FK instead of several hours). However they raise a theoretical issue as the continuation scheme may lead to a solution that is close to be singular in which case the scheme cannot work. Understanding the singularity of the kinematics of CDPR is therefore a major problem. For the time being we mix the continuation scheme with the interval one that is basically used to solve the kinematic problem when the continuation scheme detect a singularity. As a test example we have considered a difficult CDPR with 8 cables and have shown a case with up to 41 solutions for the FK [10],[14],[11]. Figure 2 shows two of these solutions.

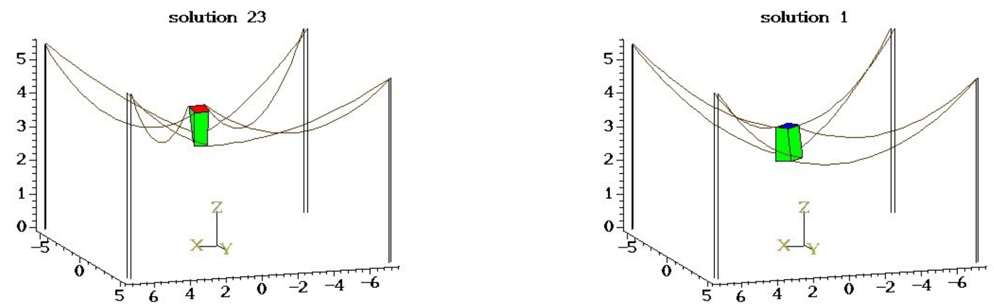


Figure 2. Two poses that are solutions of the forward kinematics, the left one being unstable while the right one is stable.

We have also investigated the calculation of cross-section of the workspace of CDPR [13]. We have shown that the border of this workspace for non deformable of purely elastic cables may be calculated rigorously by using an algorithm mixing a theoretical approach and numerical calculation. For catenary cables we have proposed a method that calculates a set of boxes that are guaranteed to lie in the workspace, getting smaller and smaller as soon as they are close to the border. Unfortunately this algorithm is highly computer intensive.

7.1.2. Cable-Driven Parallel Robots for additive manufacturing in architecture

Participant: Yves Papegay.

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 scale 1 and seems to be a good alternative to crane and industrial manipulators in the area of additive manufacturing.

Based on the proof of concept developed during the previous collaboration with CNAM and Ecole Nationale Supérieure d'Architecture Paris-Malaquais, the design of a new large scale CDPR for additive manufacturing of building based on ultra-high performance concrete has started under our supervision.

A new partnership with the the XtreeE start-up company aiming at developing a real size industrial 3D-printer of concrete has been established.

7.2. Assistance

7.2.1. Smart Environment for Human Behaviour Recognition

Participants: Mohamed Hedi Amri, Alain Coulbois, Artem Melnyk, Aurélien Massein, Yves Papegay, Odile Pourtallier [correspondante].

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.

This year we have concentrated our effort on the first step of this research that consists in being able to locate the person in his/her indoor environment.

A natural way of being able to adapt the accuracy of localization (and consequently accuracy of monitoring), is to use a partition of the monitoring area in a finite number of elementary zones ; the number of zones together with their geometry being closely related with the pursued level of monitoring. In practice these zones will be materialized by sensors barriers that detect the passage of a person from one zone to the other. Henceforth each zone are polygonal.

Several directions have been followed this year.

- monitoring system design,
- material development,
- data gathering and analysis,
- experimentation.

7.2.1.1. Monitoring system design

We aim at designing the partition of the monitoring space. Given the geometry of the monitoring area, the admissible position of the sensors barriers and a set of points of interest, the objective is to determine the positions of a minimal number of barriers such that each zone therefore defined includes at most one point of interest. The crossing of a given secession of barriers therefore allows to determine the trajectory of a person from one point of interest to another. An algorithm for solving this problem has been developed.

7.2.1.2. Material development

We initially used commercialized Infra Red barriers to detect the crossing time from one zone to an other. Nevertheless although the collected data is sufficient for the monitoring of a single person it prove not to be sufficient in a environment where there may be several persons, which is typically the case when considering retirement home for example.

Hence we have developed a multi-sensor barrier, a box containing two infra red distance sensors and two motion sensors (passive infrared type). It has been designed and created by 3D fast prototyping printer. The box is light, cheap and discreet. In addition to detecting the crossing time, it also gives the direction of crossing together with information about the speed and the size of the crossing person or object. This last information is helpful to differentiate for example a person using a wheelchair, a valid person (e.g medical staff), or an elderly.

We use phidget interface kits connected to a fit-pc for data acquisition and recording.

7.2.1.3. Data gathering and analysis

The aim of this data processing is to transform the raw data provided by the sensor barriers in a higher level data composed by the time and direction of crossing and rough estimation of the speed and size of the object or person crossing the barrier. This information can be deduced using only the data given by the distance sensors after processing. Nevertheless in real situations the barrier may be hidden by an object (food or cleaning trolley for example), and the redundant information from PIR sensors of an other closed barrier may be useful to recover the missing information.

The data are intended to be collected on large period of time (typically months). Inline filtering and averaging techniques were used to transform large and noisy raw data to get reasonable dataset size. Figure 3 shows in blue or red the general direction of the measurement of the stations (that create detection zones) and in each zone the current estimation of the number of people in each zone (a cross indicates 0, a black circle represents one person). For example the lower left zone has between 1 and 2 people.

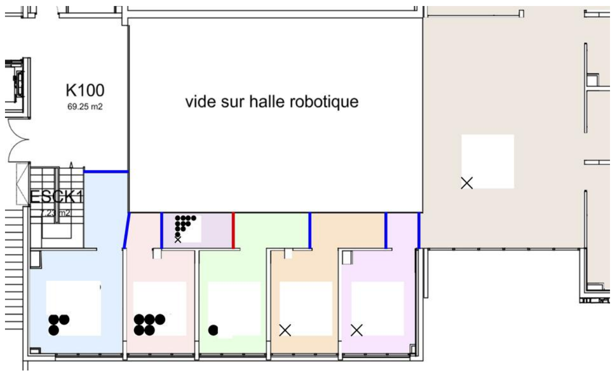


Figure 3. Occupancy of zone in a complex environment as measured by several stations

7.2.1.4. Experimentation

A monitoring system has been installed in the first floor of EHPAD Valrose in Nice. The area of monitoring was restricted to the hallway that leads to the individual rooms of six residents. Residents are proposed several activities (social or cultural activities, physical activities, meals) and have to use the hallway when participating to those activities. In addition to residents medical and service staff also use this hallway. The aim of this experiment is to determine an activity measure for each resident and to study its evolution with time. In that case the sensor placement is designed in such way that individual information may be obtained (e.g. by having stations on both side of the door of the individual room).

The installed system is composed of 10 multi sensor barriers installed on the wall of the hallway and 7 additional PIR sensors installed on the ceiling of the hallway. The data are transmitted by phidget interface kits and are processed by a fit-PC that store the daily data sets. A similar setup will be installed at the very beginning of 2017 in the Institut Claude Pompidou to monitor the activity in the corridors and in the waiting room. Here the medical community is more interested in statistical analysis than in individual analysis.

7.2.2. Sensors placement

Both economic motivations due to demographic evolution and willingness of people to live independently at home when aging, facing physical impairment or recovering from injuries has raised the need for activity monitoring at home, in rehabilitation center or in retirement home. Monitoring systems provide informations that can range from a broad measure of the daily activity to a precise analysis of the ability of a person performing a task (cooking, dressing, ...) and its evolution.

The broad range of needs and contexts, together with the large variety of available sensors implies the necessity to carefully think the design of the monitoring system. An appropriate system should be inexpensive and forgettable for the monitored person, should respect privacy but collect necessary data, and should easily adapt to stick to new needs. We aim to provide an assisting tool for designing appropriate monitoring systems.

As a second year of a PhD work, metrics have been defined to evaluate quality of sensors solutions and placement to infer people behaviors inside a smart environments. Based on these metrics, a methodology for optimal design of smart environments has been developed.

7.2.3. Rehabilitation

Participants: Alain Coulbois, Artem Melnyk, Jean-Pierre Merlet [correspondant].

We have developed the specific walking aid `ANG-med` to be used to monitor rehabilitation exercises beside performing analysis of walking pattern as any walker of the `ANG` family. The main addition for this walker are two rear looking distance sensors and two of such sensor mounted on a pan-tilt head (figure 4). These sensors have been placed under the guidance of the medical community in order to monitor and assess rehabilitation exercise such as leg flexion/extension/abduction and plantar flexion.

Figure 4. Rear view of the `ANG-med` walker with the 4 distance sensors that are used to monitor and assess rehabilitation exercises

The walker is since on year in test in the MATIA fundation in Spain. The software that is used to for this walker has been developed with the European RAPP project (see section 9.3.1.1) so that new exercise may easily be programmed and downloaded through a message passing system [9].

7.3. Miscellaneous results

7.3.1. Analysis of multi unit uniform price auction

Participant: Odile Pourtallier [correspondante].

From previous works on CO2 and electricity market we have identified relevant auction mechanism. This mechanism is strongly related with multi unit uniform price auction. In collaboration with M. Tidball (Lameta INRA) we study this mechanism using game theory models such as optimal stopping time game. The first results have been presented to the 17th ISDG conference (Urbino, Italy July 12-15 2016).

7.3.2. Symbolic tools for modeling and simulation

Participant: Yves Papegay.

This activity is the main part of a long-term ongoing collaboration with Airbus whose goal is to directly translate the conceptual work of aeronautics engineers into digital simulators to accelerate aircraft design.

An extensive modeling and simulation platform - MOSELA - has been designed which includes a dedicated modeling language for the description of aircraft dynamics models in term of formulae and algorithms, and a symbolic compiler producing as target an efficient numerical simulation code ready to be plugged into a flight simulator, as well as a formatted documentation compliant with industrial requirements of corporate memory.

Technology demonstrated by our prototype has been transferred : final version of our modeling and simulation environment has been delivered to Airbus in November 2012 and developer level know-how has been transferred in 2013 to a software company in charge of its industrialization and maintenance.

Since 2014, we are working on several enhancements and extension of functionalities, namely to enhance the performances and the numerical quality of the generated C simulation code, and ease the integration of our environment into the airbus toolbox.

In 2016, we have studied how to map modeling concepts used by other Airbus tools into our modeling concepts to allow import in MOSELA of existing models, and perform corresponding C generation [17].

LAGADIC Project-Team

7. New Results

7.1. Visual Perception

7.1.1. Micro/nano Manipulation

Participants: Le Cui, Eric Marchand.

Le Cui's Ph.D. [15] ended with a contribution related to visual tracking and estimation of the 3D pose of a micro/nano-object. It is indeed a key issue in the development of automated manipulation tasks using visual feedback. The 3D pose of the micro object can be estimated based on a template matching algorithm. Nevertheless, a key challenge for visual tracking in a scanning electron microscope (SEM) was the difficulty to observe the motion along the depth direction. We then proposed a template-based hybrid visual tracking scheme that uses luminance information to estimate the object displacement on x-y plane and uses defocus information to estimate object depth [54].

7.1.2. 3D Localization for Space Debris Removal

Participants: Aurélien Yol, Eric Marchand, François Chaumette.

This study is realized in the scope of the FP7 Removedebris project (see Section 9.3.1.1) [27]. We compared two vision-based navigation methods for tracking space debris in a low Earth orbit environment. The proposed approaches rely on a frame to frame model-based tracking in order to obtain the complete 3D pose of the camera with respect to the target [2]. The proposed algorithms robustly combine points of interest and edge features, as well as color-based features if needed. Experimental results have been obtained demonstrating the robustness of the approaches on synthetic image sequences simulating a CubeSat satellite orbiting the Earth [75].

7.1.3. 3D Localization for Airplane Landing

Participants: Noël Mériaux, François Chaumette, Patrick Rives, Eric Marchand.

This study is realized in the scope of the ANR VisioLand project (see Section 9.2.2). In a first step, we have considered and adapted our model-based tracker [2] to localize the aircraft with respect to the airport surroundings. Satisfactory results have been obtained from real image sequences provided by Airbus. In a second step, we are now considering to perform this localization from a set of keyframe images corresponding to the landing trajectory.

7.1.4. Scene Registration based on Planar Patches

Participants: Renato José Martins, Eduardo Fernandez Moral, Patrick Rives.

Image registration has been a major problem in computer vision over the past decades. It implies searching an image in a database of previously acquired images to find one (or several) that fulfill some degree of similarity, e.g. an image of the same scene from a similar viewpoint. This problem is interesting in mobile robotics for topological mapping, re-localization, loop closure and object identification. Scene registration can be seen as a generalization of the above problem where the representation to match is not necessarily defined by a single image (i.e. the information may come from different images and/or sensors), attempting to exploit all information available to pursue higher performance and flexibility. This problem is ubiquitous in robot localization and navigation. We propose a probabilistic framework to improve the accuracy and efficiency of a previous solution for structure registration based on planar representation. Our solution consists of matching graphs where the nodes represent planar patches and the edges describe geometric relationships. The maximum likelihood estimation of the registration is estimated by computing the graph similarity from a series of geometric properties (areas, angles, proximity, etc.) to maximize the global consistency of the graph. Our technique has been validated on different RGB-D sequences, both perspective and spherical [26].

7.1.5. Direct RGB-D Registration

Participants: Renato José Martins, Eduardo Fernandez Moral, Patrick Rives.

Dense direct RGB-D registration methods are widely used in tasks ranging from localisation and tracking to 3D scene reconstruction [7]. This work addresses a peculiar aspect which drastically limits the applicability of direct registration, namely the weakness of the convergence domain. In general, registration is performed only between close frames (small displacements), since dense registration tasks are particularly sensible to the local convexity of the cost error function. The main contribution of this work is an adaptive RGB-D error cost function that has a larger convergence domain and a faster convergence in both simulated and real data [67], [68]. This formulation employs the relative condition number metric to update the weighting of the RGB and depth costs. This approach is performed within a multi-resolution framework, where an efficient pixel selection for both RGB and ICP costs reduces the computational cost whilst preserving the precision. The formulation results in a larger region of attraction and faster convergence than classical RGB, ICP and RGB-D costs. Experiments was conducted using real sequences of indoor and outdoor images using perspective and spherical RGB-D sensors. Significant improvements was denoted in terms of the convergence stability and the speed of convergence in comparison with state-of-the-art methods.

7.1.6. Online localization and mapping for UAVs

Participants: Muhammad Usman, Paolo Robuffo Giordano, Eric Marchand.

Localization and mapping in unknown environments is still an open problem, in particular for what concerns UAVs because of the typical limited memory and processing power available onboard. In order to provide our quadrotor UAVs with high autonomy, we started studying how to exploit onboard cameras for an accurate (but fast) localization and mapping in unknown indoor environments. We chose to base both processes on the newly available Semi-Direct Visual Odometry (SVO) library (http://rpg.ifi.uzh.ch/software) which has gained considerable attention over the last years in the robotics community. The idea is to exploit dense images (i.e., with little image pre-processing) for obtaining an incremental update of the camera pose which, when integrated over time, can provide the camera localization (pose) w.r.t. the initial frame. In order to reduce drifts during motion, a concurrent mapping thread is also used for comparing the current view with a set of keyframes (taken at regular steps during motion) which constitute a "map" of the environment. We have started porting the SVO library to our UAVs and the preliminary results showed good performance of the localization accuracy against the Vicon ground truth. We are now planning to close the loop and base the UAV flight on the reconstructed pose from the SVO algorithm.

7.1.7. Reflectance and Illumination Estimation for Realistic Augmented Reality

Participants: Salma Jiddi, Eric Marchand.

The acquisition of surface material properties and lighting conditions is a fundamental step for photo-realistic Augmented Reality. Human visual cues remain sensitive to the global coherence within a computer-generated image. Absence or bad rendered virtual shadows, unconsidered specular reflections and/or occlusions, confused color perception such as an exuberantly bright virtual object are all elements which may not help an AR user interact and commit to a target application. In this work, we studied a new method for the estimation of the diffuse and specular reflectance properties of an indoor real static scene. Using an RGB-D sensor, we further estimate the 3D position of light sources responsible for specular phenomena and propose a novel photometry-based classification for all the 3D points. The resulting algorithm allows convincing AR results such as realistic virtual shadows as well as proper illumination and specularity occlusions [60].

7.1.8. Optimal Active Sensing Control

Participants: Paolo Salaris, Riccardo Spica, Paolo Robuffo Giordano.

This study concerns the problem of active sensing control. The objective is to improve the estimation accuracy of an observer by determining the inputs of the system that maximize the amount of information gathered by the outputs. In [9] this problem has been solved within the Structure from Motion (SfM) framework for 3D structure estimation problems, i.e. a point, a sphere and a cylinder, in the particular case where the observability property is instantaneously guaranteed. The optimal estimation strategy is hence given in terms of the instantaneous velocity direction of the camera velocities.

Recently, we have extended the optimal active sensing control to the case where the observability property is not instantaneously guaranteed. To simplify the analysis, we considered nonlinear differentially flat systems. Moreover, to quantify the richness of the acquired information the Observability Gramian (OG) has been used. We have hence defined a trajectory for the flat outputs of the system by using B-Spline curves and then, we have exploited an online gradient descent strategy to move the control points of such B-Spline in order to actively maximise the smallest eigenvalue of the OG over the whole fixed planning time horizon. While the system travels along its planned (optimized) trajectory, an Extended Kalman Filter (EKF) is used to estimate the system state. In order to keep memory of the past acquired sensory data for online re-planning, the OG is also computed on the past estimated state trajectories. This is then used for an online replanning of the optimal trajectory during the robot motion which is continuously refined by exploiting the estimated system state by the EKF. In order to show the effectiveness of our method we have considered a simple but significant case of a planar robot with a single range measurement. The simulation results show that, along the optimal path, the EKF converges faster and provides a more accurate estimate than along any other possible (non-optimal) paths. These results have been submitted to ICRA'2017.

7.2. Sensor-based Robot Control

7.2.1. Determining Singularity Configurations in IBVS

Participant: François Chaumette.

This theoretical study has been achieved through an informal collaboration with Sébastien Briot and Philippe Martinet from IRCCyN in Nantes, France. It concerns the determination of the singularity configurations of image-based visual servoing using tools from the mechanical engineering community and the concept of "hidden" robot. In a first step, we have revisited the welknown case of using three image points as visual feature, and then solved the general case of n image points [22]. The case of three image straight lines has also been solved for the first time [23].

7.2.2. Interval-based IBVS convergence domain computation

Participant: Vincent Drevelle.

This work aims to compute the set of camera poses from which IBVS will converge to the desired pose (that corresponds to the reference image). Starting from a (small) initial attraction domain of the desired pose (obtained using Lyapunov theory), we employ subpavings and guaranteed integration to iteratively increase the proven convergence domain, using a viability-based approach. Image-domain and pose-domain constraints are also enforced, like feature points visibility or workspace boundaries. First results have been obtained for a 3DOF line-scan camera IBVS case [56].

7.2.3. Visual Servoing of Humanoid Robots

Participants: Giovanni Claudio, Don Joven Agravante, Fabien Spindler, François Chaumette.

This study is realized in the scope of the BPI Romeo 2 and H2020 Comanoid projects (see Sections 9.2.7 and 9.3.1.2).

In a first step, we have considered classical kinematic visual servoing schemes for gaze control and manipulation tasks, such as can or box grasping. Two-hand manipulation has also been achieved using a master/slave approach [53], [81]. In a second step, we have designed the modeling of the visual features at the acceleration level to embed visual tasks and visual constraints in an existing QP controller [20][80]. Experimental results have been obtained on Romeo (see Section 6.9.4).

7.2.4. Model Predictive Visual Servoing

Participants: Nicolas Cazy, Paolo Robuffo Giordano, François Chaumette.

This study is realized in collaboration with Pierre-Brice Wieber, from Bipop group at Inria Rhône Alpes.

Model Predictive Control (MPC) is a powerful control framework able to take explicitly into account the presence of constraints in the controlled system (e.g., actuator saturations, sensor limitations, and so on). In this research activity, we studied the possibility of using MPC for tackling one of the most classical constraints of visual servoing applications, that is, the possibility to lose tracking of features because of occlusions, limited camera field of view, or imperfect image processing/tracking. The MPC framework depends upon the possibility to predict the future evolution of the controlled system over some time horizon, for correcting the current state of the modeled system whenever new information (e.g., new measurements) become available. We have also explored the possibility of applying these ideas in a multi-robot collaboration scenario where a UAV with a downfacing camera (with limited field of view) needs to provide localization services to a team of ground robots [13].

7.2.5. Model Predictive Control for Visual Servoing of a UAV

Participants: Bryan Penin, Riccardo Spica, François Chaumette, Paolo Robuffo Giordano.

Visual servoing is a welknown 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 underaction 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 are exploring the possibility of using techniques borrowed from Model-Predictive Control (MPC) for explicitly dealing with this kind of constraints during flight. Indeed, MPC is a class of (numerical) optimal control techniques able to explicitly take into account state and input constraints, as well as complex (and underactuated) nonlinear dynamics of the controlled system. In particular, the ability to predict, over some future time window, the behavior of the visual features on the image plane will allow the quadrotor to fly “blindly” for some limited phases, for then regaining tracking of any lost feature. This possibility will be crucial for allowing quick maneuvering guided by a direct visual feedback. We have started addressing the case of a simulated planar UAV as a representative case study, and we are now working towards an experimental validation with a real quadrotor UAV equipped with an onboard camera.

7.2.6. Visual-based shared control

Participants: Firas Abi Farraj, Nicolò Pedemonte, Paolo Robuffo Giordano.

This work concerns our activities in the context of the RoMaNS H2020 project (see Section 9.3.1.3 . Our main goal is to allow a human operator to be interfaced in an intuitive way with a two-arm system, one arm carrying a gripper (for grasping an object), and the other one carrying a camera for looking at the scene (gripper + object) and providing the needed visual feedback. The operator should be allowed to control the two-arm system in an easy way for letting the gripper approaching the target object, and she/he should also receive force cues informative of how feasible her/his commands are w.r.t. the constraints of the system (e.g., joint limits, singularities, limited camera fov, and so on).

We have started working on this topic by proposing a shared control architecture in which the operator could provide instantaneous velocity commands along four suitable task-space directions not interfering with the main task of keeping the gripper aligned towards the target object (this main task was automatically regulated). The operator was also receiving force cues informative of how much her/his commands were conflicting with the system constraints, in our case joint limits of both manipulators. Finally, the camera was always moving so as to keep both the gripper and the target object at two fixed locations on the image plane [46].

We have then extended this framework in two directions: first, by allowing the possibility of controlling a whole future trajectory for both arms (gripper+camera) while coping with the system constraints. The operator was then receiving an 'integral' force feedback along the whole planned trajectory: in this way, the operator's actions and the corresponding force cues were function of a planned trajectory (thus, carrying information over a future time window) that could be manipulated at runtime. Second, we studied how to integrate learning from demonstration into our framework by first using learning techniques for extracting statistical regularities of 'expert users' executing successful trajectories for the gripper towards the target object. Then, these learned trajectories were used for generating force cues able to guide novice users during their teleoperation task by the 'hands' of the expert users who demonstrated the trajectories in the first place. Both works have been submitted to ICRA'2017.

7.2.7. Direct Visual Servoing

Participants: Quentin Bateux, Eric Marchand.

In the direct visual servoing methods such as photometric framework, the images as a whole are used to define the control law. This can be opposed to the classical visual servoing approaches that relies on geometric features and where image processing algorithms that extract and track visual features are necessary. In [21], we proposed a generic framework to consider histograms as visual features. A histogram is an estimate of the probability distribution of a variable (for example the probability of occurrence in an intensity, color, or gradient orientation in an image). We demonstrated that the framework we proposed applies, but is not limited to, a wide set of histograms and allows the definition of efficient control laws.

Nevertheless, the main drawback for the direct visual servoing class of methods comparing to the classical geometric visual servoing methods is their comparatively limited convergence range. We then proposed in [48] a new direct visual servoing control law that relies on a particle filter to perform non-local and non-linear optimization in order to increase the convergence domain. To each particle considered we associate a virtual camera that predicts the image it should capture by using image transfer techniques. This new control law has been validated on a 6 DOF positioning task performed on our Gantry robot (see Section 6.9.1).

7.2.8. Audio-based Control

Participants: Aly Magassouba, François Chaumette.

This study is concerned with the application of sensor-based control approach to audio sensors. It is made in collaboration with Nancy Bertin from Panama group at Irisa and Inria Rennes-Bretagne Atlantique. Auditory features such as Interaural Time Difference (ITD), Interaural Level Difference (ILD), and sound energy have been modeled and integrated in various control schemes to control the motion of a mobile robot with two microphones onboard [66], [64]. Experiments with Romeo and Pepper (see Section 6.9.4) have also been achieved [65]. They show the robustness of closed loop sensor-based control with respect to coarse modeling and that explicit sound source localization is not a mandatory step for aural servoing.

7.3. Medical Robotics

7.3.1. Non-rigid Target Tracking in Ultrasound Images

Participants: Lucas Royer, Alexandre Krupa.

We pursued our work concerning the development of a real-time approach that allows tracking deformable soft tissue structures in 3D ultrasound sequences. In previous work we proposed a method which consists in estimating the target deformation by combining robust dense motion estimation and mechanical model simulation. This year we improved the robustness of our method to several image artefacts as the presence of large shadows, local illumination changes and image occlusions that occur due to the modification of the imaging gain and re-orientation of the ultrasound beam induced by probe motion. To achieve this, we proposed a new dissimilarity criterion between the current and reference images based on the Sum of Conditional Variance (SCV). Our new criterion, that we named Sum of Confident Conditional Variance (SCCV), consists in discriminating unconfident voxels thanks to the use of a pixel-wise quality measurement of the ultrasound images. This improved approach was experimentally validated on organic soft tissues and the obtained results were published in [40].

7.3.2. Optimization of Ultrasound Image Quality by Visual Servoing

Participants: Pierre Chatelain, Alexandre Krupa.

This study is realized in collaboration with Prof. Nassir Navab from the Technical University of Munich (TUM).

In previous work, we have developed ultrasound-based visual servoing methods to fulfill various tasks, such as compensating for physiological motion, maintaining the visibility of an anatomic target during ultrasound probe teleoperation, or tracking a surgical instrument. However, due to the specific nature of ultrasound images, guaranteeing a good image quality during the procedure remains a challenge. Therefore we pursued our study on the use of ultrasound confidence maps as a new modality for automatically positioning an ultrasound probe in order to improve the image quality. In addition to our visual servoing approach that optimizes the global quality of the image, this year we proposed a control fusion to optimize the acoustic window for a specific anatomical target which is tracked in the ultrasound images [50]. Recently, we extended our confidence-driven control to the out-of-plane motion of a 3D ultrasound probe and experimentally validated it on a human volunteer at TUM [14].

7.3.3. Visual Servoing using Shearlet Transform

Participants: Lesley-Ann Duflot, Alexandre Krupa.

In collaboration with the Femto-ST lab in Besançon, we proposed in a first-hand a solution to reduce the acquisition time of an Optical Coherence Tomography (OCT) 3D imaging scanner. This latter consists in sweeping a laser beam on a tissue sample of interest. To increase the frame rate of this imaging device we proposed to apply an optimal trajectory to the laser that covers entirely the image but without performing all the OCT measurements. The reconstruction of the missing data is then achieved by applying an updated Fast Iterative Soft-Thresholding Algorithm (FISTA) on a sparse representation of the image that is based on the shearlet transform [57]. In a second hand, we studied the feasibility of using the subsampled shearlet coefficients of an ultrasound image as the visual features of an image-based visual servoing. In a preliminary study we estimated numerically the interaction matrix that links the variation of the shearlet coarsest coefficients to the 6 degrees of freedom motion of the ultrasound probe and uses it in the visual servoing framework. The results obtained in cases of automatic probe positioning and phantom motion compensation demonstrated the efficiency of the shearlet-based features in terms of accuracy, repeatability, robustness and convergence behavior [59]. Then we proposed to consider a more efficient and adequate shearlet implementation that consists in a non-subsampled representation of the image. In this case the shearlet coefficients represent different images, focused on different singularities of the initial image, and we consider directly their pixel intensity values in the visual feature vector similarly to the photometry-based visual servoing approach. The modeling of the interaction matrix was analytically derived and experimental results demonstrated the reliability of the new method and its robustness to speckle noise [58].

7.3.4. 3D Steering of Flexible Needle by Ultrasound Visual Servoing

Participants: Jason Chevrie, Marie Babel, Alexandre Krupa.

The objective of this work is to provide robotic assistance during needle insertion procedures such as biopsy or ablation of localized tumor. In the past we only considered the control of the insertion and needle rotation along and around its main axis by the use of a duty-cycling control strategy. This latter consists in adapting online from visual feedback the orientation of a beveled-tip flexible needle during its insertion for controlling the needle curvature in 3D space that is induced by asymmetrical forces exerted on the bevel. However, such strategy limits the workspace of the needle tip. Therefore we proposed a new control method for flexible needle steering that combines direct base manipulation and needle tip based control. The direct base manipulation control is generated thanks to the use of a 3D model of a flexible beveled tip needle that gives the adequate motion of the needle base to obtain a given motion of the needle tip. This 3D model is based on virtual springs that characterize the needle mechanical interaction with soft tissue and is adapted online from visual tracking of the needle shape. From this model, a measure of the controllability of the needle tip degrees of freedom was proposed in order to mix the control between the direct base manipulation and the duty cycling technique

[51]. Preliminary results of an automatic needle tip positioning in a translucent gelatine phantom, observed by 2 orthogonal cameras, demonstrated the feasibility of the combination between direct base manipulation and needle tip control for reaching a desired target. This hybrid control allows better targeting capabilities in terms of larger needle workspace and reduced needle bending. In order to predict the trajectory of a needle during insertion under lateral motion of the tissue, we also improved our 3D model of the flexible needle to take into account the effect of the motion of the tissues on the needle shape. This was achieved thanks to the design of an algorithm based on an unscented Kalman filter that estimates the tissue motion. Results obtained from several needle insertions in a moving soft tissue phantom showed that our model gives good performance in terms of needle trajectory prediction. This model was also considered in a closed-loop control approach to allow automatic reaching of a target in case of tissue lateral displacement [52]. Future work will address the consideration of 3D ultrasound as visual feedback.

7.3.5. Enhancement of Ultrasound Elastography by Visual Servoing and Force Control

Participants: Pedro Alfonso Patlan Rosales, Alexandre Krupa.

Elastography imaging is performed by applying continuous stress variation on soft tissues in order to estimate a strain map of the observed tissues. It is obtained by estimating, from the RF (radio-frequency) signal along each scan line of the probe transducer, the echo time delays between pre- and post-compressed tissue. Usually, this continuous stress variation is performed manually by the user who manipulates the US probe and it results therefore in an user-dependent quality of the elastography image. To improve the ultrasound elastography imaging and provide quantitative measurement, we developed an assistant robotic palpation system that automatically moves a 2D ultrasound probe for optimizing ultrasound elastography [70]. The main originality of this work is the use of the elastography modality directly as input of the robot controller. Force measures are also considered in the probe control in order to automatically induce soft tissue deformation needed for real-time elastography imaging process.

7.4. Navigation of Mobile Robots

7.4.1. Visual Navigation from an Image Memory

Participants: Suman Raj Bista, Paolo Robuffo Giordano, François Chaumette.

This study is concerned with visual autonomous navigation in indoor environments. As in our previous works concerning navigation outdoors [4], the approach is based on a topological localization of the current image with respect to a set of keyframe images, but the visual features used for this localization as well as for the visual servoing are not composed of points of interest, but either on mutual information [71] following the idea proposed in [3], or straight lines that are more common indoors [38], or finally on a combination of points of interest and straight lines [11]. Satisfactory experimental results have been obtained using the Pioneer mobile robot (see Section 6.9.2).

7.4.2. Robot-Human Interactions during Locomotion

Participant: Julien Pettré.

In collaboration with the Gepetto team of Laas in Toulouse and the Mimetic group in Rennes, we have studied how humans avoid collision with a robot. Understanding how humans achieve such avoidance is crucial to better anticipate humans' reactions to the presence of a robot and to control the robot to adapt its trajectory accordingly. It is generally assumed that humans avoid a robot just like they avoid another human. In this work, we bring the empirical demonstration that humans actually set a specific strategy to avoid robots, and that, more precisely, they show a preference to give way to a robot which is on a collision course with them [41]. This results brings useful insight about human-robot interactions during locomotion, and provides useful guidelines to design reactive navigation techniques for mobile robots aimed at moving among humans.

7.4.3. Scene Mapping based on Intelligent Human/Robot Interactions

Participant: Patrick Rives.

For mobile robots to operate in compliance with human presence, interpreting the impact of human activities and responding constructively is a challenging goal. Towards this objective, mapping an environment allows robots to be deployed in diverse workspaces, marking this skill as a primary element in the integration of robots into human-populated environments. We proposed an effective approach for using human activity cues in order to enhance robot mapping and navigation and in particular in filtering noisy human detections, detecting passages, inferring space occupancy and allowing navigation within unexplored areas. Our contributions [36] are based on the development of intelligent interactions among conceptually different mapping levels, namely, the metric, social and semantic levels. Experiments, using the Hannibal platform (see Section 6.9.2), highlighted a number of strong dependences among these levels and the way in which they can be used to enhance individual performances and in turn the global robot operation.

7.4.4. Autonomous Social Navigation of a Wheelchair

Participants: Vishnu Karakkat Narayanan, Marie Babel.

This work is realized in collaboration with Anne Spalanzani (Chroma team - Inria Grenoble).

A key issue that hinders the adoption of assistive robotic technologies such as robotized wheelchair, in the real world, is that they need to operate in mostly human environments and among human crowds. Indeed intelligent wheelchairs need to be deployed in a human environment thereby making it essential for such robots to incorporate a sense of human-awareness. Simply put, humans are special objects that have to be perceived and acted on in a special manner by robots that interact with us humans. Thus one can define Human-aware Navigation as an intersection between human-robot interaction and robotic motion planning.

In this context we introduced a low-level velocity controller that could be employed by a social robot like a robotic wheelchair for approaching a group of interacting humans, in order to become a part of the interaction. Taking into account an interaction space that is created when at least two humans interact, a meeting point can be calculated where the robot should reach in order to equitably share space among the interacting group. We then proposed a sensor-based control law which uses the position and orientation of the humans with respect to the sensor as inputs, to reach the meeting point while respecting spatial social constraints [61]. Experiments using a mobile robot equipped with a single laser scanner, realized in collaboration with Ren Luo (Taiwan) within the Sampen Inria associated team, also proved the success of the algorithm in a noisy real world scenario [62].

In addition, a semi-autonomous framework for human-aware navigation in an intelligent wheelchair has been designed. A generalized linear control sharing framework was proposed that was able to progressively correct the user teleoperation in order to avoid obstacles and in order to avoid disturbance to humans. Meanwhile, we also proposed a Bayesian approach for user intention estimation. The formulation was partly inferred from the design of the controller for assisted doorway passing, wherein we hypothesized that predicting short term goals is sufficient for eliminating user intention uncertainty [16].

7.4.5. Semi-autonomous Control of a Wheelchair for Navigation Assistance

Participants: Louise Devigne, Vishnu Karakkat Narayanan, Marie Babel.

To address the wheelchair driving assistance issue, we proposed a unified shared control framework able to smoothly correct the trajectory of the electrical wheelchair [16]. The system integrates the manual control with sensor-based constraints by means of a dedicated optimization strategy. The resulting low-complex and low-cost embedded system is easily plugged onto on-the-shelf wheelchairs [55]. The robotic solution is currently under validation process with volunteering patients of Pôle Saint Hélier (France) who present different disabling neuro-pathologies preventing them to drive non-assisted wheelchairs.

Within the frame of ISI4NAVE associated team (see Section 9.4.1.2), this shared-control solution has been then coupled with first experimental biofeedback devices such as haptic devices. Preliminary tests have been conducted within the PAMELA facility at University College of London and within the rehabilitation center of Pôle Saint Hélier in Rennes (see Section 8.1.3). They involved regular wheelchair users as well as medical staff. We have demonstrated the ability of the framework to provide relevant assistance and now need to focus on methods to fine-tune parameters and customize/calibrate to the individual and evolving needs of each user.

7.5. Multi-robot and Crowd Motion Control

7.5.1. Advanced multi-robot control and estimation

Participant: Paolo Robuffo Giordano.

The challenge of coordinating the actions of multiple robots is inspired by the idea that proper coordination of many simple robots can lead to the fulfillment of arbitrarily complex tasks in a robust (to single robot failures) and highly flexible way. Teams of multi-robots can take advantage of their number to perform, for example, complex manipulation and assembly tasks, or to obtain rich spatial awareness by suitably distributing themselves in the environment. Within the scope of robotics, autonomous search and rescue, firefighting, exploration and intervention in dangerous or inaccessible areas are the most promising applications.

In the context of multi-robot (and multi-UAV) coordinated control, *connectivity* of the underlying graph is perhaps the most fundamental requirement in order to allow a group of robots accomplishing common goals by means of *decentralized* solutions. In fact, graph connectivity ensures the needed continuity in the data flow among all the robots in the group which, over time, makes it possible to share and distribute the needed information. We gave two contributions in this field: in the first one [35], we proposed a decentralized exploration strategy for a team of 3D agents able to guarantee exploration of a finite space in a finite amount of time while coping with the constraints of a connected sensing/communication graph for the robot group against sensing/communication constraints (limited range, occluded line-of-sight), and of obstacle and inter-robot collision avoidance. The strategy exploits a suitable state machine for assigning dynamic roles to the agents in the group for allowing completion of the exploration in finite time. Second, in [28] we studied how the choice of a leader agent in a leader-follower scenario could affect the performance of the group when tracking a desired formation (shape and gross motion). The proposed strategy allows selecting the “best leader” online as a function of the current group state (relative positions and velocities) and of the group topology (assumed connected). By cycling among several connected topologies during motion, we could show that our proposed leader selection algorithm provides the best performance among other possible choices (including the random one) while coping with the constraint of a connected (but possibly time-varying) topology.

These works were realized in collaboration with the robotics group at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, and the RIS group at Laas in Toulouse.

7.5.2. Rigidity-based methods for formation control

Participants: Fabrizio Schiano, Riccardo Spica, Andrea Peruffo, Paolo Robuffo Giordano.

Most multi-robot applications must rely on *relative sensing* among the robot pairs (rather than absolute/external sensing such as, e.g., GPS). For these systems, the concept of *rigidity* provides the correct framework for defining an appropriate sensing and communication topology architecture. Rigidity is a combinatorial theory for characterizing the “stiffness” or “flexibility” of structures formed by rigid bodies connected by flexible linkages or hinges. In a broader context, rigidity turns out to be an important architectural property of many multi-agent systems when a common inertial reference frame is unavailable. Applications that rely on sensor fusion for localization, exploration, mapping and cooperative tracking of a target, all can benefit from notions in rigidity theory. The concept of rigidity, therefore, provides the theoretical foundation for approaching decentralized solutions to the aforementioned problems using distance measurement sensors, and thus establishing an appropriate framework for relating system level architectural requirements to the sensing and communication capabilities of the system.

In the recent past, we have proposed a decentralized gradient-based rigidity maintenance action for a group of quadrotor UAVs [10]. By starting in a rigid configuration, the group of UAVs was able to estimate their relative position from sole relative distance measurements, and then use these estimated relative positions in a control action able to preserve rigidity of the whole formation despite presence of sensor limitations (maximum range and line-of-sight occlusions), possible collisions with obstacles and inter-robot collisions. This (rigidity-based) control/estimation framework has now been extended to the case of *bearing rigidity* for directed graphs: here, rather than distances the measurements are the 3D bearing vectors expressed in the local body-frame of each agent. The theory has been extended to the case of 3D agents evolving in $\mathbb{R}^3 \times \mathbb{S}^1$ by proposing a decentralized bearing controller/localization algorithm that only requires one single distance measurement (among an arbitrary pair of agents) for a correct convergence [72]. The proposed algorithm ensures stabilization towards a desired bearing formation, and allows for the possibility of actuating the motion directions in the null-space of the bearing constraints (that is, collective translations in 3D, expansion/retraction, and coordinated rotation about a vertical axis).

The need of a single distance measurement (for fixing the formation scale) has also been relaxed in [73] where an *active* scale estimation scheme has been proposed for allowing the (distributed) estimation of the various inter-agent distances online by processing the measured bearings and the known agent ego-motion (body-frame linear and angular velocities). Finally, we have also proposed an extension of the "distance" rigidity maintenance controller proposed in [10] to the case of bearing measurements (and bearing rigidity), by considering the typical sensing constraints of onboard cameras, that is, limited range, limited field of view, of possible mutual occlusions when two or more agents lie on the same line-of-sight. This work has been experimentally validated with 5 quadrotor UAVs, and has been submitted to ICRA'2017.

These works were realized in collaboration with the RIS group at Laas, Toulouse, and with Technion, Israel.

7.5.3. Cooperative localization using interval analysis

Participants: Ide Flore Kenmogne Fokam, Vincent Drevelle.

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. Ongoing work focuses on position uncertainty domain computation in image-based UAV localization [63], and its extension to cooperative localization in a multi-UAV fleet.

7.5.4. Numerical Models of Local Interactions during Locomotion

Participants: Julien Bruneau, Panayiotis Charalambous, David Wolinski, Julien Pettré.

The numerical models of local interactions are core components of reactive navigation techniques (which allows a robot to avoid dynamic obstacles) and of microscopic crowd simulation algorithms (which allows to simulate a crowd motion as a collection of agent trajectories). We have pursued our efforts to design local models of interactions which capture humans pedestrian behavior, to simulate how they adapt their trajectory so as to perform interactions with their neighbors [12]. This year, our efforts were focused on the simulation of grouping behaviors [39], and mid-term strategies human set to perform energy-efficient sequences of successive avoidance adaptations [24]. These two situations deal with complex situations of interactions, where several interactions of different kinds need to be combined to compute agents trajectories. For example, when moving in groups, agents have to keep close to the other members of their group while they should not collide with them, as well as they should avoid collision with any other agent or obstacle out of this group.

We also revisited the foundation of velocity-based models of local interaction for collision avoidance. Using a velocity-based model, a collision-free motion is computed for one agent by extrapolating the future motion of neighbor agents with respect to their current position and velocity. From this information, each agent can deduce the set of velocities (called admissible velocities) that lead to a collision-free motion in the near future. The extrapolation is generally simply based on a linear extrapolation of the future position along the current velocity vector. This is simplistic as it assumes that the current velocity vector is representative of the future motion, while it is often false when, for instance, the agent is currently performing adaptations due to ongoing collision avoidance, or when the agent is following a curvy path. To improve the accuracy of motion prediction and the resulting simulation, we have introduced a probabilistic representation of future position, that can be computed from a set of context elements such as the layout of the environment or the agents past motion [42]. We demonstrate in this work the high impact on the level of realism of resulting simulations. This work is implemented in the WarpDriver software (see Section 6.7).

Finally, we address applications of our simulators to the Computer Animation. Crowd simulation agents generally have a simplistic geometrical and kinematics models, typically, an oriented 2D circle moving on a flat surface. In Computer Animation, an animation of a crowd of 3D realistic characters can be computed on top of the agents simulation by computing their internal joints trajectories so as to perform walking motion along computed agents trajectories. However, the discrepancies between the 2D model of agents and 3D full body characters may result into residual collisions between character shapes. In this collaboration with the Mimetic team, we demonstrate that simple secondary animations for characters, such as local shoulder motions, can be efficiently triggered to camouflage those artefacts, with a very low computational overhead [29].

7.5.5. Motion Planning for Digital Characters

Participant: Julien Pettré.

Motion planning is an important component for agents and robot navigation and control, providing them the ability to perform geometrical reasoning over their environment to transform a high-level distant goal in their environment into a sequence of local motions and sub-goals to reach. This year, we have been involved into two collaborations dealing with motion planning. First collaboration was with the University of Utrecht in the Netherlands. We have proposed a method to evaluate and compare various environment decomposition techniques [74]. Environment decomposition is an important step to perform navigation planning in large static environments. Second collaboration was with the University of North Carolina in Chapel Hill (see Section 9.4.1.1). We have coupled a contact planner for virtual characters with ITOMP, a motion optimization technique to achieve complex motion in cluttered environment [69].

LARSEN Team

7. New Results

7.1. Lifelong Autonomy

7.1.1. PsyPhINe: Cogito Ergo Es

Participant: Amine Boumaza.

PsyPhINe is an interdisciplinary and exploratory project (see 8.1.1) between philosophers, psychologists and computer scientists. The goal of the project is related to cognition and behavior. Cognition is a set of processes that are difficult to unite in a general definition. The project aims to explore the idea of assignments of intelligence or intentionality, assuming that our intersubjectivity and our natural tendency to anthropomorphize play a central role: we project onto others parts of our own cognition. To test these hypotheses, our aim is to design a "non-verbal" Turing Test, which satisfies the definitions of our various fields (psychology, philosophy, neuroscience and computer science) using a robotic prototype. Some of the questions that we aim to answer are: is it possible to give the illusion of cognition and/or intelligence through such a technical device? How elaborate must be the control algorithms or "behaviors" of such a device so as to fool test subjects? How many degrees of freedom must it have?

Last year, an experimental robotic device was designed and built, and an experimental campaign with human subject was conducted. The experiments consisted in recording the interactions of the subjects with the robot when realizing a task. The results of the experiments are under analysis and will partly be presented at the second edition of the PsyPhINe workshop organized by the group, gathering top researchers from philosophy, anthropology, psychology and computer science to discuss and exchange on our methodology (see 9.1.1.1).

7.1.2. Localisation of robots on load-sensing floor

Participants: François Charpillet, Francis Colas, Vincent Thomas.

The use of floor-sensors in ambient intelligence contexts began in the late 1990's. We designed such a sensing floor in Nancy in collaboration with the Hikob company (http://www.hikob.com) and Inria SED. This is a load-sensing floor which is composed of square tiles, each equipped with two ARM processors (Cortex m3 and a8), 4 load cells, and a wired connection to the four neighboring cells. Ninety tiles cover the floor of our experimental platform (HIS).

This year, with Alexis Grall (master student from Enseirb-Matmeca), we have focused on identifying localisation and tracking scenarios involving several robots and on collecting data corresponding to instantiation of these scenarios. These data originated from the sensing tiles but also from Qualisys motion capture system in order to have information about ground-truth. We have also focused on basic algorithms (for instance, Kalman filter) to tackle the issue of tracking targets, but we plan to investigate more elaborate strategies for dealing with sensor discontinuity (for example, when the robot leaves or enters a tile) and multi-traget tracking (Joint Probability Data Association Filter algorithm [52]).

With Mohammad Rami Koujan, we also started to apply deep-learning techniques on those sequential data in order to compare model-based and model-free approaches. This work included some long-term data collection with a randomized behavior in order to have enough training data.

7.1.3. Active sensing and multi-camera tracking

Participants: Olivier Buffet, François Charpillet, Vincent Thomas.

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.

This problem we are focusing on consists of following the trajectories of persons with the help of several controllable cameras in the smart environment. This is a difficult problem since the set of cameras cannot simultaneously cover the whole environment, some persons can be hidden by obstacles or by other persons, and the behavior of each person is governed by internal variables which can only be inferred (such as his motivation or his hunger).

The approach we are working on is based on probabilistic decision processes in partial observability (POMDP - Partially Observable Markov Decision Processes) and particle filters. In the past, we have proposed an original formalism rho*-POMDP* and new algorithms for representing and solving active sensing problems [38] by tracking several persons with fixed camera based on particle filters and Simultaneous Tracking and Activity Recognition approach [45].

This year, we have focused on investigating the issue of solving the active sensing problem with controllable cameras. Approaches based on Monte-Carlo Tree Search algorithms (MCTS) like POMCP [54] are currently investigated for adressing the combinatorial explosion of the state space to consider (which is the space of probability distributions over all the possible states of the system).

7.1.4. Audio Source Localization

Participants: François Charpillet, Francis Colas, Van Quan Nguyen.

We collaborate on this subject with Emmanuel Vincent from the Multispeech team (Inria Nancy - Grand Est).

We considered, here, the task of audio source localization using a microphone array on a mobile robot. Active localization algorithms have been proposed in the literature that can estimate the 3D position of a source by fusing the measurements taken for different poses of the robot. A typical implicit assumption in the literature is that the sound source is active, but a lot of real sound sources are actually intermittent. Systems of activity detection exist but cannot reach perfect accuracy. In this work, we propose a new mixture Kalman filter that explicitly includes the discrete activity of the source in the estimated state vector, alongside the continuous states such as the position of the robot or the sound source. We take into account the imperfection of activity detection systems in order to show that we have better accuracy than the state of the art [26].

This work is led through the PhD Thesis of Van Quan Nguyen under the supervision of Emmanuel Vincent and Francis Colas.

7.1.5. Learning for damage recovery

Participants: Jean-Baptiste Mouret, Konstantinos Chatzilygeroudis, Vassilis Vassiliades, Dorian Goepp.

In 2015, we introduced a novel algorithm that allows robots to learn by trial-and-error when they are damaged [42]. In 2016, we extended this algorithm to make it easier to deploy it in real-life situations and real systems:

- We added "safety constraints" so that the learning algorithm both maximizes the post-damage performance and minimizes the probability of breaking the robot during the learning process; we demonstrated this extension with a simulation of the iCub robot, which is a fragile and expensive robot (around 250,000 euros) for which we would like to use our learning algorithms [33].
- We proposed a novel algorithm that does not require to reset the robot to its starting position between each trial [40], which allows the damaged robots to "learn while doing". We demonstrated this algorithm on our 6-legged walking robot.
- We extended the MAP-Elites algorithm, that is, the evolutionary algorithm that we use to generate prior probability distributions for our online learning algorithm, to scale-up to high-dimensional search spaces [59]. The algorithm is based on central Voronoi tesselations (CVT). In addition, we investigated the influence of the encoding (representation of the controller) on the performance of MAP-Elites [30].

7.1.6. Interactions with biology

Participant: Jean-Baptiste Mouret.

We continued our on-going collaborations with biologists.

- *The Evolutionary Origins of Hierarchy.* Hierarchical organization—the recursive composition of sub-modules—is ubiquitous in biological networks, including neural, metabolic, ecological, and genetic regulatory networks, and in man-made systems such as large organizations and the Internet. In this contribution, we showed that the pressure to minimize the connection costs in network can explain the evolution of hierarchical and modular biological networks [21]; this result extends our previous work on the evolutionary origins of modularity in biological networks [41]. (Collaboration with Jeff Clune, University of Wyoming, USA).
- *Animal-robot interaction.* We worked with a team based in Rennes to perform preliminary experiments about animal-robot attachment (here with a gallinaceous bird) [16].

7.1.7. Learning for whole-body motions

Participants: Serena Ivaldi, Valerio Modugno.

Within the European project CoDyCo, we studied how to combine learning, dynamics, and control for redundant robots. In [25], we proposed a framework to automatically optimize the evolution in time of soft task priorities for multi-task controllers. The motivation of the work was to propose a way to automatate the manual optimization procedure of task priorities and weights, that is classically done by control experts and is time consuming. In [24], we improved the framework by using constrained stochastic optimization algorithms to optimize the task priorities while ensuring that the system constraints (robot and problem setting) are never violated. We showed the results on our robot iCub. Our master student Ugo Chervet contributed to the simulations of this paper.

7.2. Natural Interaction with Robotic Systems

7.2.1. Human Activity recognition on load-sensing floor

Participant: François Charpillet.

In the framework of a collaboration with Lebanese University and CRIStAL laboratory, Lille, we have evaluated this year the capability of the load-sensing floor that we have designed in Nancy, to adress fall detection and activity recognition for elderly people living alone at Home.

The Inria-Nancy sensing floor consists of 104 tiles (60*60 cm). Each tile is equipped with a 3-axis accelerometer in the center of the tile, and four force sensors (strain gauge load cells) positioned at each corner.

The pressure sensors measure the load forces exerted on the floor that can be used to determine, for example, the center of pressure of objects, robots or human beeing on the floor.

This year we have demonstrated that we can also determine the posture or activiy of the monitored person (walking, sitting, standing, falling, etc.) by combining the pressure amount, the pressure duration on a tile, the 3-axis acceleration using a relatively simple algorithm [10], [11].

7.2.2. Human Activity recognition with depth camera

Participants: François Charpillet, Xuan Son Nguyen.

This year, we proposed a new local descriptor for action recognition in depth images. The proposed descriptor relies on surface normals in 4D space of depth, time, spatial coordinates and higher-order partial derivatives of depth values along spatial coordinates. In order to classify actions, we follow the traditional Bag-of-words (BoW) approach, and propose two encoding methods termed Multi-Scale Fisher Vector (MSFV) and Temporal Sparse Coding based Fisher Vector Coding (TSCFVC) to form global representations of depth sequences. The high- dimensional action descriptors resulted from the two encoding methods are fed to a linear SVM for efficient action classification. Our proposed methods are evaluated on two public benchmark datasets, MSRAction3D and MSRGesture3D. The experimental result shows the effectiveness of the proposed methods on both the datasets.

7.2.3. Human Posture Recognition

Participants: François Charpillet, Abdallah Dib, Alain Filbois, Thomas Moinel.

Human pose estimation in realistic world conditions raises multiple challenges such as foreground extraction, background update and occlusion by scene objects. Most of existing approaches were demonstrated in controlled environments. In this work, we propose a framework to improve the performance of existing tracking methods to cope with these problems. To this end, a robust and scalable framework is provided composed of three main stages. In the first one, a probabilistic occupancy grid updated with a Hidden Markov Model used to maintain an up-to-date background and to extract moving persons. The second stage uses component labelling to identify and track persons in the scene. The last stage uses a hierarchical particle filter to estimate the body pose for each moving person. Occlusions are handled by querying the occupancy grid to identify hidden body parts so that they can be discarded from the pose estimation process. We provide a parallel implementation that runs on CPU and GPU at 4 frames per second. We also validate the approach on our own dataset that consists of synchronized motion capture with a single RGB-D camera data of a person performing actions in challenging situations with severe occlusions generated by scene objects. We make this dataset available online (http://www0.cs.ucl.ac.uk/staff/M.Firman/RGBDdatasets/).

7.2.4. Evaluation of control interfaces by non-experts

Participants: Serena Ivaldi, François Charpillet.

In this work, we address the question of user preference for a robotic interface by non-experts (or naive users without training in robotics), after one single evaluation of such an interface on a simple task. This refers to situations when non-experts face the decision of adopting a robot for episodic use (i.e., not a regular continuous use as workers in factories): the ease of use of an interface is crucial for the robot acceptance. We also probe the possible relation between user performance and individual factors. After a focus group study, we chose to compare the robotic arm joystick and a graphical user interface. Then, we studied the user performance and subjective evaluation of the interfaces during an experiment with the robot arm Jaco and 40 healthy adults. Our results show that the user preference for a particular interface does not seem to depend on their performance in using it: for example, many users express their preference for the joystick while they are better performing with the graphical interface. Contrary to our expectations, this result does not seem to relate to the user's individual factors that we evaluate, namely desire for control and negative attitude towards robots.

The preliminary results of this work are published in [23]. A journal paper with the complete results is in preparation. The work was conducted with the master students Sebastian Marichal and Adrien Malaisé.

7.2.5. Robot acceptance and trust

Participant: Serena Ivaldi.

We continued our collaboration with psychologists.

- *Trust as a measure of robot acceptance*: together with the research group of Elisabetta Zibetti (Université de Paris 8), we proposed trust as a main indicator of acceptance in decision-making tasks characterized by perceptual uncertainty (e.g., evaluating the weight of two objects) and socio-cognitive uncertainty (e.g., evaluating which is the most suitable item in a specific context). We measured trust by the participants' conformation to the iCub's answers to specific questions. We found that participants conformed more to the iCub's answers when their decisions were about functional issues than when they were about social issues. Moreover, the few participants conforming to the iCub's answers for social issues also conformed less for functional issues. Trust in the robot's functional knowledge does not thus seem to be a pre-requisite for trust in its social knowledge. Finally, desire for control, attitude towards social influence of robots and type of interaction scenario did not influence the trust in iCub. The results have been published in [13].
- *Acceptance of assistance robots in EHPADs by professional caregivers*: together with Sophie Nertomb (Université de Lorraine), we started a dialogue with professional caregivers to probe their acceptance and positive/negative attitude towards an assistance robot as a collaborator in an EHPAD.

From the first focus group, we found that caregivers are rather positive in adopting a robot to get assistance in some daily tasks with the patients, and they would prefer a social robot such as Pepper rather than a functional robot arm, because they believe it could be more useful and would be better accepted by patients. The results of our preliminary investigation were presented in [22].

7.2.6. Individual factors and social/physical signals

Participant: Serena Ivaldi.

We finalized our study about the influence of individual factors in the production of social signals during human-humanoid interaction on a collaborative assembly task. We found that the more people are extrovert, the more and longer they tend to talk with the robot, and the more people have a negative attitude towards robots, the less they will look at the robot face and the more they will look at the robot hands where the assembly and the contacts occur. Our results confirm and provide evidence that the engagement models classically used in human-robot interaction should take into account attitudes and personality traits. The results are published in [15].

We started to study the influence of individual factors on physical signals and collaborative movement. We made interesting observations, for example the influence of age and negative attitude towards robots in the amount of exchanged forces. Part of the analysis was performed by the master student Anthony Voilqué. A paper describing our findings is in preparation.

7.2.7. Learning gait models with cheap sensors for applications in EHPADs

Participants: Serena Ivaldi, François Charpillet, Olivier Rochel.

Thanks to the MITACS-Inria grant, we started a collaboration with Prof. Dana Kulic in University of Waterloo on the topic of learning gait models with cheap sensors. Jamie Waugh, master student, visited us for 3 months to start a data collection protocol where several sensors are used to monitor the human gait under different conditions. The aim is to learn gait parameters with different sensors, such as IMUs and Kinect cameras, and to provide quantitative comparison of the accuracy of the estimation provided by the different sensors. As ground truth, the Qualisys motion capture and the Gaitrite walking mat are used. The final goal of the project is to be able to deliver algorithms for estimating gait based on cheap sensors that could be used on a daily basis in healthcare facilities such as EHPADs.

PERVASIVE INTERACTION Team

6. New Results

6.1. Simulating Haptic Sensations

Participants: Jingtao Chen, Sabine Coquillart, 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.

RITS Project-Team

6. New Results

6.1. Low Speed Vehicle Localization using WiFi-FingerPrinting

Participants: Dinh-Van Nguyen, Myriam Vaca Recalde, Fawzi Nashashibi.

Recently, the problem of fully autonomous navigation of vehicle has gained major interest from research institutes and private companies. In general, these researches rely on GPS in fusion with other sensors to track vehicle in outdoor environment. However, as indoor environment such as car park is also an important scenario for vehicle navigation, the lack of GPS poses a serious problem. In [39] we present an approach to use WiFi Fingerprinting as a replacement for GPS information in order to allow seamlessly transition of localization architecture from outdoor to indoor environment. Often, movement speed of vehicle in indoor environment is low (10-12km/h) in comparison to outdoor scene but still surpasses human walking speed (3-5km/h, which is usually maximum movement speed for effective WiFi localization). We propose an ensemble classification method together with a motion model in order to deal with the above issue. Experiments show that proposed method is capable of imitating GPS behavior on vehicle tracking.

6.2. Free navigation space estimation

Participants: Raoul de Charette, Rafael Colmenares Prieto, Alexis Meyer, Fawzi Nashashibi.

Autonomous vehicles need to know where they can physically drive. In the past, lane detection was used to bound the driving area of the vehicle but road markings do not exist in many urban scenario thus perception needs to estimate the free navigation space with other means.

To contrast with the state of the art two approaches were developed and will be published soon. The first approach is using a monocular setup and use an absurd logic to identify the flow of the scene and extract the ego motion. The second method still under research is to develop a hybrid approach to segment the navigation space using energy minimization to label the scene assuming learning on the go.

6.3. Pedestrian Recognition using Convolutional Neural Networks

Participants: Danut-Ovidiu Pop, Fawzi Nashashibi.

Pedestrian detection is of highly importance for a large number of applications, especially in the elds of automotive safety, robotics and surveillance. In spite of the widely varying methods developed in recent years, pedestrian detection is still an open challenge whose accuracy and robustness has to be improved. This year we focused on the improvement of the classification component in the pedestrian detection task by adopting two approaches: 1) by combining three image modalities (intensity, depth and ow) to feed a unique convolutional neural network (CNN) and 2) by fusing the results of three independent CNNs. The evaluations have been performed on the Daimler stereo vision data set.

6.4. Reliability estimation and information redundancy for accurate localization

Participants: Zayed Alsayed, Anne Verroust-Blondet, Fawzi Nashashibi.

Our goal is to improve localization systems performances in order to be able to navigate in urban and peri-urban environments. For this purpose, we choose to study the reliability of a SLAM method that incrementally builds a map of the surrounding environment from an information given by a set of 2D laser points.

This year, we focused on SLAM failure and non-failure scenarios.

- Experimental data acquired on the VEDECOM demonstrator in the context of ITS Bordeaux demonstrations in 2015 were analyzed. This evaluation showed in [30] that the SLAM concept seems better suited to urban scenarios, while algorithms such as lane marking detection could offer a good alternative in peri-urban environments.
- In parallel, we worked on designing a reliability measure associated to the pose given by our SLAM considering the geometrical configurations of the 2D laser points describing the environment and the computations done in the maximum likelihood matching process.

6.5. Feature Selection for road obstacles classification

Participants: Itheri Yahiaoui, Pierre Merdrignac, Anne Verroust-Blondet.

In order to ensure the ability of an automated vehicle to be autonomous in a real environment, we must equip it with tools (hardware and software) to meet the requirement of such an application as safety, real-time processing, understanding and intelligence, etc. To contribute to these objectives a perception system is of vital importance. The one on road obstacles detection and classification is of particular interest for us. In this work a large number of geometric features have been proposed to describe different class objects like vehicles, pedestrians, cyclists and static obstacles from 2D laser points. A binary classification was performed with an Adaboost algorithm. In order to improve this work and enhance the classification rate, we have constructed new binary and multiclass classifiers, using SVM and logistic regression, with optimal choices of kernel parameters and models. We have defined several decision strategies by tracking objects in the video sequences, which lead to obtain the most probable target object. On the other hand, we have studied different dependence measures between the proposed features and the classes, leading the selection of the best set of features. As measures of dependence, we have used nonparametric estimate of mutual information, Fisher information and Pearson correlation. We have used also the Akaïke criterion in order to select the best models (the best subset of features) in logistic regression.

6.6. Motion planning techniques

Participants: David González Bautista, Fernando Garrido Carpio, Vicente Milanés, Fawzi Nashashibi, Myriam Vaca Recalde, Jose Emilio Traver Becerra.

The latest developments in the Intelligent Transportation Systems (ITS) field allow emerging technologies to show promising results at increasing passengers comfort and safety, while decreasing energy consumption, emissions and travel time. Despite of great efforts, fully automated driving still remains unsolved, where research challenges such as navigation in urban dynamic environments with obstacle avoidance capabilities–i.e. Vulnerable Road Users (VRU) and vehicles–and cooperative maneuvers among automated and semi-automated vehicles, still need further efforts for a real environment implementation. A deep state-of-the-art review has been conducted to find the gaps in this important topic into the autonomous vehicle field, with special attention to overtaking and obstacle avoidance maneuvers [21].

Having this in mind, a novel local path planning algorithm combining both off-line and real-time generation has been proposed in [32], providing a significant reduction on the computational time with respect to prior implementations from RITS team. This new local planning architecture for urban environments benefits from *a priori* knowledge of the geometry of the road layout, vehicle's kinematics and dynamics, among others, to produce local smooth path for the vehicle to navigate. The planner relies on several databases containing optimized trajectories for a G^2 continuous path generation. Four different type of databases have been generated to provide our system with a naturalistic driving style, allowing the car to maintain smooth trajectories according to the characteristics of the road [33].

Based on the accuracy of current digital maps, it is possible to know before-hand the way-points that define the route by which the vehicle will pass to reach a predefined destination. Furthermore, the original route can be generated in real-time and modified on-demand according to the user needs through the use of Automatic Global Planners (AGP) [42]. That way, since urban scenarios can present several consecutive curves in a short period of time, a smoother and more comfortable path generation can be done by extending the planning horizon up to two curves. There, a set of paths are analyzed by considering the angles of the curves and the distances to them in order to find the best joint point for the consecutive curves.

In this sense, a speed planning algorithm has also been designed to increase passenger comfort and set continuous speed profiles [35]. The approach permits to improve the comfort in automated vehicles by integrating the speed profile with the previously computed path, constraining the global acceleration in the whole ride (longitudinal and lateral accelerations according to ISO 2631-1). It also minimizes distance error problems by associating the speed profile w.r.t. distance in the path instead of the time. The planner has been tested against other techniques in the state-of-the-art, providing better results.

The proposed architecture has been validated both on simulation (with Pro-Sivic and RTMaps) and on the Inria Rocquencourt terrain. The results showed a smoother tracking of the curves, reduction on the execution times and reduced global accelerations increasing comfort. Future works will improve the capacity to deal with dynamic obstacles, conducting avoidance maneuvers if possible, or returning to the original lane if not. The maneuver will be decided by building an occupancy grid with the information given by the perception system. It will provide the best point near the obstacle to carry out the avoidance trajectory by loading the pre-computed curves.

6.7. Plug&Play control for highly non-linear systems: Stability analysis of autonomous vehicles

Participants: Francisco Navas, Vicente Milanés, Fawzi Nashashibi.

The final stage for automating a vehicle relies on the control algorithms. They are in charge of providing the proper behavior and performance to the vehicle, leading to provide fully automated capabilities. Controllability and stability of dynamic complex systems are the key aspects when it comes to design intelligent control algorithms for vehicles.

Nowadays, the problem is that control systems are “monolithic”. That means that a minor change in the system could require the entire redesign of the control system. It addresses a major challenge, a system able to adapt the control structure automatically when a change occurred.

An autonomous vehicle is built by combining a set-of-sensors and actuators together with sophisticated algorithms. Since sensors and actuators are prone to intermittent faults, the use of different sensors is better and more cost effective than duplicating the same sensor type. The problem is to deal with the different availability of each sensor/actuator and how the vehicle should react to these changes. A methodology that improves the security of autonomous driving systems by providing a framework managing different sensor/actuator setups should be carried out. New trends are proposing intelligent algorithms able to handle any unexpected circumstances as unpredicted uncertainties or even fully outages from sensors. This is the case of Plug&Play control, which is able to provide stability responses for autonomous vehicles under uncontrolled circumstances, including modifications on the input/output sensors.

In order to meet with the idea of automatically handling those changes into the system, different research lines should be followed:

- Reconfiguration of existing controllers whenever changes are introduced in the system being controlled. In that line, the already commercially available Adaptive Cruise Controller (ACC) system, and its evolution by adding vehicle-to-vehicle communication (CACC) are examined. Plug&Play control is used for providing stable transitions between both controllers when the vehicle-to-vehicle communication link is changing from available to available or vice versa. More detail can be found in [38]. Gain scheduling approaches can be achieved by using the same structure. An Advanced-CACC is developed by using it. Hybrid behaviors between controllers with different head times are carried out depending on the traffic situation.

- Online closed loop identification of the vehicle and its components. Plug&Play control also provides a way for doing online closed loop identification of any system as open loop like systems. Here, the obtained models for the vehicle will be compared with the physical lateral model (Bicycle and 2GDL) and the longitudinal model together with the tire models (Pacejka, Dugoff and Buckhardt). It is also possible to identify new sensors or actuators connected to the system.
- Automatic control reconfiguration to achieve optimal performance together with identification of the new situation. Once a new situation has been identified in the system, the controller should be reconfigured to achieve the optimal performance of the autonomous vehicle.

6.8. Using Fractional Calculus for Cooperative Car Following Control

Participants: Carlos Flores, Vicente Milanés, Fawzi Nashashibi.

In the field of Advanced Driver Assistance Systems (ADAS), there are two main types of systems: passive and active ones. Specifically the active ADAS, they are capable of taking partial or complete control of the vehicle. Among these techniques, Car-Following has arisen as one important solution to traffic jams, driver comfort and safety.

Scoping on the evolution of the control involved in Car Following, it can be remarked the improved version of the cruise control system, Adaptive Cruise Control (ACC). This system allows the vehicle to maintain a desired distance gap measured by raging sensors (LiDAR, radars, etc), by controlling longitudinally the vehicle through the throttle and brake.

Afterward, the addition of Vehicle to Vehicle (V2V) communication links allowed the vehicles to maintain even shorter distances between each of the string members, by performing a Cooperative ACC (CACC). Focusing on CACC formations, a control structure must be conceived to guarantee stability and string stability as well. As a core of the control structure, the controller must be able to maintain the vehicle in the desired spacing in a stable, robust and comfortable.

Towards achieving this goals, it is proposed to use fractional order calculus to gain a more flexible frequency response and at the same time satisfy more demanding design requirements. This mathematical has been used for years for different applications providing good results and outperforming classical techniques in the industrial control field, due to its capability of describing systems more accurately than integer order calculus. Several research lines are stated to achieve these objectives:

- An exhaustive identification process of the experimental platforms dynamics. Allowing further comparison between the empirical identified dynamics of the real vehicle and a theoretical mathematical dynamic model. Such permits to design much more effective and stable control algorithms for both the lateral and longitudinal command of the vehicle.
- Conception of a Car-Following gap regulation controller using fractional order calculus, which has been proven that yields a more accurate description of real processes. The controller should satisfy more demanding design requirements [31], allowing to extend the scope of Car Following controllers' design. This controller should be framed into an appropriate control structure both for ACC and CACC
- Further investigation on the effects of communication delays and latency in the V2V links, as well as study different control structures that react not with the preceding vehicle's behavior but also other string members.

6.9. Decision making for automated vehicles in urban environments

Participants: Pierre de Beaucorps, Thomas Streubel, Anne Verroust-Blondet, Fawzi Nashashibi.

The development of automated vehicles in urban environments requires a robust sensing system followed by an adaptive situation assessment. This is the basis for smart decision making in the driving process without collisions or taking high risks. We address this aspect of automated driving in a project with the sensor developer VALEO. The focus is on complex urban traffic scenarios, e.g. intersections and roundabouts, including multiple road users.

In a first step, we developed a new multi-agent driving simulation as a tool to explore human behavior in relevant traffic scenarios. We conducted a study with 10 test persons driving in a scene with one dummy car to acquire data and understand the human decision process in risky situations. This data was used to retrieving speed profiles for the trajectory planning. The path planning was established with Bezier curves. Further, a robust decision making algorithm utilizes the trajectory planning coupled with a risk assessment. The latter is estimating the post-encroachment time (PET), which is the time between one vehicle leaving a collision zone in an intersection area and the other car entering this same zone. Based on this estimation a risk is assigned to every predetermined speed profile and the one with lowest acceptable risk is chosen to be send to the controller of the automated vehicle. The results showed better performance than the drivers in our study. The so equipped automated vehicle is integrated in our simulation environment and was presented to our project partners in several intersection and roundabout scenarios with a real driver in the same scene.

6.10. Transposition of autonomous vehicle architecture

Participants: Raoul de Charette, Pablo Marin Plaza, Fawzi Nashashibi.

With the development of autonomous vehicles, many software and hardware architectures exist in the world to handle perception, control, decision, planning. Studies were conducted to see how an alien software architecture could be transposed to our Cycabs platforms. Lightweight Communications and Marshalling has been implemented on our platforms to communicate fully with the Carlos 3 architecture, allowing the alien software pipeline to control fully our vehicle. Results and studies include stability of the communication, impact on the control quality, and planning comparison.

6.11. Fusion of Perception and V2P Communication Systems for Safety of Vulnerable Road Users

Participants: Pierre Merdrignac, Oyunchimeg Shagdar, Fawzi Nashashibi.

With cooperative intelligent transportation systems (C-ITS), vulnerable road users (VRU) safety can be enhanced by multiple means.

On one hand, perception systems are based on embedded sensors to protect VRUs. However, such systems may fail due to the sensors' visibility conditions and imprecision. On the other hand, Vehicle-to-Pedestrian (V2P) communication can contribute to the VRU safety by allowing vehicles and pedestrians to exchange information. This solution is, however, largely affected by the reliability of the exchanged information, which most generally is the GPS data. Since perception and communication have complementary features, we can expect that a fusion between these two approaches can be a solution to the VRU safety.

In this work, we proposed a cooperative system that combines the outputs of communication and perception. After introducing theoretical models of both individual approaches, we developed a probabilistic association between perception and V2P communication information by means of multi-hypothesis tracking (MHT).

Experimental studies were conducted to demonstrate the applicability of this approach in real-world environments. Our results showed that the cooperative VRU protection system can benefit of the redundancy coming from the perception and communication technologies both in line-of-sight (LOS) and non-LOS (NLOS) conditions. We established that the performances of this system are influenced by the classification performances of the perception system and by the accuracy of the GPS positioning transmitted by the communication system.

More detail can be fund in [24]

6.12. Study and Evaluation of Laser-based Perception and Light Communication for a Platoon of Autonomous Vehicles

Participants: Mohammad Abualhoul, Pierre Merdrignac, Oyunchimeg Shagdar, Fawzi Nashashibi.

Visible Light Communication (VLC) is a new emerging technology that is being proposed as a reliable and supportive choice for short range communications in ITS.

On the same context, Laser Range Finders (LRF) sensors are used for the vehicular environment perception. Compared to VLC, LRF can provide more coverage range and extended viewing angle.

To take the full advantages of both technologies features, we have studied and demonstrated the proposal of using VLC for information exchange among the platoon members and LRF for inter-vehicle distance estimation. A handover algorithm was proposed to manage the switching process for any failure occurrence by assessing LRF and VLC performance using three different metrics: LRF confidence value, vehicles angular orientation, and the VLC link latency.

The evaluation of the proposed system is verified using VLC prototype and Pro-SiVIC Simulator driving platoon of two autonomous vehicles over different curvature scenarios. Our results showed that the proposed combination are extending the VLC limitations and satisfying the platooning requirement. However, in the very sharp curvature, LRF was capable of driving the platoon except for the 90° curve scenario, the system experienced non-stable behavior due to the LRF area of interest limitation.

More detail can be fund in [27].

6.13. Solutions for Safety-Critical Communications in IVNs

Participant: Gérard Le Lann.

In 2016, we have followed a divide-and-conquer approach. Rather than considering medium-range omnidirectional communications, we have split the problem space in two sub-domains, longitudinal short-range SC communications and lateral short-range communications. Our research has been directed at MAC protocols, string-wide message dissemination based on longitudinal communications, and distributed agreement algorithms based on longitudinal and lateral communications. New results are:

- A rigorous characterization of what is meant by SC communications in IVNs: the space-time bounds acceptability (STBA) requirements, as follows:
  - $STBA_1$: a MAC protocol is acceptable if and only if the distance traveled in λ time units by any vehicle involved in a SC scenario is an order of magnitude smaller than average vehicle size.
  - $STBA_2$: a string-wide message dissemination algorithm, or a string-wide distributed agreement algorithm, is acceptable if and only if the distance traveled in Δ time units by any vehicle involved in a SC scenario is smaller than average vehicle size.
- Specification of SWIFT (Synchronous Wireless Interference-Free Transmissions), a collision-free MAC protocol that solves the BCAD and the TBMA problems introduced in [48] (no solutions given in this publication), and that also achieves fast string-wide acknowledged message dissemination,
- Analytical formulae of worst-case upper bounds λ and Δ achieved with SWIFT [36],
- Specification of Fast Distributed Agreement (FastDA), a problem that arises in IVNs in the presence of conflicting concurrent SC events (e.g., lane changes and brutal braking), under two instances, single-lane (longitudinal) agreement and multilane (lateral and longitudinal) agreement [37],
- Specifications of solutions to FastDA: the Eligo algorithm for the single-lane string-wide agreement (SLA), and the LHandshake protocol for the multilane agreement (MLA),
- Analytical formulae of worst-case upper bounds Δ_{SLA} and Δ_{MLA} achieved with Eligo and LHandshake, respectively [37],
- Verification that SWIFT, Eligo and LHandshake meet the STBA requirements.

It turns out that SWIFT, Eligo, and LHandshake outperform existing stochastic solutions.

6.14. Large scale simulation interfacing

Participants: Ahmed Soua, Jean-Marc Lasgouttes, Oyunchimeg Shagdar.

In order to efficiently design and validate a cooperative intelligent transportation system, a complete simulation environment handling both mobility and communication is required. 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 802.11p protocol. This simulation tool is be 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. In fact, the messages emitted by the vehicles located outside the interference zone are not useful for the simulation of the ego-vehicle, and therefore limiting the transmission area to a useful one reduces obviously the number of nodes involved in the transmission operation and thus reduces the processing time of messages. To corroborate the efficiency of our proposal, we compare it with an already existing simulation tool called COLOMBO. The simulation results have shown that our technique outperforms COLOMBO in terms of simulation execution time in the case of large scale simulations (when the number of vehicles exceeds 2400 nodes).

6.15. Belief propagation inference for traffic prediction

Participant: Jean-Marc Lasgouttes.

This work [50], in collaboration with Cyril Furtlehner (TAO, Inria), deals with real-time prediction of traffic conditions in a setting where the only available information is floating car data (FCD) sent by probe vehicles. The main focus is on finding a good way to encode some coarse information (typically whether traffic on a segment is fluid or congested), and to decode it in the form of real-time traffic reconstruction and prediction. Our approach relies in particular on the belief propagation algorithm.

The work about the theoretical aspects of encoding real valued variables into a binary Ising model has now been published [23].

Moreover, following an agreement signed with the city of Vienna (Austria) and the company SISTeMA ITS (Italy), we obtained access to large amounts of data. We are now working on assessing the performance of our techniques in real-world city networks.

6.16. Random Walks in Orthants

Participant: Guy Fayolle.

The Second Edition of the Book [45] *Random walks in the Quarter Plane*, prepared in collaboration with R. Iasnogorodski (St-Petersburg, Russia) and V. Malyshev (MGU, Moscow), is complete and now in the Springer Production Department. It will be published in the collection *Probability Theory and Stochastic Processes*. **Part II** of this second edition borrows specific case-studies from queuing theory, and enumerative combinatorics. Five chapters have been added, including examples and applications of the general theory to enumerative combinatorics. Among them:

- Explicit criterion for the finiteness of the group, both in the genus 0 and genus 1 cases.
- Chapter *Coupled-Queues* shows the first example of a queuing system analyzed by reduction to a BVP in the complex plane.

- Chapter *Joining the shorter-queue* analyzes a famous model, where maximal homogeneity conditions do not hold, hence leading to a system of functional equations.
- Chapter *Counting Lattice Walks* concerns the so-called *enumerative combinatorics*. When counting random walks with small steps, the nature (rational, algebraic or holonomic) of the generating functions can be found and a precise classification is given for the basic (up to symmetries) 79 possible walks.

6.17. Facing ADAS validation complexity with usage oriented testing

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 maybe 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. The COVADEC project (type FUI/FEDER 15) aims to provide methods and techniques to deal with these problems. The test cases automatic generation relies on a *Model Based Testing (MBT)* approach. The tool used for MBT is the software MaTeLo (Markov Test Logic), developed by the company All4Tec. MaTeLo is an MBT tool, which makes it possible to build a model of the expected behaviour of the system under test and then to generate, from this model, a set of test cases suitable for particular needs. MaTeLo is based on Markov chains, and, for non-deterministic generation of test cases, uses the Monte Carlo methods. To cope with the inherent combinatorial explosion, we couple the graph generated by MaTeLo to an ad hoc *random scan Gibbs sampler (RSGS)*, which converges at geometric speed to the target distribution. Thanks to these test acceleration techniques, MaTeLo also makes it possible to obtain a maximal coverage of system validation by using a minimum number of test cases. As a consequence, the number of driving kilometers needed to validate an ADAS is reduced, see [40], [41].

6.18. Broadcast Transmission Networks with Buffering

Participant: Guy Fayolle.

In collaboration with P. Muhlethaler, we analyzed the so-called back-off technique of the IEEE 802.11 protocol in broadcast mode with waiting queues. In contrast to existing models, packets arriving when a station (or node) is in back-off state are not discarded, but are stored in a buffer of infinite capacity. As in previous studies, the key point of our analysis hinges on the assumption that the time on the channel is viewed as a random succession of transmission slots (whose duration corresponds to the length of a packet) and mini-slots during which the back-off of the station is decremented. These events occur independently, with given probabilities. The state of a node is represented by a two-dimensional Markov chain in discrete-time, formed by the back-off counter and the number of packets at the station. Two models are proposed both of which are shown to cope reasonably well with the physical principles of the protocol. Stability (ergodicity) conditions are obtained and interpreted in terms of maximum throughput. Several approximations related to these models are also discussed in [20].

AYIN Team

6. New Results

6.1. Fusion of multitemporal and multiresolution remote sensing data and application to natural disasters

Participants: Ihsen Hedhli, Josiane Zerubia [contact].

This work was carried out in collaboration with Prof. Gabriele Moser and Prof. Sebastiano Serpico from DITEN departement (http://www.dibe.unige.it/index.php?lang=en), University of Genoa, Italy.

In this work we address the problem of constructing statistical models of images using Hierarchical Hidden Markov modeling techniques for high resolution remotely sensed image classification of urban areas. The main difficulty is to develop a classifier that jointly utilizes the benefits of multi-band and multi-resolution input data while maintaining a good trade-off between accuracy and computation time. In this framework, Markov random field (MRF) models are widely used in classification problems since they provide a convenient and consistent way of integrating contextual information into the classification scheme. Furthermore, MRF models defined according to hierarchical structures exhibit good methodological and application-oriented properties including causality, thanks to the use of appropriate graphs such as a quadtree structure [1]. The input satellite images are inserted in a hierarchical structure on the basis of their spatial resolution. This approach is aimed at both exploiting multi-scale information, which is known to play a crucial role in high-resolution image analysis, and supporting contextual information at each scale. However, hierarchical MRFs on quad-trees rely on a causality concept captured by the factorization of the prior distribution in terms of causal transition probabilities [2]. In practice, this structure tends to generate "blocky" effects in the final classification map. Due to this disadvantage, a new hierarchical MRF based on a Symmetric Markov Mesh Random Field (SMMRF) is proposed in this work, to overcome these limitations from both mathematical and practical points of view, and to establish a causal and symmetrical model. This can be accomplished by scanning the lattice at each level of the hierarchical model based on the visiting scheme shown in Fig 1 . Then, for each scale of the quad-tree, the causal SMMRF is integrated into the hierarchical structure. Accordingly, each node s at each scale level of the quad-tree, except at the root, is linked to one parent (in the upper level) and three neighbors (in the same level). For each pixel at the root level, there is no parent and only the neighbors remain. The shapes of the neighborhoods of the pixels at the top and left borders of each lattice, at each scale level of the pyramid, are obviously adapted to the image borders [8].

We applied the developed hierarchical classification approach to a multi-resolution dataset that consists of a panchromatic and a multi-spectral Pléiades images acquired over Port-au-Prince (Haiti). Experimental results with HR satellite imagery of a very high-resolution urban scene suggest that the method allows to effectively incorporate spatial information in the hierarchical classification process and provides higher accuracies than previous techniques. Indeed, it is confirmed experimentally (see Fig. 2) that MMRFs and their lattice models are corner-dependent, and that the proposed approach is effective in circumventing this drawback by using a Symmetric Markov Mesh Random Field. The proposed method, in the application to a challenging urban area classification problem, is able to combine the computational and modeling benefits of hierarchical and symmetric mesh MRF models, while preventing their individual artifacts.

6.2. Multitemporal change detection on image sequences with a False Discovery Rate approach

Participant: Josiane Zerubia [contact].

This work was carried out in collaboration with Dr. Vladimir Krylov, Prof. Gabriele Moser and Prof. Sebastiano Serpico from DITEN departement (http://www.dibe.unige.it/index.php?lang=en), University of Genoa, Italy.

a b

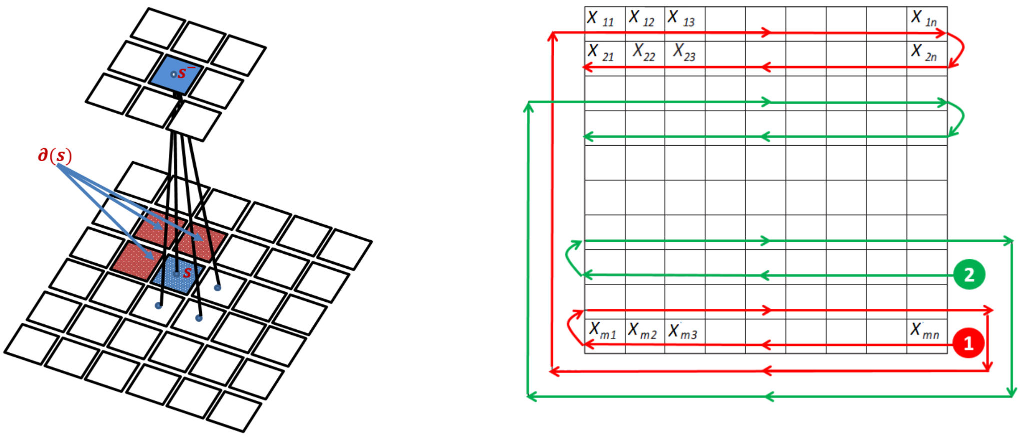


Figure 1. (a) Hybrid structure that combines a spatial grid using an SMMRF and a hierarchical MRF via a quad-tree. (b) Regular rectangular lattice S of size m x n: the "past" of site $s_{i,j}$ is the gray area, arrow lines show raster scan.

a b c d

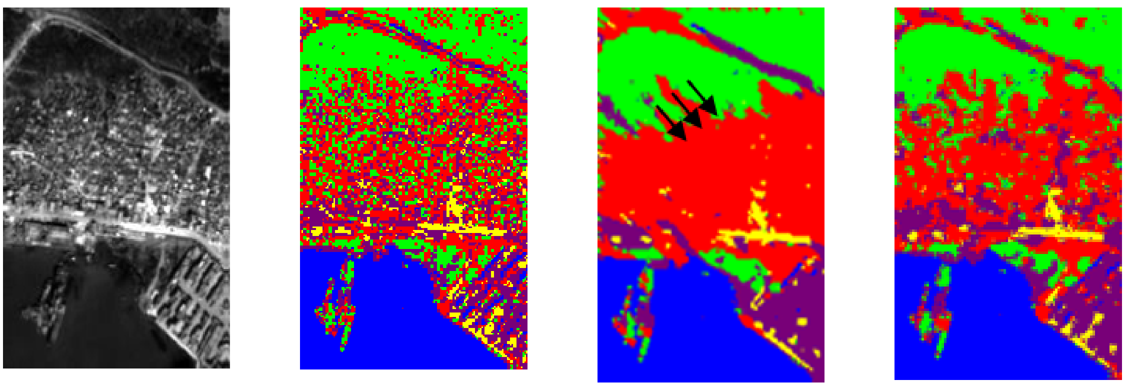


Figure 2. classification maps of optical (Pléiades) image (a) using the original Laferté method (b), the previous method in [2] (c) and the new proposed method (d).

Multitemporal change detection on image sequences is one of the fundamental image processing problems and multiple detection, monitoring and tracking applications rely on its accurate and timely performance. To address this problem we develop an approach that gives a unified statistical thresholding procedure to perform change detection based on statistical features that have a known distribution under the no-change hypothesis. The proposed False Discovery Rate (FDR) formulation is based on the control of the proportion of false alarms among all detections [3]. This efficient technique for large scale hypotheses testing allows to use the wide range of statistical tests developed in the state-of-the-art by adjusting to the dependence structure present in the images and the patch-based samples. The developed approach involves only a few parameters and is highly parallelizable. We propose several rank-based statistical features that report accurate experimental results and the corresponding detectors positively compare with benchmark techniques in three different applications. Further features can be easily constructed to elaborate application-specific change detectors.

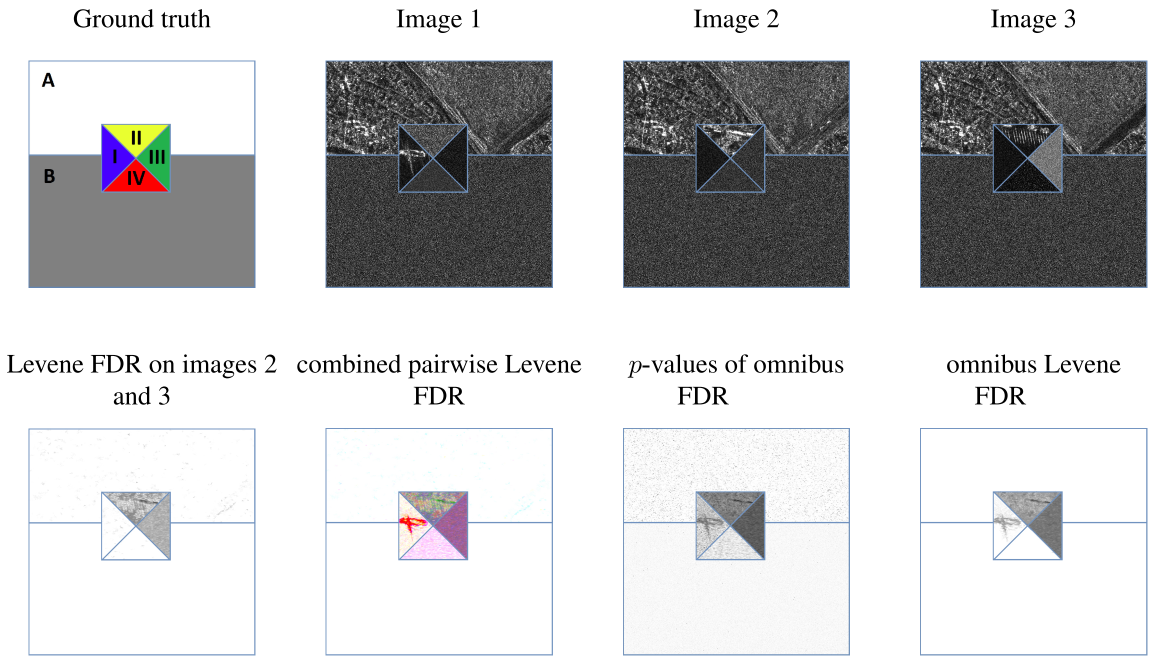


Figure 3. Semi-synthetic 3-image SAR sequence based on COSMO-SkyMed (©ASI) images of Haiti in 2011: (a) ground truth and (b)-(d) images, results of (e), (f) pairwise and (g), (h) omnibus Levene FDR-detection.

In Fig. 3 we demonstrate a typical result of the FDR-based change detector on a semi-synthetic 3-image synthetic aperture radar sequence based on COSMO-SkyMed (©ASI) images of Haiti (April, May and August 2011). In this experiments the Levene multisample statistic with a 9-by-9 local window is employed. A comparison of the omnibus test (formulating the hypothesis for all three images simultaneously) with pairwise tests, demonstrates that the latter are more sensitive to changes. This sensitivity can be a disadvantage as is the case with the detection noise in (e) and (f). The omnibus test on the other hand did not suffer from the same mistake due to a generally higher level of tolerance to pairwise fluctuations. Hence, from the SAR-change detection point of view, the results reported by omnibus version of Levene-statistic are considered more adequate.

6.3. Solving inverse problems related to FUV image processing for ICON mission

Participant: Josiane Zerubia [contact].

This work has been conducted in collaboration with Prof. Farzad Kamalabadi, Dr. Jianqi Qin and Dr. Mark Butala from Coordinated Science Laboratory (CSL, http://www.csl.illinois.edu/) at University of Illinois at Urbana Champaign (UIUC, http://illinois.edu/)

ICON (Ionospheric Connection Explorer) is a satellite which is part of the NASA Explorer missions (see http://icon.ssl.berkeley.edu/) and is planned to be launched in 2017 (see Fig. 4). The main goal of ICON is to study the area where terrestrial weather meets space weather in order to understand the behavior of our planet's upper atmosphere, including what causes disruptions in this region, such as those that can significantly affect radio transmissions.

There will be 4 instruments on board. One of them is the FUV: Far UltraViolet spectrographic imager. Prof Kamalabadi is responsible to process the FUV data from this instrument. During Josiane Zerubia's stay at CSL, UIUC, she worked with Prof. Farzad Kamalabadi team on defining a proper energy function using Bayesian theory (i.e. defining a data term + various priors for regularizing the solution) in order to be able to take into account the geometry of the information and also to deal with optical transmission function. This inverse problem is highly non linear. We will continue in the future to work on the problem of the error estimation (or bound derivation) as far as the estimation of distribution parameters is concerned.

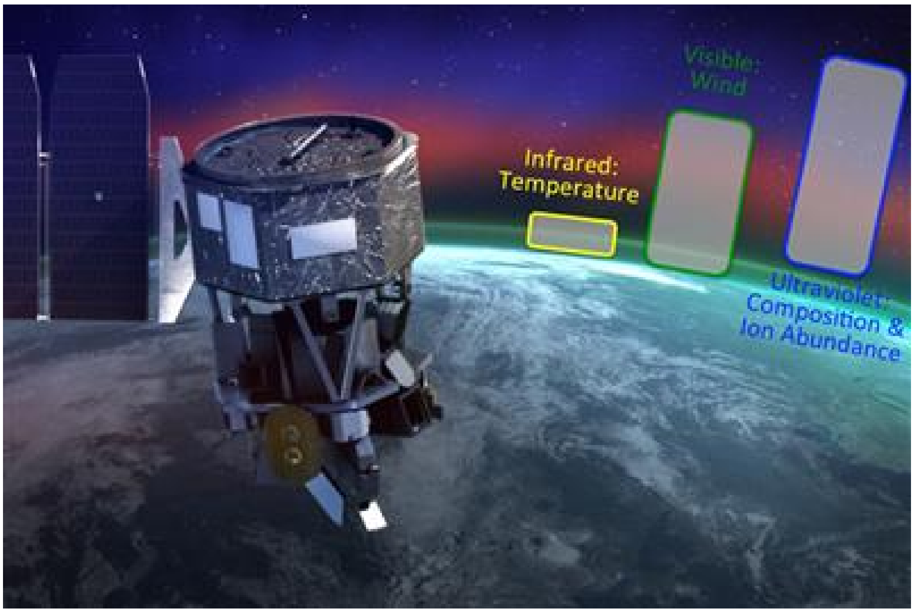


Figure 4. Ionospheric Connection Explorer satellite ©NASA.

6.4. Hyperspectral Image Processing for Detection and Grading of Skin Erythema

Participant: Josiane Zerubia [contact].

This work was carried out in collaboration with Ali Madooei (Simon Fraser University (https://www.sfu.ca/), Canada), Ramy Abdlaty (McMaster University (https://www.mcmaster.ca/), Canada), Lilian Doerwald-Munoz (Hamilton Health Sciences - General Hospital (http://www.hamiltonhealthsciences.ca/), Canada), Joseph Hayward (Hamilton Health Sciences - General Hospital, Canada), Mark Drew (Simon Fraser University, Canada) and Qiyin Fang (McMaster University, Canada).

This study focused on detection and grading of skin erythema using hyperspectral image processing. The ultimate objective is to build a system for monitoring radiation response in individuals using hyperspectral imaging technology and image processing. The present project was to investigate the possibility of monitoring the degree of skin erythema. To this aim, we proposed an image processing pipeline and conducted controlled experiments to demonstrate the efficacy of the proposed approach for (1) reproducing clinical assessments, and (2) outperforming RGB imaging data (see Fig. 5). We combined the problem of erythema detection and grading into a multi-class classification problem where each pixel is classified as one of the four erythema classes or a non-erythema class. We used a weighted LDA (linear discriminant analysis) classifier to deal with noisy labels. Moreover, we devised pre-processing steps to deal with noisy measurements. We evaluate the system against the clinical assessment of an experienced clinician. We also compare the performance to that of using digital photography (instead of hyperspectral images) [9]. The results from this preliminary study are encouraging and indicate that hyperspectral image data contain relevant information, and indeed outperform imaging photography. In future, we want to extend the technique to further detect other skin responses to radiation (such as dry/moist desquamation, skin necrosis, etc.) and also to experiment with real patients undergoing radiotherapy.

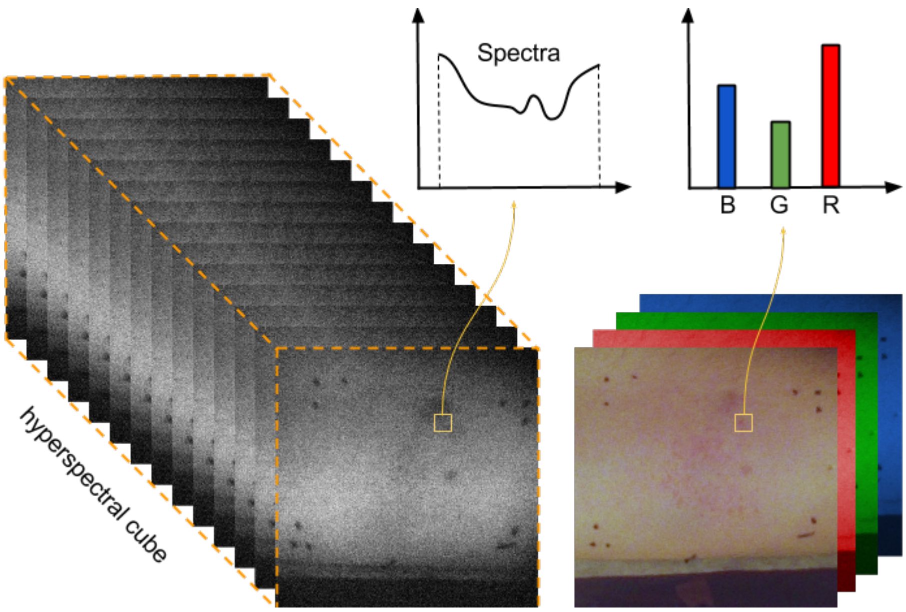


Figure 5. A schematic representation of hyperspectral vs. RGB image data. The image shows artificially induced erythema over the inside of the forearm of a volunteer.

LINKMEDIA Project-Team

7. New Results

7.1. Unsupervised motif and knowledge discovery

7.1.1. Multimodal person discovery in TV broadcasts

Participants: Guillaume Gravier, Gabriel Sargent, Ronan Sicre.

Work in collaboration with Silvio J. Guimarães, Gabriel B. de Fonseca and Izabela Lyon Freire, PUC Minas, in the framework of the Inria Associate Team MOTIF.

Pursuing efforts initiated in 2015 in the framework of the MediaEval benchmark on Multimodal Person Discovery, we investigated graph-based approaches to name the persons on screen and speaking in TV broadcasts with no prior information, leveraging text overlays, speech transcripts as well as face and voice comparison. We adopted a graph-based representation of speaking faces and investigated two tag-propagation approaches to associate overlays co-occurring with some speaking faces to other visually or audiovisually similar speaking faces. Given a video, we first build a graph from the detected speaking faces (nodes) and their audiovisual similarities (edges). Each node is associated to its co-occurring overlays (tags) when they exist. Then, we consider two tag-propagation approaches, respectively based on a random walk strategy and on Kruskal's minimum spanning tree algorithm for node clustering [28].

7.1.2. Efficient similarity self-join for near-duplicate video detection

Participants: Laurent Amsaleg, Guillaume Gravier.

Work in collaboration with Henrique B. da Silva, Silvio J. Guimarães, Zenilto do Patrocino Jr., PUC Minas, and Arnaldo de A. Araújo, UFMG, in the framework of the Inria Associate Team MOTIF.

The huge amount of redundant multimedia data, like video, has become a problem in terms of both space and copyright. Usually, the methods for identifying near-duplicate videos are neither adequate nor scalable to find pairs of similar videos. Similarity self-join operation could be an alternative to solve this problem in which all similar pairs of elements from a video dataset are retrieved. Methods for similarity self-join however exhibit poor performance when applied to high-dimensional data. In [33], we propose a new approximate method to compute similarity self-join in sub-quadratic time in order to solve the near-duplicate video detection problem. Our strategy is based on clustering techniques to find out groups of videos which are similar to each other.

7.1.3. Recommendation systems with matrix factorization

Participants: Raghavendran Balu, Teddy Furon.

Matrix factorization is a prominent technique for approximate matrix reconstruction and noise reduction. Its common appeal is attributed to its space efficiency and its ability to generalize with missing information. For these reasons, matrix factorization is central to collaborative filtering systems. In the real world, such systems must deal with million of users and items, and they are highly dynamic as new users and new items are constantly added. Factorization techniques, however, have difficulties to cope with such a demanding environment. Whereas they are well understood with static data, their ability to efficiently cope with new and dynamic data is limited. Scaling to extremely large numbers of users and items is also problematic. In [10], we propose to use the count sketching technique for representing the latent factors with extreme compactness, facilitating scaling.

In [11], we discovered that sketching techniques implicitly provide differential privacy guarantees thanks to the inherent randomness of the data structure. Collaborative filtering is a popular technique for recommendation system due to its domain independence and reliance on user behavior data alone. But the possibility of identification of users based on these personal data raise privacy concerns. Differential privacy aims to minimize these identification risks by adding controlled noise with known characteristics. The addition of noise impacts the utility of the system and does not add any other value to the system other than enhanced privacy.

7.2. Multimedia content description and structuring

7.2.1. Hierarchical topic structuring

Participants: Guillaume Gravier, Pascale Sébillot.

In [37], we investigated the potential of a topical structure of text-like data that we recently proposed [55] in the context of summarization and anchor detection in video hyperlinking. This structure is produced by an algorithm that exploits temporal distributions of words through word burst analysis to generate a hierarchy of topically focused fragments. The obtained hierarchy aims at filtering out non-critical content, retaining only the salient information at various levels of detail. For the tasks we choose to evaluate the structure on, the lost of important information is highly damaging. We show that the structure can actually improve the results of summarization or at least maintain state-of-the-art results, while for anchor detection it leads us to the best precision in the context of the Search and Anchoring in Video Archives task at MediaEval. The experiments were carried on written text and a more challenging corpus containing automatic transcripts of TV shows.

7.2.2. Multimedia-inspired descriptors for time series classification

Participant: Simon Malinowski.

The SIFT framework has shown to be effective in the image classification context. Recently, we designed a bag-of-words approach based on an adaptation of this framework to time series classification. It relies on two steps: SIFT-based features are first extracted and quantized into words; histograms of occurrences of each word are then fed into a classifier. In [38], we investigated techniques to improve the performance of bag-of-temporal-SIFT-words: dense extraction of keypoints and different normalizations of Bag-of-Words histograms. Extensive experiments have shown that our method significantly outperforms nearly all tested standalone baseline classifiers on publicly available UCR datasets. In [23], we also investigate the use of convolutional neural networks (CNN) for time series classification. Such networks have been widely used in many domains like computer vision and speech recognition, but only a little for time series classification. We have designed a convolutional neural network that consists of two convolutional layers. One drawback with CNN is that they need a lot of training data to be efficient. We propose two ways to circumvent this problem: designing data-augmentation techniques and learning the network in a semi-supervised way using training time series from different datasets. These techniques are experimentally evaluated on a benchmark of time series datasets.

7.2.3. Early time series classification

Participant: Simon Malinowski.

In time series classification, two antagonist notions are at stake. On the one hand, in most cases, the sooner the time series is classified , the higher the reward. On the other hand, an early classification is more likely to be erroneous. Most of the early classification methods have been designed to take a decision as soon as a sufficient level of reliability is reached. However, in many applications, delaying the decision with no guarantee that the reliability threshold will be met in the future can be costly. Recently, a framework dedicated to optimizing the trade-off between classification accuracy and the cost of delaying the decision was proposed, together with an algorithm that decides online the optimal time instant to classify an incoming time series. On top of this framework , we have built in [29] two different early classification algorithms that optimize the trade-off between decision accuracy and the cost of delaying the decision. These algorithms are non-myopic in the sense that, even when classification is delayed, they can provide an estimate of when the optimal classification time is likely to occur. Our experiments on real datasets demonstrate that the proposed approaches are more robust than existing methods.

7.3. Content-based information retrieval

7.3.1. Bi-directional embeddings for cross-modal content matching

Participants: Guillaume Gravier, Christian Raymond, Vedran Vukotić.

Common approaches to problems involving multiple modalities (classification, retrieval, hyperlinking, etc.) are early fusion of the initial modalities and crossmodal translation from one modality to the other. Recently, deep neural networks, especially deep autoencoders, have proven promising both for crossmodal translation and for early fusion via multimodal embedding. In [31], we propose a flexible cross-modal deep neural network architecture for multimodal and crossmodal representation. By tying the weights of two deep neural networks, symmetry is enforced in central hidden layers thus yielding a multimodal representation space common to the two original representation spaces. The proposed architecture is evaluated in multimodal query expansion and multimodal retrieval tasks within the context of video hyperlinking. In [32], we extend the approach, focusing on the evaluation of a good single-modal continuous representations both for textual and for visual information. word2vec and paragraph vectors are evaluated for representing collections of words, such as parts of automatic transcripts and multiple visual concepts, while different deep convolutional neural networks are evaluated for directly embedding visual information, avoiding the creation of visual concepts. We evaluate methods for multimodal fusion and crossmodal translation, with different single-modal pairs, in the task of video hyperlinking.

7.3.2. Intrinsic dimensions in language 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. Recently, several authors have studied the real dimensionality of the datasets, called intrinsic dimensionality, in specific parts of these spaces. In [34], we propose to revisit this notion through a coefficient called α in the specific case of IR and to study its use in IR tasks. More precisely, we show how to estimate α from IR similarities and to use it in representation spaces used for documents and words. Indeed, we prove that α may be used to characterize difficult queries. We moreover show that this intrinsic dimensionality notion, applied to words, can help to choose terms to use for query expansion.

7.3.3. Evaluation of distributional thesauri

Participants: Vincent Claveau, Ewa Kijak.

With the success of word embedding methods, all the fields of distributional semantics have experienced a renewed interest. Beside the famous word2vec, recent studies have presented efficient techniques to build distributional thesaurus, including our work on information retrieval (IR) tools and concepts to build a thesaurus [14]. In [13], we address the problem of the evaluation of such thesauri or embedding models. Several evaluation scenarii are considered: direct evaluation through reference lexicons and specially crafted datasets, and indirect evaluation through a third party tasks, namely lexical subsitution and Information Retrieval. Through several experiments, we first show that the recent techniques for building distributional thesaurus outperform the word2vec approach, whatever the evaluation scenario. We also highlight the differences between the evaluation scenarii, which may lead to very different conclusions when comparing distributional models. Last, we study the effect of some parameters of the distributional models on these various evaluation scenarii.

7.3.4. Scaling group testing similarity search

Participants: Laurent Amsaleg, Ahmet Iscen, Teddy Furon.

The large dimensionality of modern image feature vectors, up to thousands of dimensions, is challenging high dimensional indexing techniques. Traditional approaches fail at returning good quality results within a response time that is usable in practice. However, similarity search techniques inspired by the group testing framework have recently been proposed in an attempt to specifically defeat the curse of dimensionality. Yet, group testing does not scale and fails at indexing very large collections of images because its internal procedures analyze an excessively large fraction of the indexed data collection. In [16], we identify these difficulties and proposes extensions to the group testing framework for similarity searches that allow to handle larger collections of feature vectors. We demonstrate that it can return high quality results much faster compared to state-of-the-art group testing strategies when indexing truly high-dimensional features that are indeed hardly indexable with traditional indexing approaches.

We also discovered that group testing helps in enforcing security and privacy in identification. We detail a particular scheme based on embedding and group testing. Whereas embedding poorly protects the data when used alone, the group testing approach makes it much harder to reconstruct the data when combined with embedding. Even when curious server and user collude to disclose the secret parameters, they cannot accurately recover the data. Our approach reduces as well the complexity of the search and the required storage space. We show the interest of our work in a benchmark biometrics dataset [17], where we verify our theoretical analysis with real data.

7.3.5. Large-scale similarity search using matrix factorization

Participants: Ahmet Iscen, Teddy Furon.

Work in collaboration with Michael Rabbat, McGill University, Montréal.

We consider the image retrieval problem of finding the images in a dataset that are most similar to a query image. Our goal is to reduce the number of vector operations and memory for performing a search without sacrificing accuracy of the returned images. In [18], we adopt a group testing formulation and design the decoding architecture using either dictionary learning or eigendecomposition. The latter is a plausible option for small-to-medium sized problems with high-dimensional global image descriptors, whereas dictionary learning is applicable in large-scale scenarios. Experiments with standard image search benchmarks, including the Yahoo100M dataset comprising 100 million images, show that our method gives comparable (and sometimes better) accuracy compared to exhaustive search while requiring only 10 % of the vector operations and memory. Moreover, for the same search complexity, our method gives significantly better accuracy compared to approaches based on dimensionality reduction or locality sensitive hashing.

7.4. Linking, navigation and analytics

7.4.1. Opinion similarity and target extraction

Participants: Vincent Claveau, Grégoire Jadi.

Work in collaboration with Laura Monceaux and Béatrice Daille, LINA, Nantes.

In [19], we propose to evaluate the lexical similarity information provided by word representations against several opinion resources using traditional information retrieval tools. Word representation have been used to build and to extend opinion resources, such as lexicon and ontology, and their performance have been evaluated on sentiment analysis tasks. We question this method by measuring the correlation between the sentiment proximity provided by opinion resources and the semantic similarity provided by word representations using different correlation coefficients. We also compare the neighbors found in word representations and list of similar opinion words. Our results show that the proximity of words in state-of-the-art word representations is not very effective to build sentiment similarity.

In [20], we present the development of an opinion target extraction system in English and transpose it to the French language. In addition, we realize an analysis of the features and their effectiveness in English and French which suggest that it is possible to build an opinion target extraction system independant of the domain. Finally, we propose a comparative study of the errors of our systems in both English and French and propose several solutions to these problems.

7.4.2. Reinformation and fake detection in social networks

Participants: Vincent Claveau, Ewa Kijak, Cédric Maigrot.

Traditional media are increasingly present on social networks, but these usual sources of information are confronted with other sources called reinformation sources. These last ones sometimes tend to distort the information relayed to match their ideologies, rendering it partially or totally false. In [25], we conduct a study pursuing two goals: first, we present a corpus containing Facebook messages issued from both types of media sources; secondly, we propose some experiments in order to automatically detect reinformation messages. In particular, we investigate the influence of shallow features versus features more specifically describing the message content. We also developed a multi-modal hoax detection system composed of text, source, and image analysis [24]. As hoax can be very diverse, we want to analyze several modalities to better detect them. This system is applied in the context of the Verifying Multimedia Use task of MediaEval 2016. Experiments show the performance of each separated modality as well as their combination.

7.4.3. Multimodal video hyperlinking

Participants: Rémi Bois, Guillaume Gravier, Christian Raymond, Pascale Sébillot, Ronan Sicre, Vedran Vukotić.

Pursuing previous work on video hyperlinking and recent advances in multimodal content matching [32], we benchmarked a full video hyperlinking system in the framework of the TRECVid international benchmark [12]. The video hyperlinking task aims at proposing a set of video segments, called targets, to complement a query video segment defined as anchor. The 2016 edition of the task encouraged participants to use multiple modalities. In this context, we chose to submit four runs in order to assess the pros and cons of using two modalities instead of a single one and how crossmodality differs from multimodality in terms of relevance. The crossmodal run performed best and obtained the best precision at rank 5 among participants. In parallel, we also demonstrated that, in this framework, multimodal and crossmodal approaches offer significantly more diversity in the set of target proposed than classical information retrieval based approaches where all modalities are combined. We compared bidirectionnal multimodal embeddings [31] with multimodal LDA approaches as experimented last year in TRECVid [49]. The former offers more accurate matching, the latter exhibiting slighlty more diveristy.

7.4.4. User-centric evaluation of hyperlinked news content

Participants: Rémi Bois, Guillaume Gravier, Pascale Sébillot, Arnaud Touboulic.

Work in collaboration with Éric Jamet, Martin Ragot and Maxime Robert, CRPCC, Rennes.

Following our study of professional user needs in multimedia news analytics [15], we developed a prototype news analytics interface that facilitates the exploration of collections of multimedia documents by journalists. The application, based on standard web technology, enriches classical functionalities for this type of applications (e.g., keyword highlights, named entity detection, keyword search, etc.) with navigation-based functionalities. The latter exploit a graph-based organization of the collection, established from content-based similarity graphs on which community detection is performed along with basic link characterization. We performed usage tests on students in journalism and on journalists where each user was asked to write a synthesis article on a given topic. Preliminary results indicate that the graph-based navigation improves the completeness of the synthesis by exposing users to more content than with a standard search engine.

7.5. Miscellaneous

In parallel with mainstream research activities, LINKMEDIA has a number of contribtions in other domains based on the expertise of the team members.

7.5.1. Bidirectional GRUs in spoken dialog

Participants: Christian Raymond, Vedran Vukotić.

Recurrent neural networks recently became a very popular choice for spoken language understanding (SLU) problems. They however represent a big family of different architectures that can furthermore be combined to form more complex neural networks. In [30], we compare different recurrent networks, such as simple recurrent neural networks, long short-term memory networks, gated memory units and their bidirectional versions, on the popular ATIS dataset and on MEDIA, a more complex French dataset. Additionally, we propose a novel method where information about the presence of relevant word classes in the dialog history is combined with a bidirectional gated recurrent unit (GRU).

7.5.2. Kernel principal components analysis with extreme learning machines

Participant: Christian Raymond.

Work in collaboration with M'Sila University, Algeria.

Nowadays, wind power and precise forecasting are of great importance for the development of modern electrical grids. In [26], we investigate a prediction system for time series based on kernel principal component analysis (KPCA) and extreme learning machine (ELM). Comparison with standard dimensionality reduction techniques show that the reduction of the original input space affects positively the prediction output.

7.5.3. Pronunciation adaptation for spontaneous speech synthesis

Work in collaboration with Gwénolé Lecorvé and Damien Lolive, IRISA, Rennes.

In [36], we present a new pronunciation adaptation method which adapts canonical pronunciations to a spontaneous style. This is a key task in text-to-speech as those pronunciation variants bring expressiveness to synthetic speech, thus enabling new potential applications. The strength of the method is to solely rely on linguistic features and to consider a probabilistic machine learning framework, namely conditional random fields, to produce the adapted pronunciations.

7.6. Participation in benchmarking initiatives

- Video hyperlinking, TRECVid
- Search and anchoring, Mediaeval Multimedia International Benchmark
- Multimodal person discovery in broadcast TV, Mediaeval Multimedia International Benchmark
- DeFT 2015 text-mining challenge

MAGRIT Project-Team

6. New Results

6.1. Matching and localization

Participants: Marie-Odile Berger, Antoine Fond, Pierre Rolin, Gilles Simon, Frédéric Sur.

Pose initialization

Estimating the pose of a camera from a model of the scene is a challenging problem when the camera is in a position not covered by the views used to build the model, because feature matching is difficult in such a situation. Several viewpoint simulation techniques have been recently proposed in this context. They generally come with a high computational cost, are limited to specific scenes such as urban environments or object-centered scenes, or need an initial guess for the pose. In [24], we have proposed a viewpoint simulation method well suited to most scenes and query views. Two major problems have been addressed: the positioning of the virtual viewpoints with respect to the scene, and the synthesis of geometrically consistent patches. Experimental results showed that patch synthesis dramatically improves the accuracy of the pose in case of difficult registration, with a limited computational cost.

Facade detection and matching

Planar building facades are semantically meaningful city-scale landmarks. Such landmarks are essential for localization and guidance tasks in GPS-denied areas which are common in urban environments. Detection of facades is also key in augmented reality systems that allow for the annotation of prominent features in the user's view. We introduced several "facadeness" measures of image regions and showed how to combine them to generate building facade proposals in images of urban environments [26]. We demonstrated the interest of this procedure through two applications. First, a convolutional neural network (CNN) was proposed to detect facades from a restricted list of facade proposals. We showed that this method outperforms the state-of-the-art techniques in term of adequation of the detected facades with a ground truth. In addition, the computational time is compatible with the navigation requirements. Second, we investigated image matching based on facade proposals. Considering a large set of data extracted from Google Street View, we showed that matching based on Euclidean distances between CNN descriptors outperforms the classical SIFT matching based on RANSAC-homography calculation. This work has been submitted to IEEE ICRA'2017.

A preliminary step in facade detection is the image rectification process. For that purpose, we introduced a simple and effective method to detect orthogonal vanishing points in Manhattan scenes. A key element of this approach is to explicitly detect the horizon line *before* detecting the vanishing points, which is done by exploiting accumulations of oriented segments around the horizon line. This results in a significant reduction in computation times, while keeping an accuracy comparable or superior to more complex approaches. A paper reporting on this work was published and an oral presentation was made at Eurographics'2016 [25].

6.2. Handling non-rigid deformations

Participants: Marie-Odile Berger, Jaime Garcia Guevara, Pierre-Frédéric Villard.

Simultaneous pose estimation and augmentation of elastic surface

We have proposed an original method to estimate the pose of a monocular camera while simultaneously modeling and capturing the elastic deformation of the object to be augmented [22]. Our method tackles a challenging problem where ambiguities between rigid motion and non-rigid deformation are present. This issue represents a major barrier for the establishment of an efficient surgical augmented reality where the endoscopic camera moves and organs deform. Using an underlying physical model to estimate the low stressed regions our algorithm separates the rigid body motion from the elastic deformations using polar decomposition of the strain tensor. Following this decomposition, a constrained minimization, that encodes both the optical and the physical constraints, is resolved at each frame. Results on real and simulated data proved the effectiveness of our approach.

Fusing US and CT data
3D ultrasound (3D US) is an ideal imaging modality for hepatic image-guided interventions. Yet, its limited field of view and poor in-depth image quality reduce its usefulness. Within J. Guevara's PhD thesis, we propose to reduce these limitations by augmenting the intraoperative 3D US view with a preoperative image. Our approach is automatic and does not require manual initialization or a tracking device for the 3D US probe. Moreover, by using an underlying biomechanical model, the proposed method handles significant liver deformation, even when it occurs outside the 3D US field of view. The method relies on the segmentation of a vascular tree from the preoperative and intraoperative images, and their transformation into graphs. The preoperative and partial intraoperative graphs are then matched using an algorithm based on a combined Gaussian Process regression approach and biomechanical model. The model is used to robustly select a correct match from several hypotheses generated by the Gaussian Process. Once the two graphs are matched, a deformation of the preoperative liver is driven by the local displacement field computed from the partial graph match.

Individual-specific heart valve modeling
We developed a method to semi-automatically build a mitral valve computational model from micro CT (computed tomography) scans: after manually picking fiducial points on the chordae, the leaflets were segmented and the boundary conditions as well as the loading conditions were automatically defined. Fast Finite Element Method (FEM) simulation was carried out using Simulation Open Framework Architecture (SOFA) to reproduce leaflet closure at peak systole. We developed three metrics to evaluate simulation results. We validated our method on three explanted porcine hearts and showed that our model performs well. We evaluated the sensitivity of our model to changes in various parameters. We also measured the influence of the positions of the chordae tendineae on simulation results.

6.3. Interventional neuroradiology

Participants: Marie-Odile Berger, Charlotte Delmas, Erwan Kerrien, Raffaella Trivisonne.

Tools reconstruction for interventional neuro-radiology Minimally invasive techniques impact surgery in such ways that, in particular, an imaging modality is required to maintain a visual feedback. Live X-ray imaging, called fluoroscopy, is used in interventional neuroradiology. Such images are very noisy, and cannot show any brain tissue except the vasculature. In particular, since at most only two projective fluoroscopic views are available, containing absolutely no depth hint, the 3D shape of the micro-tool (guidewire, micro-catheter or micro-coil) can be very difficult, if not impossible to infer, which may have an impact on the clinical outcome of the procedure.

In collaboration with GE Healthcare, we aim at devising ways to reconstruct the micro-tools in 3D from fluoroscopy images. Charlotte Delmas has been working as a PhD CIFRE student on this subject since April 2013. A setup was designed in a view to reconstruct in 3D a deploying coil in as little X-ray dose and time as possible. It combines a fast rotation of both X-ray planes around the patient's head and a tomographic reconstruction combining an l_1-constraint to promote sparsity together with diffusion filters that promote the curvilinear nature of the coil. During this final year of her PhD thesis, various acquisition strategies and diffusion filters were evaluated [20].

Image driven simulation We consider image-driven simulation, applied to interventional neuroradiology as a coupled system of interactive computer-based simulation (interventional devices in blood vessels) and on-line medical image acquisitions (X-ray fluoroscopy). The main idea is to use the live X-ray images as references to continuously refine the parameters used to simulate the blood vessels and the interventional devices (micro-guide, micro-catheter, coil).

Raffaella Trivisonne started her PhD thesis in November 2015 (co-supervised by Stéphane Cotin, from MIMESIS team in Strasbourg) to address this research topic. Both projective and mechanical constraints were integrated in an augmented Lagrangian framework to solve the dynamical system. Experiments based on synthetic and phantom data were indicative that the shape from template problem could be solved without the need for considering collisions with the vessel surface, if an efficient tracking of the catheter in the X-ray images is available. These results were submitted for publication at a conference.

6.4. Assessing metrological performances in experimental mechanics

Participant: Frédéric Sur.

Progress was made during this year on several aspects of our collaboration with Institut Pascal on experimental mechanics. As mentioned in Section 4.3, the surface of the specimens under study are marked either by a regular grid, or by a random speckle. Displacement and strain maps are estimated by comparing images taken before and after deformation: through spectral methods (named here "the grid method") in the first case and through digital image correlation (DIC) in the latter.

Our contributions to the grid method are twofolds. First, we carefully analyzed the effect of digital sampling which causes aliasing [17]. We have proposed simple guidelines to minimize the effect of aliasing on strain maps. Second, we have mathematically characterized the properties of the analysis windows commonly used for processing grid images through the grid method [18]. It turns out that a Gaussian window has to be used, mainly because of its good concentration in both spatial and spectral domains in the sense of the Wigner-Ville transform. We eventually published a comprehensive review paper on the use of grid methods in experimental mechanics [15]

We also contributed to DIC-based methods. We have proposed new predictive formulas for the resolution of the displacement maps provided by DIC, which is mainly limited by sensor noise. These formulas take interpolation into account [12]. Indeed, displacement amplitude being often much smaller than one pixel, it is crucial to analyze the effect of the interpolation scheme. We have also proposed an experimental validation of these formula. This requires to take into account the heteroscedastic nature of sensor noise and rigid body motions caused by unavoidable vibrations [13].

MORPHEO Project-Team

7. New Results

7.1. Cotemporal Multi-View Video Segmentation

We address the problem of multi-view video segmentation of dynamic scenes in general and outdoor environments with possibly moving cameras. Multi-view methods for dynamic scenes usually rely on geometric calibration to impose spatial shape constraints between viewpoints. In this paper, we show that the calibration constraint can be relaxed while still getting competitive segmentation results using multi-view constraints. We introduce new multi-view cotemporality constraints through motion correlation cues, in addition to common appearance features used by co-segmentation methods to identify co-instances of objects. We also take advantage of learning based segmentation strategies by casting the problem as the selection of monocular proposals that satisfy multi-view constraints. This yields a fully automated method that can segment subjects of interest without any particular pre-processing stage, as depicted in Figure 2 . Results on several challenging outdoor datasets demonstrate the feasibility and robustness of our approach.

This work has been presented at the International Conference on 3D Vision (3DV) 2016 [9].

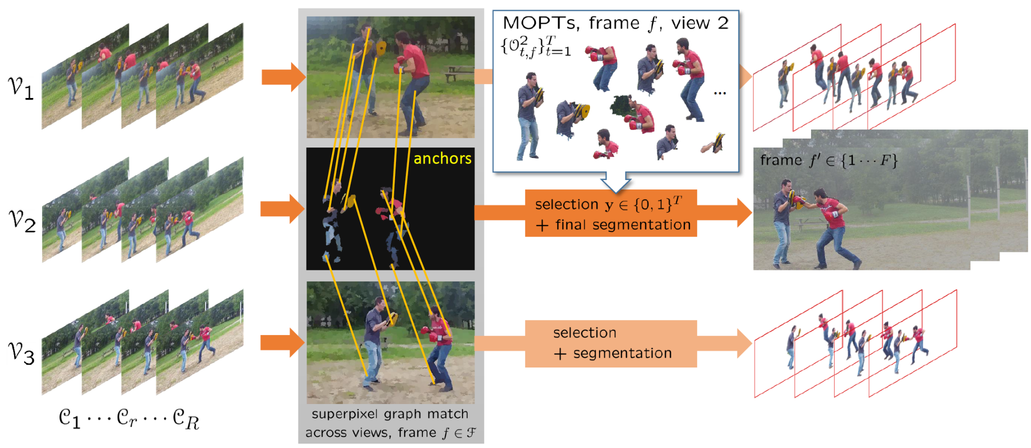


Figure 2. Overview of multiview segmentation pipeline

7.2. Volumetric Shape Reconstruction from Implicit Forms

In this work we evaluate volumetric shape reconstruction methods that consider as input implicit forms in 3D. Many visual applications build implicit representations of shapes that are converted into explicit shape representations using geometric tools such as the Marching Cubes algorithm. This is the case with image based reconstructions that produce point clouds from which implicit functions are computed, with for instance a Poisson reconstruction approach. While the Marching Cubes method is a versatile solution with proven efficiency, alternative solutions exist with different and complementary properties that are of interest for shape modeling. In this paper, we propose a novel strategy that builds on Centroidal Voronoi Tessellations (CVTs). These tessellations provide volumetric and surface representations with strong regularities in addition to provably more accurate approximations of the implicit forms considered. In order to compare the existing strategies, we present an extensive evaluation that analyzes various properties of the main strategies for implicit to explicit volumetric conversions: Marching cubes, Delaunay refinement and CVTs, including accuracy and shape quality of the resulting shape mesh.

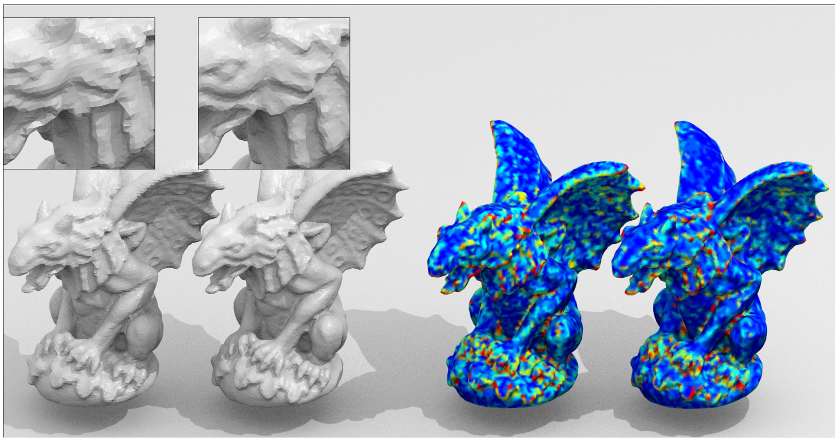


Figure 3. Poisson volumetric reconstructions from a Gargoyle point cloud with Marching Cubes (left) and CVT (right) [16]. Distances to the implicit form are color encoded on the right, from low (blue) to high (red).

This work has been presented at the ECCV 2016 conference [16].

7.3. Bayesian 3D imaging from X-rays and video

A new method for estimating 3D dense attenuation of moving samples such as body parts from multiple video and a single planar X-ray device has been devised [12]. Most dense modeling methods consider samples observed with a moving X-ray device and cannot easily handle moving samples. We proposed a novel method that uses a surface motion capture system associated to a single low-cost/low-dose planar X-ray imaging device for dense in-depth attenuation information. Our key contribution is to rely on Bayesian inference to solve for a dense attenuation volume given planar radioscopic images of a moving sample. The approach enables multiple sources of noise to be considered and takes advantage of 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 simulated and in-vivo data, as illustrated in Figure 4 .

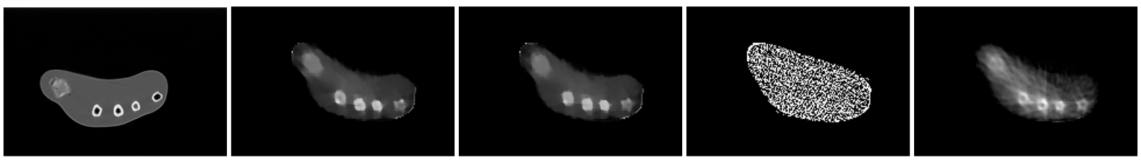


Figure 4. Results of the proposed method on a forearm phantom (2 selected slices). Left-to-right: ground-truth CT scan, proposed method, without optical flow, without TVL_1prior, ART. Without optical flow, artefacts are visible, for example in the bone cavities. The ART method produces much noisier results.

7.4. Robust Multilinear Model Learning Framework for 3D Faces

Statistical models are widely used to represent the variations of 3D human faces. Multilinear models in particular are common as they decouple shape changes due to identity and expression. Existing methods to learn a multilinear face model degrade if not every person is captured in every expression, if face scans are noisy or partially occluded, if expressions are erroneously labeled, or if the vertex correspondence is inaccurate. These limitations impose requirements on the training data that disqualify large amounts of available 3D face data from being usable to learn a multilinear model. To overcome this, we have developed an effective framework to robustly learn a multilinear model from 3D face databases with missing data, corrupt data, wrong semantic correspondence, and inaccurate vertex correspondence. To achieve this robustness to erroneous training data, our framework jointly learns a multilinear model and fixes the data. This framework is significantly more efficient than prior methods based on linear statistical models. This work was presented at CVPR 2016 [7].

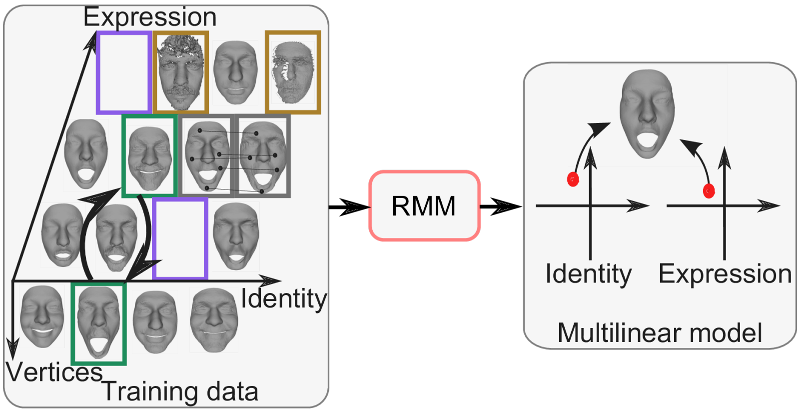


Figure 5. Overview of our robust multilinear model (RMM) learning framework that is robust to missing data (purple), corrupt data (brown), wrong semantic correspondence (green), and inaccurate vertex correspondence (gray).

7.5. Segmentation of Tree Seedling Point Clouds into Elementary Units

We propose a new semi-automatic method to cluster TLS data into meaningful sets of points to extract plant components. The approach is designed for small plants with distinguishable branches and leaves, such as tree seedlings. It first creates a graph by connecting each point to its most relevant neighbours, then embeds the graph into a spectral space, and finally segments the embedding into clusters of points. The process can then be iterated on each cluster separately. The main idea underlying the approach is that the spectral embedding of the graph aligns the points along the shape's principal directions. A quantitative evaluation of the segmentation accuracy, as well as of leaf area estimates, is provided on a poplar seedling mock-up. It shows that the segmentation is robust with false positive and false negative rates around 1%. Qualitative results on four contrasting plant species with three different scan resolution levels each are also shown in the paper, which has been published in the International Journal of Remote Sensing [2].

7.6. Estimation of Human Body Shape in Motion with Wide Clothing

Estimating 3D human body shape in motion from a sequence of unstructured oriented 3D point clouds is important for many applications. We propose the first automatic method to solve this problem that works in the presence of loose clothing. The problem is formulated as an optimization problem that solves for identity and posture parameters in a shape space capturing likely body shape variations. The automation is achieved by leveraging a recent robust pose detection method Stitched Puppet. To account for clothing, we take advantage of motion cues by encouraging the estimated body shape to be inside the observations. The method is evaluated on a new benchmark containing different subjects, motions, and clothing styles that allows to quantitatively measure the accuracy of body shape estimates. Furthermore, we compare our results to existing methods that require manual input and demonstrate that results of similar visual quality can be obtained.

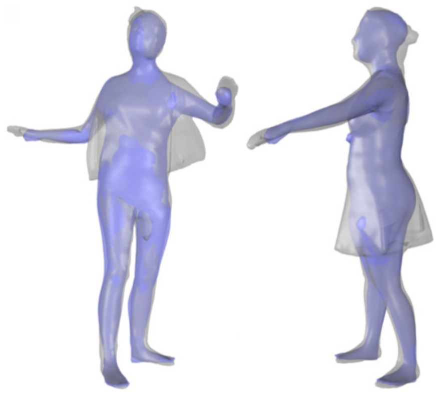


Figure 6. Two frames of an input point cloud sequence (in gray) with the estimated body shape shown in blue [14].

This work has been presented at the ECCV 2016 conference [14].

7.7. Computing Temporal Alignments of Human Motion Sequences in Wide Clothing using Geodesic Patches

In this work, we address the problem of temporal alignment of surfaces for subjects dressed in wide clothing, as acquired by calibrated multi-camera systems. Most existing methods solve the alignment by fitting a single surface template to each instant's 3D observations, relying on a dense point-to-point correspondence scheme, e.g. by matching individual surface points based on local geometric features or proximity. The wide clothing situation yields more geometric and topological difficulties in observed sequences, such as apparent merging of surface components, misreconstructions, and partial surface observation, resulting in overly sparse, erroneous point-to-point correspondences, and thus alignment failures. To resolve these issues, we propose an alignment framework where point-to-point correspondences are obtained by growing isometric patches from a set of reliably obtained body landmarks. This correspondence decreases the reliance on local geometric features subject to instability, instead emphasizing the surface neighborhood coherence of matches, while improving density given sufficient landmark coverage. We validate and verify the resulting improved alignment performance in our experiments.
This work has been presented at the International Conference on 3D Vision (3DV) 2016 [13].

7.8. A 3D+t Laplace Operator for Temporal Mesh Sequences

The Laplace operator plays a fundamental role in geometry processing. Several discrete versions have been proposed for 3D meshes and point clouds, among others. We have defined a discrete Laplace operator for temporally coherent mesh sequences, which allows to process mesh animations in a simple yet efficient way. This operator is a discretization of the Laplace-Beltrami operator using Discrete Exterior Calculus on CW complexes embedded in a four-dimensional space. A parameter is introduced to tune the influence of the motion with respect to the geometry. This enables straightforward generalization of existing Laplacian static mesh processing works to mesh sequences. An application to spacetime editing has been provided as example.

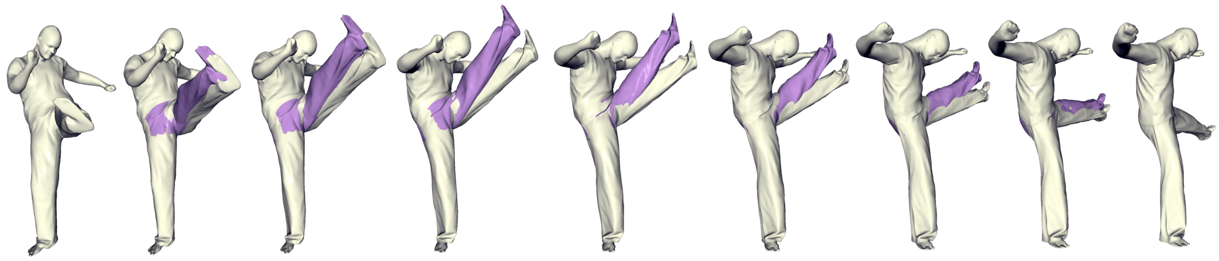


Figure 7. Spacetime editing of a temporal mesh sequence using the proposed 3D+t Laplace operator [1].

This work has been published in Computer & Graphics [1] and presented at the Shape Modeling International (SMI) 2016 conference.

7.9. Volumetric 3D Tracking by Detection

In this collaboration with TU Munich, we investigated a new solutions for 3D tracking by detection based on fully volumetric representations. On one hand, 3D tracking by detection has shown robust use in the context of interaction (Kinect) and surface tracking. On the other hand, volumetric representations have recently been proven efficient both for building 3D features and for addressing the 3D tracking problem. We leveraged these benefits by unifying both families of approaches into a single, fully volumetric tracking-by-detection framework. We used a centroidal Voronoi tessellation (CVT) representation to compactly tessellate shapes with optimal discretization, construct a feature space, and perform the tracking according to the correspondences provided by trained random forests (see figure 8). Our results show improved tracking and training computational efficiency and improved memory performance. This in turn enables the use of larger training databases than state of the art approaches, which we leveraged by proposing a cross-tracking subject training scheme to benefit from all subject sequences for all tracking situations, thus yielding better detection and less overfitting. The approach has been presented at CVPR 2016 [10].

7.10. Eigen Appearance Maps of Dynamic Shapes

In this work, we considered the problem of building efficient appearance rep- resentations of shapes observed from multiple viewpoints and in several movements. Multi-view systems now allow the acquisition of spatio-temporal models of such moving objects. While efficient geometric representations for these models have been widely studied, appearance information, as provided by the observed images, is mainly considered on a per frame basis, and no global strategy yet addresses the case where several temporal sequences of a shape are available. We proposed a per subject representation that builds on PCA to identify the underlying manifold structure of the appearance information relative to a shape. The resulting eigen representation encodes shape appearance variabilities due to viewpoint and motion, with Eigen textures, and due to local inaccuracies in the geometric model, with Eigen warps. In addition to providing compact representations, such decompositions also allow for appearance interpolation and appearance completion. We evaluated their performances over different characters and with respect to their ability to reproduce compelling appearances in a compact way. This work was presented at ECCV 2016.

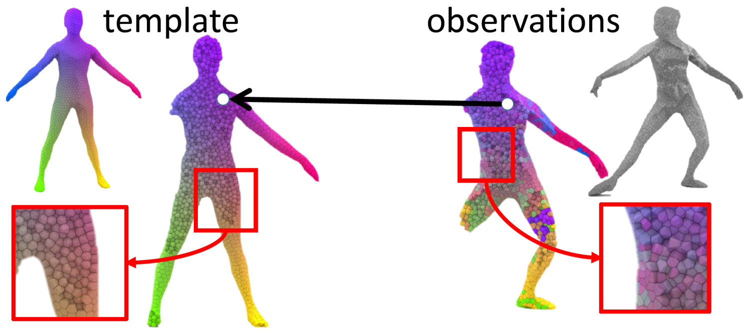


Figure 8. 3D shapes are represented using centroidal Voronoi tessellations. The volumetric cells of the observations are matched to cells of the template.

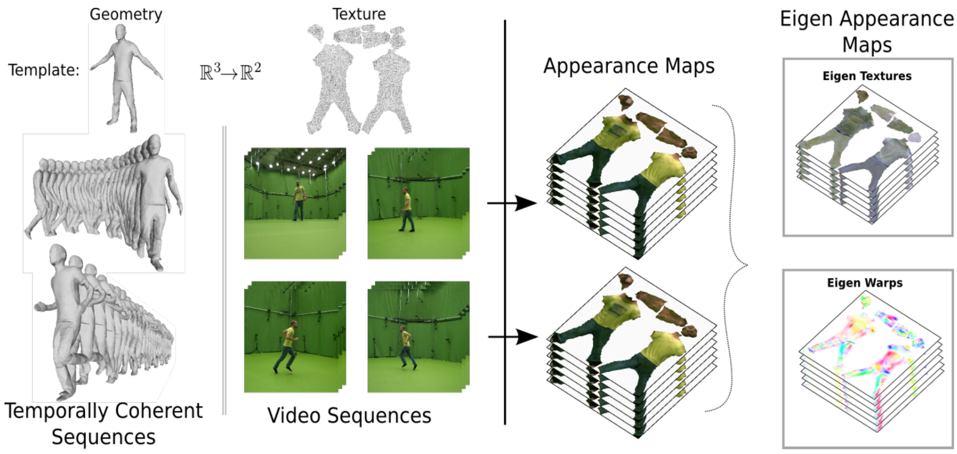


Figure 9. Given time consistent shape models and their appearance maps, our method exploits the manifold structure of these appearance information through PCA decomposition to generate the Eigen appearance maps relative to a shape.

7.11. Visual Contrast Sensitivity and Discrimination for 3D Meshes

In this work, we first introduce an algorithm for estimating the visual contrast on a 3D mesh. We then perform a series of psychophysical experiments to study the effects of contrast sensitivity and contrast discrimination of the human visual system for the task of differentiating between two contrasts on a 3D mesh. The results of these experiments allow us to propose a perceptual model that is able to predict whether a change in local contrast on 3D mesh, induced by a local geometric distortion, is visible or not. Finally, we illustrate the utility of the proposed perceptual model in a number of applications: we compute the Just Noticeable Distortion (JND) profile for smooth-shaded 3D meshes and use the model to guide mesh processing algorithms.
This work has been published in Computer Graphics Forum [] and has received the best paper award at the Pacific Graphics 2016 conference.

PERCEPTION Project-Team

6. New Results

6.1. Audio-Source Localization

In previous years we have developed several *supervised* sound-source localization algorithms. The general principle of these algorithms was based on the learning of a mapping (regression) between binaural feature vectors and source locations [5], [7]. While fixed-length wide-spectrum sounds (white noise) are used for training to reliably estimate the model parameters, we show that the testing (localization) can be extended to variable-length sparse-spectrum sounds (such as speech), thus enabling a wide range of realistic applications. Indeed, we demonstrate that the method can be used for audio-visual fusion, namely to map speech signals onto images and hence to spatially align the audio and visual modalities, thus enabling to discriminate between speaking and non-speaking faces. We released a novel corpus of real-room recordings that allow quantitative evaluation of the co-localization method in the presence of one or two sound sources. Experiments demonstrate increased accuracy and speed relative to several state-of-the-art methods. During the period 2015-2016 we extended this method to an arbitrary number of microphones based on the *relative transfer function – RTF* (between any channel and a reference channel). Then we extended this work and developed a novel transfer function that contains the direct path between the source and the microphone array, namely the *direct-path relative transfer function* [29], [36].

Websites:
https://team.inria.fr/perception/research/acoustic-learning/
https://team.inria.fr/perception/research/binaural-ssl/
https://team.inria.fr/perception/research/ssl-rtf/

6.2. Audio-Source Separation

We address the problem of separating audio sources from time-varying convolutive mixtures. We proposed an unsupervised probabilistic framework based on the local complex-Gaussian model combined with non-negative matrix factorization [33], [28]. The time-varying mixing filters are modeled by a continuous temporal stochastic process. This model extends the case of static filters which corresponds to static audio sources. While static filters can be learnt in advance, e.g. [5], time-varying filters cannot and therefore the problem is more complex. We present a variational expectation-maximization (VEM) algorithm that employs a Kalman smoother to estimate the time-varying mixing matrix, and that jointly estimates the source parameters. The sound sources are then separated by Wiener filters constructed with the estimators provided by the VEM algorithm. Extensive experiments on simulated data show that the proposed method outperforms a block-wise version of a state-of-the-art baseline method. This work is part of the PhD topic of Dionyssos Kounades Bastian and is conducted in collaboration with Sharon Gannot (Bar Ilan University) and Xavier Alameda Pineda (University of Trento). Our journal paper [28] is an extended version of a paper presented at IEEE WASPAA in 2015 which received the best student paper award.

Website:
https://team.inria.fr/perception/research/vemove/
https://team.inria.fr/perception/research/nmfig/

6.3. Single-Channel Audio Processing

While most of our audio scene analysis work involves microphone arrays, it is important to develop single-channel (one microphone) signal processing methods as well. In particular, it is important to detect speech signal (or voice) in the presence of various types of noise (stationary or non-stationary). In this context, we developed the following methods [39], [37]:

- Statistical likelihood ratio test is a widely used voice activity detection (VAD) method, in which the likelihood ratio of the current temporal frame is compared with a threshold. A fixed threshold is always used, but this is not suitable for various types of noise. In this work, an adaptive threshold is proposed as a function of the local statistics of the likelihood ratio. This threshold represents the upper bound of the likelihood ratio for the non-speech frames, whereas it remains generally lower than the likelihood ratio for the speech frames. As a result, a high non-speech hit rate can be achieved, while maintaining speech hit rate as large as possible.
- Estimating the noise power spectral density (PSD) is essential for single channel speech enhancement algorithms. We propose a noise PSD estimation approach based on regional statistics which consist of four features representing the statistics of the past and present periodograms in a short-time period. We show that these features are efficient in characterizing the statistical difference between noise PSD and noisy-speech PSD. We therefore propose to use these features for estimating the speech presence probability (SPP). The noise PSD is recursively estimated by averaging past spectral power values with a time-varying smoothing parameter controlled by the SPP. The proposed method exhibits good tracking capability for non-stationary noise, even for abruptly increasing noise level.

Website:
https://team.inria.fr/perception/research/noise-psd/

6.4. Tracking Multiple Persons

Object tracking is an ubiquitous problem in computer vision with many applications in human-machine and human-robot interaction, augmented reality, driving assistance, surveillance, etc. Although thoroughly investigated, tracking multiple persons remains a challenging and an open problem. In this work, an online variational Bayesian model for multiple-person tracking is proposed. This yields a variational expectation-maximization (VEM) algorithm. The computational efficiency of the proposed method is made possible thanks to closed-form expressions for both the posterior distributions of the latent variables and for the estimation of the model parameters. A stochastic process that handles person birth and person death enables the tracker to handle a varying number of persons over long periods of time [24], [30].

Website:
https://team.inria.fr/perception/research/ovbt/

6.5. Audio-Visual Speaker Detection, Localization, and Diarization

Any multi-party conversation system benefits from speaker diarization, that is, the assignment of speech signals among the participants. More generally, in HRI and CHI scenarios it is important to recognize the speaker over time. We propose to address speaker detection, localization and diarization using both audio and visual data. 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 [25], 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 data-sets that are available from recent contributions which are used as baselines for comparison [26].

Websites:

https://team.inria.fr/perception/research/wdgmm/
https://team.inria.fr/perception/research/speakerloc/
https://team.inria.fr/perception/research/speechturndet/
https://team.inria.fr/perception/research/avdiarization/

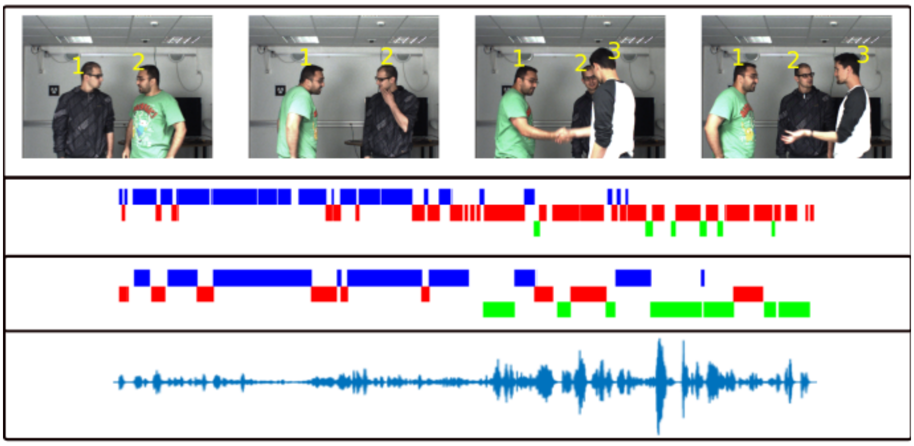


Figure 3. This figure illustrates the audiovisual tracking and diarization method that we have recently developed. First row: A number is associated with each tracked person. Second row: diarization result. Third row: the ground truth diarization. Fourth row: acoustic signal recorded by one of the two microphones.

6.6. Head Pose Estimation and Tracking

Head pose estimation is an important task, because it provides information about cognitive interactions that are likely to occur. Estimating the head pose is intimately linked to face detection. We addressed the problem of head pose estimation with three degrees of freedom (pitch, yaw, roll) from a single image and in the presence of face detection errors. Pose estimation is formulated as a high-dimensional to low-dimensional mixture of linear regression problem [6]. We propose a method that maps HOG-based descriptors, extracted from face bounding boxes, to corresponding head poses. To account for errors in the observed bounding-box position, we learn regression parameters such that a HOG descriptor is mapped onto the union of a head pose and an offset, such that the latter optimally shifts the bounding box towards the actual position of the face in the image. The performance of the proposed method is assessed on publicly available datasets. The experiments that we carried out show that a relatively small number of locally-linear regression functions is sufficient to deal with the non-linear mapping problem at hand. Comparisons with state-of-the-art methods show that our method outperforms several other techniques [42]. This work is part of the PhD of Vincent Drouard and it received the best student paper award (second place) at the IEEE ICIP'15. Currently we investigate a temporal extension of this model.

Website:
https://team.inria.fr/perception/research/head-pose/

6.7. Estimation of Eye Gaze and of Visual Focus of Attention

We address the problem of estimating the visual focus of attention (VFOA), e.g. who is looking at whom? This is of particular interest in human-robot interactive scenarios, e.g. when the task requires to identify targets of interest and to track them over time. We make the following contributions. We propose a Bayesian temporal model that links VFOA to eye-gaze direction and to head orientation. Model inference is cast into a switching Kalman filter formulation, which makes it tractable. The model parameters are estimated via training based on manual annotations. The method is tested and benchmarked using a publicly available dataset. We show that both eye-gaze and VFOA of several persons can be reliably and simultaneously estimated and tracked over time from observed head poses as well as from people and object locations [40].

Website:
https://team.inria.fr/perception/research/eye-gaze/.

6.8. High-Resolution Scene Reconstruction

We addressed the problem of range-stereo fusion for the construction of high-resolution depth maps. In particular, we combine time-of-flight (low resolution) depth [27] data with high-resolution stereo data, in a maximum a posteriori (MAP) formulation. Unlike existing schemes that build on MRF optimizers, we infer the disparity map from a series of local energy minimization problems that are solved hierarchically, by growing sparse initial disparities obtained from the depth data. The accuracy of the method is not compromised, owing to three properties of the data-term in the energy function. Firstly, it incorporates a new correlation function that is capable of providing refined correlations and disparities, via sub-pixel correction. Secondly, the correlation scores rely on an adaptive cost aggregation step, based on the depth data. Thirdly, the stereo and depth likelihoods are adaptively fused, based on the scene texture and camera geometry. These properties lead to a more selective growing process which, unlike previous seed-growing methods, avoids the tendency to propagate incorrect disparities. The proposed method gives rise to an intrinsically efficient algorithm, which runs at 3FPS on 2.0MP images on a standard desktop computer. The strong performance of the new method is established both by quantitative comparisons with state-of-the-art methods, and by qualitative comparisons using real depth-stereo data-sets [8]. This work is funded by the ANR project MIXCAM.

Website:
https://team.inria.fr/perception/research/dsfusion/

6.9. 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 [43].

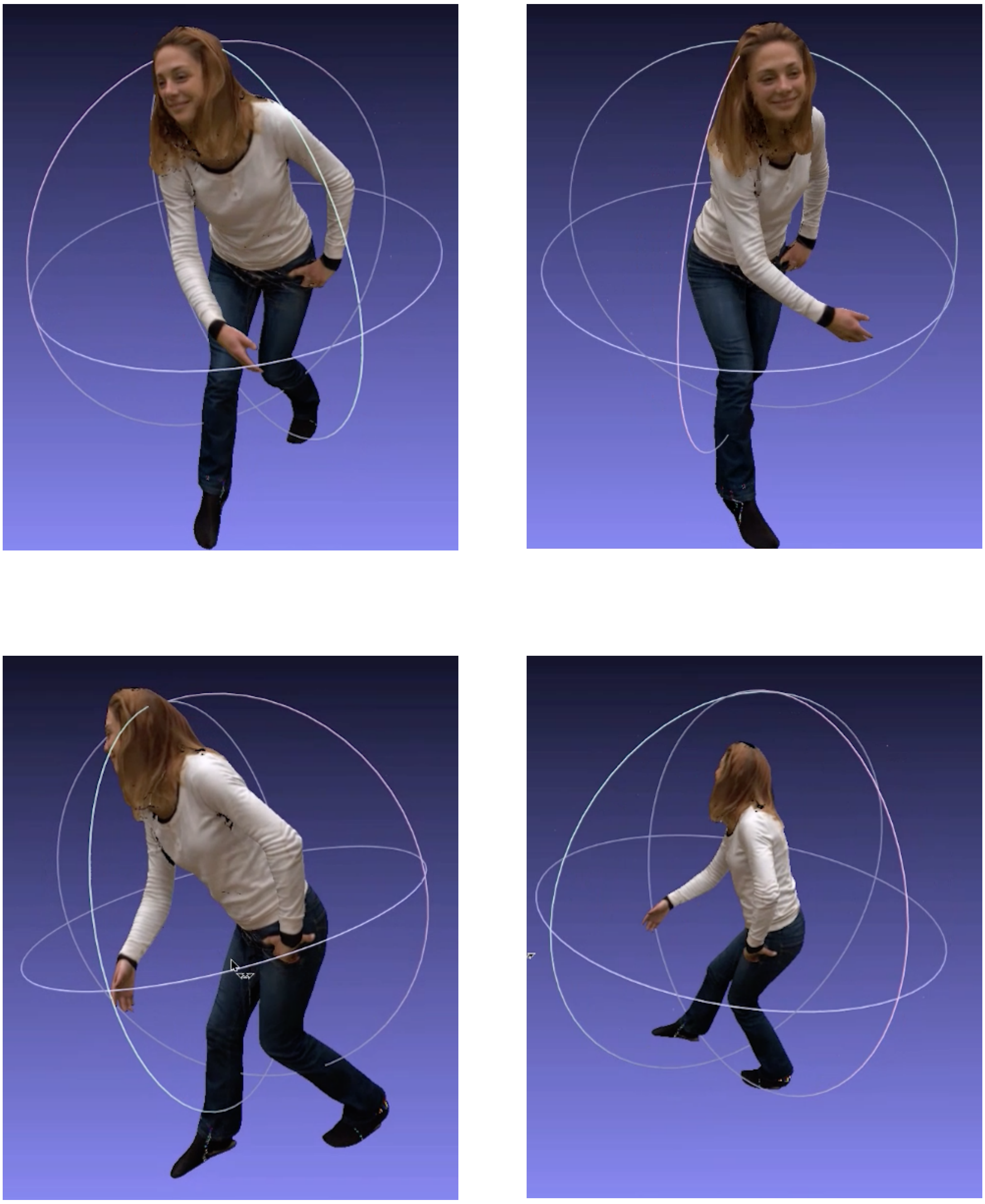


Figure 4. Four views of a 3D person reconstructed with our algorithm. In this example we used a large number of high-resolution cameras and the rendering was performed by the software of 4D View Solutions.

SIROCCO Project-Team

7. New Results

7.1. Analysis and modeling for compact representation

3D modelling, multi-view plus depth videos, light-fields, 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

7.1.1. Visual attention

Participant: Olivier Le Meur.

Visual attention is the mechanism allowing to focus our visual processing resources on behaviorally relevant visual information. Two kinds of visual attention exist: one involves eye movements (overt orienting) whereas the other occurs without eye movements (covert orienting). Our research activities deal with the understanding and modeling of overt attention.

Saccadic model: Previous research showed the existence of systematic tendencies in viewing behavior during scene exploration. For instance, saccades are known to follow a positively skewed, long-tailed distribution, and to be more frequently initiated in the horizontal or vertical directions. In 2016, we investigated the fact that these viewing biases are not universal, but are modulated by the semantic visual category of the stimulus. We showed that the joint distribution of saccade amplitudes and orientations significantly varies from one visual category to another. These joint distributions turn out to be, in addition, spatially variant within the scene frame. We demonstrated that a saliency model based on this better understanding of viewing behavioral biases and blind to any visual information outperforms well-established saliency models. We also proposed an extension of the saccadic model developed in 2015. The improvement consists in accounting for spatially-variant and context-dependent viewing biases. This model outperforms state-of-the-art saliency models, and provides scanpaths in close agreement with human behavior.

Inference of age from eye movements: We have presented evidence that information derived from eye gaze can be used to infer observers' age. From simple features extracted from the sequence of fixations and saccades, we predict the age of an observer. To reach this objective, we used the eye data from 101 observers split in 4 age groups (adults, 6-10 year-old, 4-6 year-old. and 2 year-old) to train a computational model. Participant's eye movements were monitored while participants were instructed to explore color pictures taken from children books for 10 seconds. The analysis of eye gaze provided evidence of age-related differences in viewing patterns. Fixation durations decreased with age while saccades turned out to be shorter when comparing children with adults. We combine several features, such as fixation durations, saccade amplitudes, and learn a direct mapping from those features to age using Gentle AdaBoost classifiers. Experimental results show that the proposed method succeeds in predicting reasonably well the observer's age.

7.1.2. Graph structure in the rays space for fast light fields segmentation

Participants: Christine Guillemot, Matthieu Hog.

In collaboration with Technicolor (Neus Sabater), we have introduced a novel graph representation for interactive light field segmentation using Markov Random Field (MRF). The greatest barrier to the adoption of MRF for light field processing is the large volume of input data. The proposed graph structure exploits the redundancy in the ray space in order to reduce the graph size, decreasing the running time of MRF-based optimisation tasks. The concepts of free rays and ray bundles with corresponding neighbourhood relationships are defined to construct the simplified graph-based light field representation. We have then developed a light field interactive segmentation algorithm using graph-cuts based on such ray space graph structure, that guarantees the segmentation consistency across all views. Our experiments with several datasets show results that are very close to the ground truth, competing with state of the art light field segmentation methods in terms of accuracy and with a significantly lower complexity. They also show that our method performs well on both densely and sparsely sampled light fields [18] (see Figure 1).

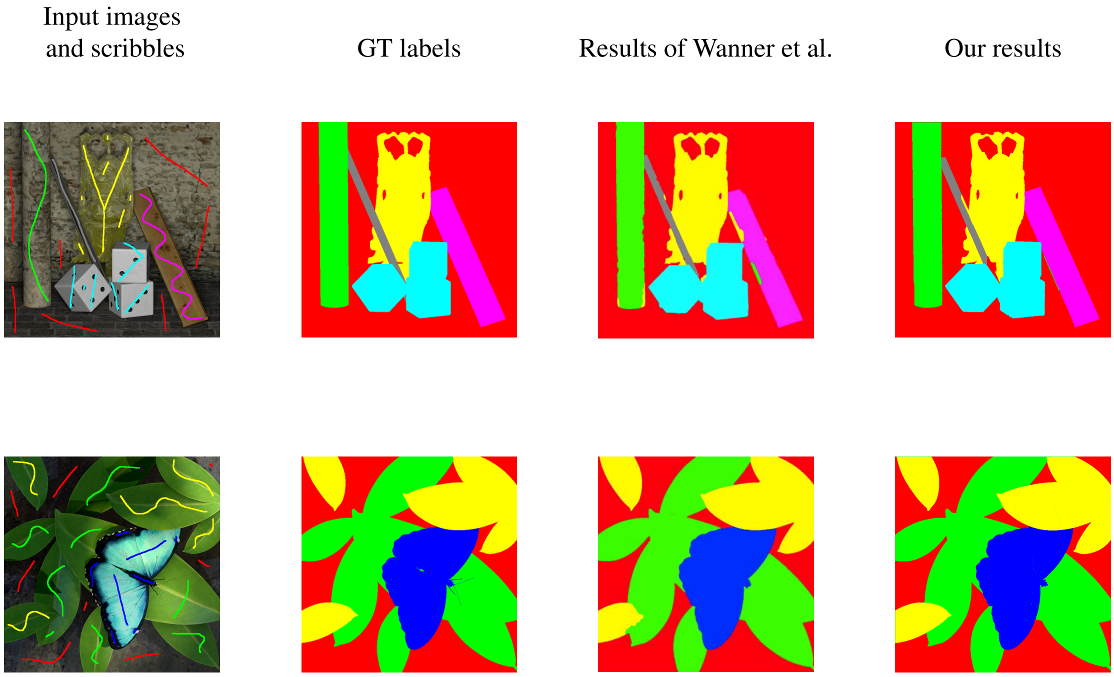


Figure 1. Light-field segmentation results obtained with synthetic light-fields. From left to right, we show, the input central view with scribbles, the ground truth labelling, the results of Wanner et al. and our results.

7.2. Rendering, inpainting and super-resolution

image-based rendering, inpainting, view synthesis, super-resolution

7.2.1. Joint color and gradient transfer through Multivariate Generalized Gaussian Distribution

Participants: Hristina Hristova, Olivier Le Meur.

Multivariate generalized Gaussian distributions (MGGDs) have aroused a great interest in the image processing community thanks to their ability to describe accurately various image features, such as image gradient fields, wavelet coefficients, etc. However, so far their applicability has been limited by the lack of a transformation between two of these parametric distributions. In collaboration with FRVSense (Rémi Cozot and Kadi Bouatouch), we have proposed a novel transformation between MGGDs, consisting of an optimal transportation of the second-order statistics and a stochastic-based shape parameter transformation. We employ the proposed transformation in both color and gradient transfers between images. We have also proposed a new simultaneous transfer of color and gradient.

7.2.2. High-Dynamic-Range Image Recovery from Flash and Non-Flash Image Pairs

Participants: Hristina Hristova, Olivier Le Meur.

In 2016, in collaboration with FRVSense (Rémi Cozot and Kadi Bouatouch), we have proposed a novel method for creating High Dynamic Range (HDR) images from only two images - flash and non-flash images. The proposed method consists of two main steps, namely brightness gamma correction and bi-local chromatic adaptation transform (CAT). First, the brightness gamma correction performs series of increases and decreases of the brightness of the non-flash image and that way yields multiple images with various exposure values. Second, a proposed CAT method, called bi-local CAT enhances the quality of the computed images, by recovering details in the under-/over-exposed regions, using detail information from the flash image. The final multiple exposure images are then merged together to compute an HDR image. Evaluation shows that our HDR images, obtained by using only two LDR images, are close to HDR images, obtained by combining five manually taken multi-exposure images. The proposed method does not require the usage of a tripod and it is suitable for images of non-still objects, such as people, candle flames, etc. Figure 2 illustrates some results of the proposed method. The HDR-VDP-2 color-coded map (right-most image) shows the main luminance differences (the red areas) between our HDR result and the real HDR image. Snippets (a) and (b) show that the proposed method sharpens fine details, e.g. the net on the lamp. The net on the lamp of the real HDR image is blurry, due to a movement in the real multi-exposure images.

Figure 2. HDR image recovery from two input images, i.e. flash and non-flash images. Our HDR result and the real HDR image are tone-mapped for visualization on an LDR display.

7.2.3. Depth inpainting

Participant: Olivier Le Meur.

To tackle the disocclusion inpainting of RGB-D images appearing when synthesizing new views of a scene by changing its viewpoint, in collaboration with Pierre Buyssens from the Greyc laboratory from the Caen University, we have developed a new examplar-based inpainting method of depth map. The proposed method is based on two main components. First, a novel algorithm to perform the depth-map disocclusion inpainting has been proposed. In particular, this intuitive approach is able to recover the lost structures of the objects and to inpaint the depth-map in a geometrically plausible manner. Then, a depth-guided patch-based inpainting method has been defined in order to fill-in the color image. Depth information coming from the reconstructed depth-map is added to each key step of the classical patch-based algorithm from Criminisi et al. in an intuitive manner. Relevant comparisons to state-of-the-art inpainting methods for the disocclusion inpainting of both depth and color images have illustrated the effectiveness of the proposed algorithms.

7.2.4. Super-resolution and inpainting for face recognition

Participants: Reuben Farrugia, Christine Guillemot.

Most face super-resolution methods assume that low- and high-resolution manifolds have similar local geometrical structure, hence learn local models on the low-resolution manifold (e.g. sparse or locally linear embedding models), which are then applied on the high-resolution manifold. However, the low- resolution manifold is distorted by the one-to-many relationship between low- and high- resolution patches.

We have developed a method which learns linear models based on the local geometrical structure on the high-resolution manifold rather than on the low-resolution manifold. For this, in a first step, the low-resolution patch is used to derive a globally optimal estimate of the high-resolution patch. The approximated solution is shown to be close in Euclidean space to the ground-truth but is generally smooth and lacks the texture details needed by state-of-the-art face recognizers. Unlike existing methods, the sparse support that best estimates the first approximated solution is found on the high-resolution manifold. The derived support is then used to extract the atoms from the coupled dictionaries that are most suitable to learn an upscaling function between the low- and high- resolution patches.

The proposed solution has also been extended to compute face super-resolution of non-frontal images. Experimental results show that the proposed method out- performs six face super-resolution and a state-of-the-art cross- resolution face recognition method. These results also reveal that the recognition and quality are significantly affected by the method used for stitching all super-resolved patches together, where quilting was found to better preserve the texture details which helps to achieve higher recognition rates. The proposed method was shown to be able to super-resolve facial images from the IARPA Janus Benchmark A (IJB-A) dataset which considers a wide range of poses and orientations.

A method has also been developed to inpaint occluded facial regions with unconstrained pose and orientation. This approach first warps the facial region onto a reference model to synthesize a frontal view [15]. A modified Robust Principal Component Analysis (RPCA) approach is then used to suppress warping errors. It then uses a novel local patch-based face inpainting algorithm which hallucinates missing pixels using a dictionary of face images which are pre-aligned to the same reference model. The hallucinated region is then warped back onto the original image to restore missing pixels. Experimental results on synthetic occlusions demonstrate that the proposed face inpainting method has the best performance achieving PSNR gains of up to 0.74dB over the second-best method. Moreover, experiments on the COFW dataset and a number of real-world images show that the proposed method successfully restores occluded facial regions in the wild even for Closed-Circuit Television (CCTV) quality images.

7.2.5. *Light-field inpainting*

Participants: Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

Building up on the advances in low rank matrix completion, we have developed a novel method for propagating the inpainting of the central view of a light field to all the other views. After generating a set of warped versions of the inpainted central view with random homographies, both the original light field views and the warped ones are vectorized and concatenated into a matrix. Because of the redundancy between the views, the matrix satisfies a low rank assumption enabling us to fill the region to inpaint with low rank matrix completion. To this end, a new matrix completion algorithm, better suited to the inpainting application than existing methods, has also been developed. Unlike most of the existing light field inpainting algorithms, our method does not require any depth prior. Another interesting feature of the low rank approach is its ability to cope with color and illumination variation between the input views of the light field (see Fig.3 . As it can be seen in Figure 3 , the proposed method yields inpainting consistency across views.

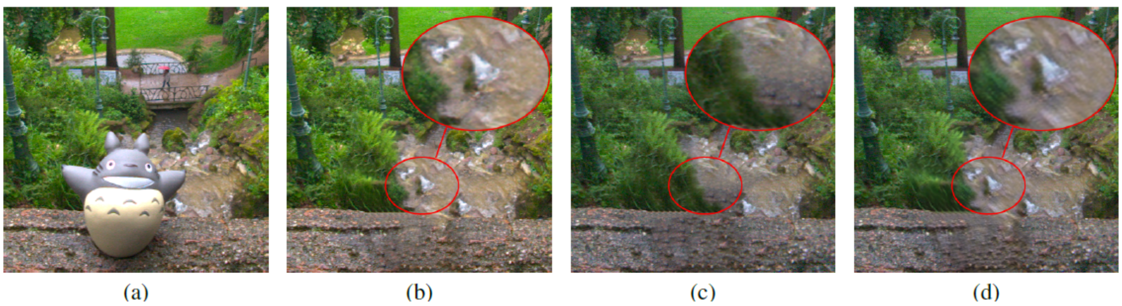


Figure 3. Illustration of our inpainting propagation method : (a) Original central view. (b) Inpainted central view. (c) Another view of the light field inpainted with a state-of-the-art 2D image inpainting method. (d) Propagated inpainting from central view to a different view with the developed low rank method.

7.3. Representation and compression of large volumes of visual data

Sparse representations, data dimensionality reduction, compression, scalability, perceptual coding, rate-distortion theory

7.3.1. *Graph-based multi-view video representation*

Participants: Christine Guillemot, Thomas Maugey, Mira Rizkallah, Xin Su.

One of the main open questions in multiview data processing is the design of representation methods for multiview data, where the challenge is to describe the scene content in a compact form that is robust to lossy data compression. Many approaches have been studied in the literature, such as the multiview and multiview plus depth formats, point clouds or mesh-based techniques. All these representations contain two types of data: i) the color or luminance information, which is classically described by 2D images; ii) the geometry information that describes the scene 3D characteristics, represented by 3D coordinates, depth maps or disparity vectors. Effective representation, coding and processing of multiview data partly rely on a proper representation of the geometry information. The multiview plus depth (MVD) format has become very popular in recent years for 3D data representation. However, this format induces very large volumes of data, hence the need for efficient compression schemes. On the other hand, lossy compression of depth information in general leads to annoying rendering artefacts especially along the contours of objects in the scene. Instead of lossy compression of depth maps, we consider the lossless transmission of a geometry representation that captures only the information needed for the required view reconstructions.

The goal is thus to develop a Graph-Based Representation (GBR) for geometry information, where the geometry of the scene is represented as connections between corresponding pixels in different views. In this representation, two connected pixels are neighboring points in the 3D scene. The graph connections are derived from dense disparity maps and provide just enough geometry information to predict pixels in all the views that have to be synthesized. GBR drastically simplifies the geometry information to the bare minimum required for view prediction. This "task-aware" geometry simplification allows us to control the view prediction accuracy before coding compared to baseline depth compression methods. In 2015, we have first considered multi-view configurations, in which cameras are parallel.

In 2016, we have developed the extension of GBR to complex camera configurations. In [21], Xin Su has implemented a generalized Graph-Based Representation handling two views with complex translations and rotations between them (Fig. 4). The proposed approach uses the epipolar segments to have a row-wise description of the geometry that is as simple as for rectified views. This generalized GBR has been further extended to handle multiple views and scalable description of the geometry, *i.e.*, a geometry data that is coded as a function of the user navigation among the views.

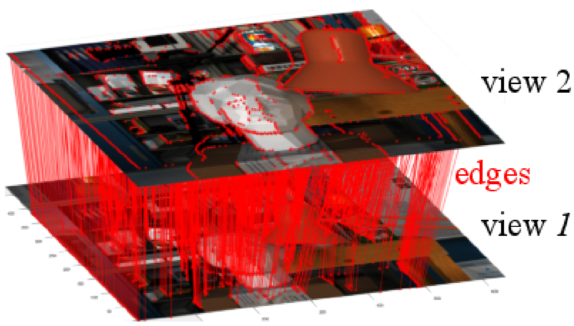


Figure 4. The proposed GBR (i) provides edges describing the geometry information and (ii) link pixels that are neighbors in the 3D scene.

The graph described above links neighboring pixels in the 3D scene as 3D meshes do. This meaningful structure might be used to code the color pixels lying on it. This can be done thanks to the new processing tools developed for signals lying on graphs. These tools rely however on covariance models that are assumed to be suited for the processed data. The PhD work of Mira Rizkallah is currently focussing on the effect of errors in the correlation models on the efficiency of the graph-based transforms.

7.3.2. Sparse and low rank approximation of light fields

Participants: Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

We have studied the problem of low rank approximation of light fields for compression. A homography-based approximation method has been proposed which jointly searches for homographies to align the different views of the light field together with the low rank approximation matrices. We have first considered a global homography per view and shown that depending on the variance of the disparity across views, the global homography is not sufficient to well-align the entire images. In a second step, we have thus considered multiple homographies, one per region, the region being extracted using depth information. We have first shown the benefit of the joint optimization of the homographies together with the low-rank approximation. The resulting compact representation compressed using HEVC yields compression performance significantly superior to those obtained by directly applying HEVC on the light field views re-structured as a video sequence.

7.3.3. Deep learning, autoencoders and neural networks for sparse representation and compression

Participants: Thierry Dumas, Christine Guillemot, Aline Roumy.

Deep learning is a novel research area that attempts to extract high level abstractions from data by using a graph with multiple layers. One could therefore expect that deep learning might allow efficient image compression based on these high level features. However, deep learning, as classical machine learning, consists in two phases: (i) build a graph that can make a good representation of the data (i.e. find an architecture usually made with neural nets), and (ii) learn the parameters of this architecture from large-scale data. As a consequence, neural nets are well suited for a specific task (text or image recognition) and require one training per task. The difficulty to apply machine learning approach to image compression is that it is important to deal with a large variety of patches, and with also various compression rates. To test the ability of neural networks to compress images, we studied shallow sparse autoencoders (AE) for image compression in [14]. A performance analysis in terms of rate-distortion trade-off and complexity is conducted, comparing sparse AEs with LARS-Lasso, Coordinate Descent (CoD) and Orthogonal Matching Pursuit (OMP). A Winner Take All Auto-encoder (WTA AE) is proposed where image patches compete with one another when computing their sparse representation. This allows to spread the sparsity constraint on the whole image. Since the learning is made for this WTA AE, the neural network also learns to deal with various patches, which helps building a general-purpose AE. Finally, we showed that, WTA AE achieves the best rate-distortion trade-off, is robust to quantization noise and it is less complex than LARS-Lasso, CoD and OMP.

7.3.4. Data geometry aware local basis selection

Participants: Julio Cesar Ferreira, Christine Guillemot.

Local learning of sparse image models has proven to be very effective to solve a variety of inverse problems in many computer vision applications. To learn such models, the data samples are often clustered using the K-means algorithm with the Euclidean distance as a dissimilarity metric. However, the Euclidean distance may not always be a good dissimilarity measure for comparing data samples lying on a manifold.

In 2015, we have developed, in collaboration with Elif Vural (now Prof. at METU in Ankara, former postdoc in the team), two algorithms for determining a local subset of training samples from which a good local model can be computed for reconstructing a given input test sample, where we take into account the underlying geometry of the data. The first algorithm, called Adaptive Geometry-driven Nearest Neighbor search (AGNN), is an adaptive scheme which can be seen as an out-of-sample extension of the replicator graph clustering method for local model learning. The second method, called Geometry-driven Overlapping Clusters (GOC), is a less complex nonadaptive alternative for training subset selection. The AGNN and GOC methods have been evaluated in image super-resolution, deblurring and denoising applications and shown to outperform spectral clustering, soft clustering, and geodesic distance based subset selection in most settings. The selected patches are used for learning good local bases using the traditional PCA method. PCA is considered an efficient tool to recover the tangent space of the patch manifold when the manifold is sufficiently regular.

However, when the patch manifold has high curvature, which is observed to be the case for images with high frequencies, PCA may not be suitable. It can be seen in Figure 5 that the PCA basis with respect to a manifold fails to approximate the tangent space as the manifold bends over itself. In other words, PCA basis is not adapted when the curvature is too high. On the other hand, it can be seen in Figure 5 that a union of subspaces with respect to a manifold might generate a local model that yields a more efficient local representation of data.

In 2016, we have proposed a strategy to choose between these two kinds of bases locally depending on the local data geometry. This function is defined as the variability of the tangent space in each cluster.

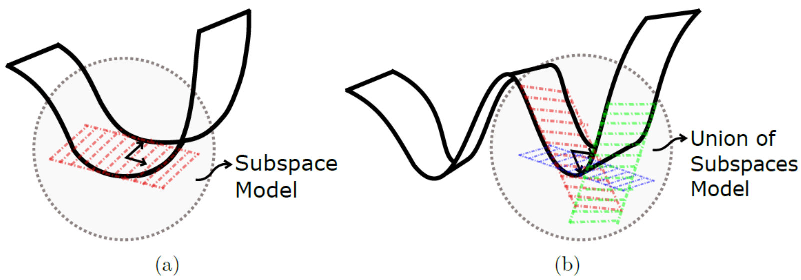


Figure 5. Subspaces computed with data sampled from a neighborhood on a manifold; (a): PCA basis which fails to approximate the subspace as the manifold curvature is too high; (b): union of subspaces generating a local model more coherent with the manifold geometry.

7.3.5. Rate-distortion optimized tone curves for HDR video compression

Participants: David Gommelet, Christine Guillemot, Aline Roumy.

High Dynamic Range (HDR) images contain more intensity levels than traditional image formats. Instead of 8 or 10 bit integers, floating point values requiring much higher precision are used to represent the pixel data. These data thus need specific compression algorithms. In collaboration with Ericsson [17], we have developed a novel compression algorithm that allows compatibility with the existing Low Dynamic Range (LDR) broadcast architecture in terms of display, compression algorithm and datarate, while delivering full HDR data to the users equipped with HDR display. The developed algorithm is thus a scalable video compression offering a base layer that corresponds to the LDR data and an enhancement layer, which together with the base layer corresponds to the HDR data. The novelty of the approach relies on the optimization of a mapping called Tone Mapping Operator (TMO) that maps efficiently the HDR data to the LDR data. The optimization has been carried out in a rate-distortion sense: the distortion of the HDR data is minimized under the constraint of minimum sum datarate (for the base and enhancement layer), while offering LDR data that are closed to some “aesthetic” a priori. Taking into account the aesthetic of the scene in video compression is indeed novel, since video compression is traditionally optimized to deliver the smallest distortion with the input data at the minimum datarate.

7.3.6. Cloud-based image compression

Participants: Jean Begaint, Christine Guillemot.

The emergence of cloud applications and web services has led to an increasing use of online resources. Image processing applications can benefit from this vast storage and distribution capacity. In collaboration with Technicolor, we investigate the use of this mass of redundant data to enhance image compression schemes. A region-based registration algorithm has been developped to capture complex deformations between two images. The registration method is then used to exploit both global and local correspondences between pairs of images of the same scene. The region-based registration yields a better prediction (hence reduced prediction errors, see Fig.6) which in turn yields a significant rate-distortion performance gain compared to current image coding solutions.

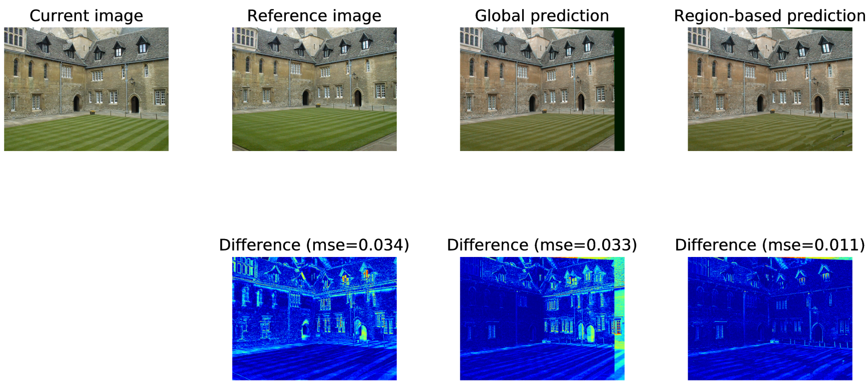


Figure 6. Image registration with global and region-based homographies and corresponding prediction error.

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. Interactive Coding for Navigation in 3D scenes (ICON 3D)

Participants: Thomas Maugey, Aline Roumy.

In order to have performing FTV systems, the data transmission has to take into account the interactivity of the user, *i.e.,* the viewpoint that is requested. In other words, a FTV system transmits to the visualisation support only what needs to be updated when a user changes its viewpoint angle (*i.e.*, the new information appearing in its vision field). The Sirocco has recently proposed some promising work using channel coding for interactive data coding. This coding scheme focusses on multi-view plus depth format only. In order to extend this approach to other formats, we have started a collaboration with the I3S laboratory in Nice, expert in 3D mesh compression.

The project ICON 3D funded by the GdR-Isis will be divided into two parts. First, we will study and develop new geometry prediction algorithms for surface meshes. Given a part of a mesh, the prediction algorithm should be able to estimate a neighboring mesh subset corresponding to the one newly visible after user viewpoint angle change (Fig. 7). The prediction error will be characterized. Then, we will study the channel coding method that should be developed to correct this error.

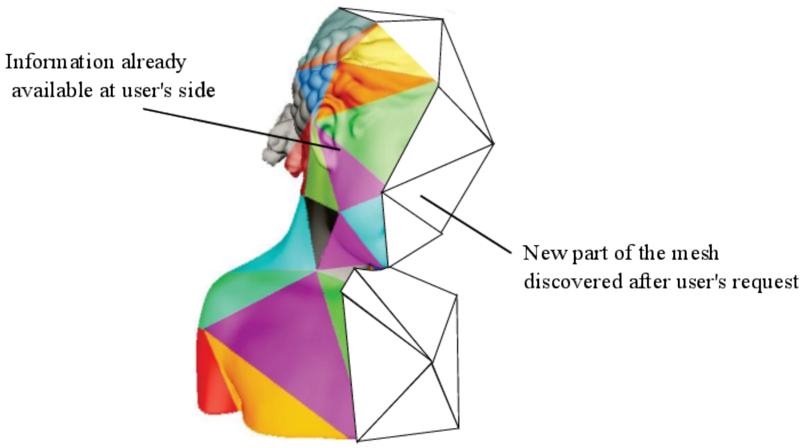


Figure 7. When a user changes his viewpoint angle, he discovers new part of the mesh that has to be transmitted.

STARS Project-Team

6. New Results

6.1. 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.

6.1.1. Perception for Activity Recognition

Participants: Piotr Bilinski, François Brémond, Etienne Corvée, Antitza Dancheva, Furqan Muhammad Khan, Michal Koperski, Thi Lan Anh Nguyen, Javier Ortiz, Remi Trichet, Jana Trojanova, Ujjwal Ujjval.

The new results for perception for activity recognition are:

- Exploring Depth Information for Head Detection with Depth Images (see 6.2)
- Modeling Spatial Layout of Features for Real World Scenario RGB-D Action Recognition (see 6.3)
- Multi-Object Tracking of Pedestrian Driven by Context (see 6.4)
- Pedestrian detection: Training set optimization (see 6.5)
- Pedestrian Detection on Crossroads (see 6.6)
- Automated Healthcare: Facial-expression-analysis for Alzheimer's patients in Musical Mnemotherapy (see 6.7)
- Hybrid Approaches for Gender estimation (see 6.8)
- Unsupervised Metric Learning for Multi-shot Person Re-identification (see 6.9)

6.1.2. Semantic Activity Recognition

Participants: François Brémond, Carlos Fernando Crispim Junior, Michal Koperski, Farhood Negin, Thanh Hung Nguyen, Philippe Robert.

For this research axis, the contributions are :

- Semi-supervised Understanding of Complex Activities in Large-scale Datasets (see 6.10)
- On the Study of the Visual Behavioral Roots of Alzheimer's disease (see 6.11)
- Uncertainty Modeling with Ontological Models and Probabilistic Logic Programming (see 6.12)
- A Hybrid Framework for Online Recognition of Activities of Daily Living In Real-World Settings (see 6.13)
- Praxis and Gesture Recognition (see 6.14)

6.1.3. Software Engineering for Activity Recognition

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Ines Sarray, Daniel Gaffé, Rachid Guerchouche, Matias Marin, Etienne Corvée, Julien Badie, Manikandan Bakthavatchalam, Vasanth Bathrinarayanan, Ghada Balhoul, Anais Ducoffe, Jean Yves Tigli, François Brémond.

The contributions for this research axis are:

- Scenario Recognition (see 6.15)
- The CLEM Workflow (see 6.16)
- Safe Composition in Middleware for Internet of Things (see 6.17)
- Verification of Temporal Properties of Neuronal Archetype (see 6.18)
- Dynamic Reconfiguration of Feature Models (see 6.19)
- Setup and management of SafEE devices (see 6.20)
- Brick & Mortar Cookies (see 6.21)

6.2. Exploring Depth Information for Head Detection with Depth Images

Participants: Thanh Hung Nguyen, Siyuan Chen.

Head detection may be more demanding than face recognition and pedestrian detection in the scenarios where a face turns away or body parts are occluded in the view of a sensor, but when locating people is needed. This year [29], we introduce an efficient head detection approach for single depth images at low computational expense. First, a novel head descriptor was developed and used to classify pixels as head or non-head. We used depth values to guide each window size, to eliminate false positives of head centers, and to cluster head pixels, which significantly reduce the computation costs of searching for appropriate parameters. High head detection performance was achieved in experiments with 90% accuracy for our dataset containing heads with different body postures, head poses, and distances to a Kinect2 sensor, and above 70% precision on a public dataset composed of a few daily activities, which is better than using a head-shoulder detector with HOG feature for depth images (see Figure 5)

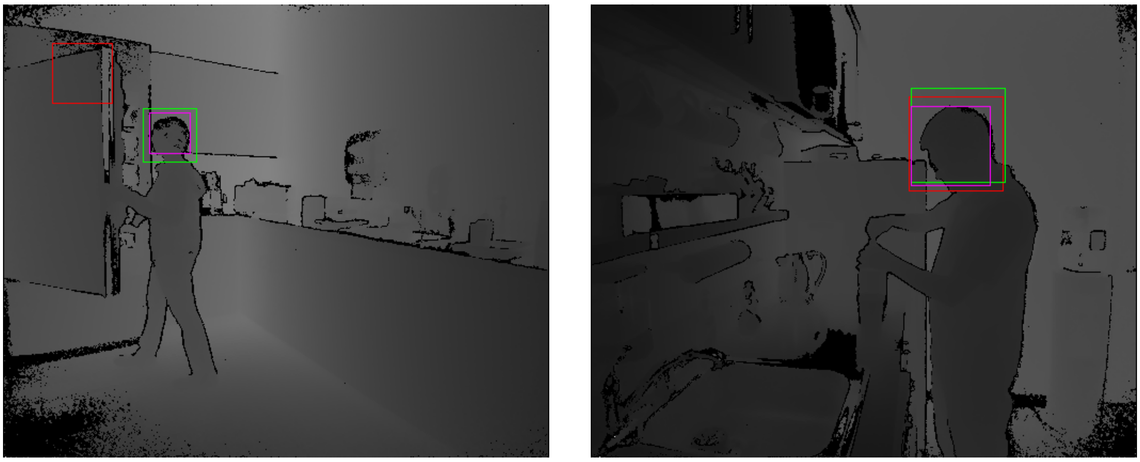


Figure 5. Examples of head detection where our algorithm successfully detects head. Pink square represents the ground truth, green rectangle represents our algorithm.

6.3. Modeling Spatial Layout of Features for Real World Scenario RGB-D Action Recognition

Participants: Michal Koperski, François Brémond.

keywords: computer vision, action recognition

Challenges in action representation in real-world scenario using RGB-D sensor

With RGB-D sensor it is easy to take advantage of real-time skeleton detection. Using skeleton information we can model not only dynamics of action, but also static features like pose. Skeleton-based methods have been proposed by many authors, and have reported superior accuracy on various daily activity data-sets. But the main drawback of skeleton-based methods is that they cannot make the decision when skeleton is missing.

We claim that in real world scenario of daily living monitoring, skeleton is very often not available or is very noisy. This makes skeleton based methods unpractical. There are several reasons why skeleton detection fails in real-world scenario. Firstly, the sensor has to work outside of it's working range. Since daily living monitoring is quite an unconstrained environment, the monitored person is very often too far from sensor, or is captured from non-optimal viewpoint angle. In Figure 6 we show two examples where skeleton detection fails. In the first example, the person on the picture wears black jeans which interferes with sensor. In such a case depth information from lower body parts is missing, making skeleton detection inaccurate. In the second example (see Figure 7) the person is too far from sensor and basically disappears in the background. In this case depth information is too noisy, thus skeleton detection fails. All disadvantages mentioned above will affect skeleton-based action recognition methods, because they strictly require skeleton detection.

On the other hand, local points-of-interest methods do not require skeleton detection, nor segmentation. That is why they received great amount of interest in RGB based action recognition where segmentation is much more difficult than with RGB-D. Those methods rely mostly on detection of points-of-interest usually based on some motion features (eg optical flow). The features are either based on trajectory of points-of-interest or descriptors are computed around the points-of-interest. One of the main disadvantage of those methods is fact that they fail when they cannot "harvest" enough points-of-interest. It happens when action has low dynamics eg "reading a book" or "writing on paper". Such actions contain very low amount of motion coming from hand when writing or turning the pages. In addition local points-of-interest methods very often ignore the spatial layout of detected features.

Proposed method

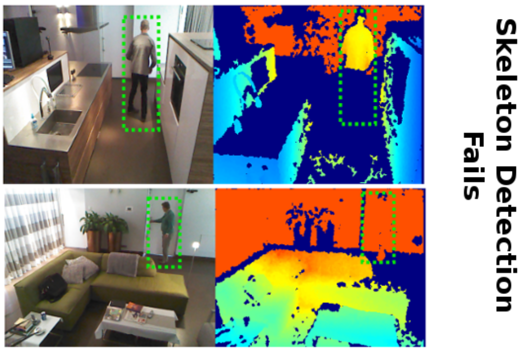


Figure 6. We show two examples where skeleton detection methods fail. Pictures on the left show RGB frame, pictures on the right show depth map (dark blue indicates missing depth information).

To address those problems we propose to replace skeleton detection by RGB-D based people detector. Note that person detection is much easier than skeleton detection. In addition we propose to use two people detectors: RGB and depth based - to take advantage of two information streams.

We propose to model spatial layout of motion features obtained from a local points-of-interest based method. We use Dense Trajectories [99] as a point of interest detector and MBH (Motion Boundary Histogram [62]) as a descriptor. To improve the discriminating power of MBH descriptor we propose to model spatial-layout of visual words computed based on MBH (Figure 7). We divide a person bounding box into Spatial Grid (SG) and we compute Fisher Vector representation in each cell. In addition, we show that other spatial-layout encoding methods also improve recognition accuracy. We propose 2 alternative spatial-layout encoding methods an we compare them with Spatial Grid.

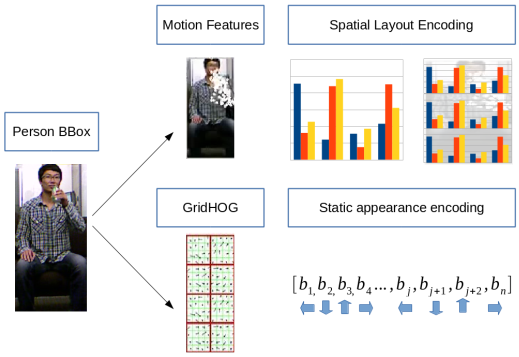


Figure 7. We show proposed method where we use people detection in place of skeleton. Next we propose to encode spatial-layout of visual words computed from motion features. In addition we propose GridHOG descriptor which encodes static appearance information.

To improve recognition of actions with low amount of motion we propose a descriptor which encodes rough static appearance (Figure 7). This can be interpreted as rough pose information. We divide the detected person bounding box into grid cells. Then we compute HOG [61] descriptor inside each cell to form the GHOG (GridHog) descriptor.

Further details can be find in the paper [37]. The contributions of this paper can be listed as follows:

- We propose to use two people detectors (RGB and depth based) to obtain person bounding box instead of skeleton.
- We propose to use Spatial Grid (SG) inside person bounding box. To model spatial-layout of MBH features.
- We propose to encode static information by using novel GHOG descriptor.
- We propose two other methods which model spatial-layout of MBH features and we compare them with Spatial Grid.

Experiments

We evaluate our approach on three daily activity data-sets: MSRDailyActivity3D, CAD-60 and CAD-120. The experiments show that we outperform most of the skeleton-based methods without requiring difficult in real-world scenario skeleton detection and thus being more robust (see Table1 , Table2 and Table3).

6.4. Multi-Object Tracking of Pedestrian Driven by Context

Participants: Thi Lan Anh Nguyen, François Brémond, Jana Trojanova.

Keywords: Tracklet fusion, Multi-object tracking

Multi-object tracking (MOT) is essential to many applications in computer vision. As so many trackers have been proposed in the past, one would expect the tracking task as solved. It is true for scenarios containing solid background with a low number of objects and few interactions. However, scenarios with appearance changes due to pose variation, abrupt motion changes, and occlusion still represent a big challenge.

Table 1. Recognition Accuracy Comparison for MSRDailyActivity3D data-set. corresponds to methods which require skeleton detection.

Method	Accuracy [%]
NBNN [94]	*70.00*
HON4D [87]	*80.00*
STIP + skeleton [106]	*80.00*
SSFF [95]	*81.90*
DSCF [102]	*83.60*
Actionlet Ensemble [101]	*85.80*
RGGP + fusion [79]	*85.60*
Super Normal [80]	*86.26*
BHIM [74]	86.88
DCSF + joint [102]	*88.20*
Our Approach	**85.95**

Table 2. Recognition Accuracy Comparison for CAD-60 data-set. corresponds to methods which require skeleton detection.

Method	Accuracy [%]
STIP [106]	62.50
Order Sparse Coding [86]	*65.30*
Object Affordance [75]	*71.40*
HON4D [87]	*72.70*
Actionlet Ensemble [101]	*74.70*
JOULE-SVM [72]	*84.10*
Our Approach	**80.36**

Table 3. Recognition Accuracy Comparison for CAD-120 data-set. corresponds to methods which require skeleton detection.

Method	Accuracy [%]
Salient Proto-Objects [92]	78.20
Object Affordance [75]	*84.70*
STS [76]	*93.50*
Our Approach	**85.48**

In the state of the art, some sets of efficient methods are proposed to face this challenge: data association (local and global) and tracking parameter adaptation. A very popular method for local data association is the bipartite matching. The exact solution can be found via Hungarian algorithm [85]. These methods are computationally inexpensive, but can deal only with short term occlusion. An example of global method is the extension of the bipartite matching into network flow [104]. Given the objects detections at each frame, the direct acyclic graph is formed and the solution is found through minimum-cost flow algorithm. The algorithms reduce trajectory fragments and improve trajectory consistency but lack robustness to identity switches of close or intersecting trajectories.

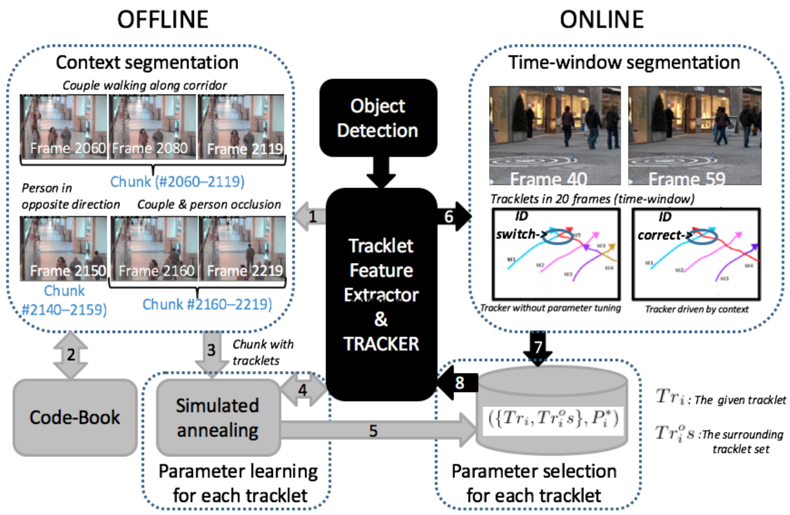


Figure 8. Our proposed framework.

Another set of methods for MOT is online parameter adaptation [56]. They tune automatically the tracking parameters based on the context information, while methods mentioned above use one appearance and/or one motion feature for the whole video. In [56], the authors learn the parameters for the scene context offline. In online phase the tracking parameters are selected from database based on the current context of the scene. These parameters are applied to all objects in the scene. Such a concept assumes discriminative appearance and trajectories among individuals, which is not always the case in real scenarios.

In order to overcome these limitations, we propose a new long term tracking framework. This framework has several dominant contributions:

- We introduce new long term tracking framework which combines short data association and the online parameter tuning for individual tracklets. In contrast to previous methods that used the same setting for all tracklets.
- We show that large number of parameters can be efficiently tuned via multiple simulated annealing, whereas previous method could tune only a limited number of parameters and fix the rest to be able to do exhaustive search.
- We define the surrounding context around each tracklet and similarity metric among tracklets allowing us to match learned context with unseen video set.

The proposed framework was trained on 9 public video sequences and tested on 3 unseen sets. It outperforms the state-of-art pedestrian trackers in scenarios of motion changes, appearance changes and occlusion of objects as shown in Table 4 . The paper is accepted in conference AVSS-2016 [39].

Table 4. Tracking performance. The best values are printed in red.

Dataset	Method	MOTA	MOTP	GT	MT	PT	ML
PETS2009	Shitrit et al. [52]	0.81	0.58	21	–	–	–
	***Bae et al.**-global association* [50]	0.73	0.69	23	*100*	0	*0.0*
	Chau et al. [57]	0.62	0.63	21	–	–	–
	Chau [58]([57] + parameter tuning for whole video context)	0.85	0.71	21	–	–	–
	Ours ([57] + Proposed approach)	*0.86*	*0.73*	21	76.2	14.3	9.5
TUD-Stadtmitte	Andriyenko et al. [47]	0.62	0.63	9	60.0	20.0	10.0
	Milan et al. [81]	*0.71*	*0.65*	9	*70.0*	20.0	*0.0*
	Chau et al. [57]	0.45	0.62	10	60.0	40.0	*0.0*
	Chau [58]([57] + parameter tuning for whole video context)	–	–	10	*70.0*	10.0	20.0
	Ours ([57] + Proposed approach)	0.47	*0.65*	10	*70.0*	30.0	*0.0*
TUD-Crossing	Tang et al. [96]	–	–	11	*53.8*	38.4	7.8
	Chau et al. [57]	0.69	0.65	11	46.2	53.8	*0.0*
	Ours ([57] + Proposed approach)	*0.72*	*0.67*	11	*53.8*	46.2	*0.0*

6.5. Pedestrian detection: Training set optimization

Participants: Remi Trichet, Javier Ortiz.

keywords: computer vision, pedestrian detection, classifier training, data selection, data generation, data weighting

The emphasis of our work is on data selection. Training for pedestrian detection is, indeed, a peculiar task. It aims to differentiate a few positive samples with relatively low intra-class variation and a swarm of negative samples picturing everything else present in the dataset. Consequently, the training set lacks discrimination and is highly imbalanced. Due to the possible creation of noisy data while oversampling, and the likely loss of information when undersampling, balancing positive and negative instances is a rarely addressed issue in the literature.

Bearing these data selection principles in mind, we introduce a new training methodology, grounded on a two-component contribution. First, it harnesses an expectation-maximization scheme to weight important training data for classification. Second, it improves the cascade-of-rejectors [105][54] classification by enforcing balanced train and validation sets every step of the way, and optimizing separately for recall and precision. A new data generation technique was developped for this purpose.

The training procedure unfolds as follows. After the initial data selection, we balance the negative and positive sample cardinalities. Then, a set of n negative data rejectors are trained and identified negative data are discarded. The validation set negative samples are iteratively oversampled after each training to ensure a balanced set. The final classifier is learned after careful data selection. Figure 9 illustrates the process.

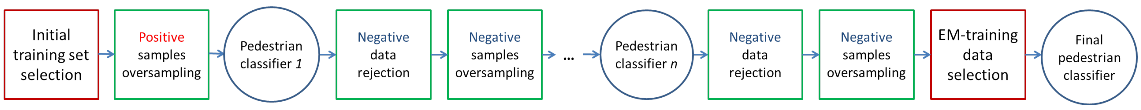


Figure 9. Training pipeline.

Experiments carried out on the Inria [61] and PETS2009 [69] datasets, demonstrate the effectiveness of the approach, leading to a simple HoG-based detector to outperform most of its near real-time competitors.

Table 5. Comparison with the state-of-the-art on the Inria dataset. Our approach is in italic. Computation time are calculated according to 640×480 resolution frames. The used metric is the log-average miss rate (the lower the better).

Method	Inria	Speed
HoG [61]	46%	21fps
DPM-v1 [68]	44%	< 1fps
HoG-LBP [98]	39%	Not provided
MultiFeatures [100]	36%	< 1fps
FeatSynth [51]	31%	< 1fps
MultiFeatures+CSS [97]	25%	No
FairTrain - HoG + Luv	*25%*	*11fps*
FairTrain - HoG	*25%*	*16fps*
Channel Features [65]	21%	0.5fps
FPDW [64]	21%	2-5fps
DPM-v2 [67]	20%	< 1fps
VeryFast [53]	18%	8fps(CPU)
VeryFast [53]	18%	135fps(GPU)
WordChannels [60]	17%	8fps(GPU)

Table 6. Comparison with the state-of-the-art on the PETS2009 S2.L1 sequence. Our approach is in italic. The used metric is the MODA (the higher the better).

Method	PETS2009	Speed
Arsic [48]	44%	n.a.
Alahi [46]	73%	n.a.
Conte [59]	85%	n.a.
FairTrain - HoG	*85.38%*	*29fps*
FairTrain - HoG + Luv	*85.49%*	*18fps*
Breitenstein [55]	89%	n.a.
Yang [103]	96%	n.a.

6.6. Pedestrian Detection on Crossroads

Participants: Ujwal Ujwal, François Brémond.

Pedestrian detection has a specific relevance in the space of object detection problems in computer vision. Due to increasing role of automated surveillance systems in increasing areas, demands for a highly robust and accurate pedestrian detection system is increasing day after day. Recently, deep learning has emerged as an important paradigm to tackle complex object detection problems. This year, we performed our initial studies on pedestrian detection using deep learning techniques. These studies form an important basis for us to extend our work in the future.

Evaluation Metrics

The relative comparison of different pedestrian detection systems was done using evaluation metrics. In the area of pedestrian detection, the most widely used evaluation metric is that of *miss rate*(MR). *Miss rate* is related to the concept of *recall*, which is another very commonly used metric in computer vision, especially in problems related to retrieval of images and concepts. *Miss Rate* is defined as follows:

$$Miss\ Rate = \frac{False\ Negatives}{True\ Positives + False\ Negatives} \tag{1}$$

		Pedestrian Detector	
		Pedestrian	Other
Ground Truth	Pedestrian	True Positive (TP)	False Negative (FN)
	Other	False Positive (FP)	True Negative (TN)

Figure 10. True and False Positives in pedestrian detection

In equation 1 , *True Positives*(TP) and *False Negatives*(FN) can be understood from figure 10 . A good pedestrian detector should not miss many people in a scene and this aspect is reflected in the definition of equation 1 . A good pedestrian detector is required to detect as few *False Positives*(FP) as possible. This is expressed in the literature usually in the form of False Positives Per Image(FPPI). FPPI is basically a per-image average of total number of FP detections.

Pedestrian detection systems usually work with a number of parameters. Different values of these parameters may tune a system to different MR and FPPI value. This is usually expressed in the form of a *Precision-recall*(PR) curve. This curve is created by varying a control parameter of a system and plotting MR and FPPI values. In literature it is customary to report MR value at 0.1 FPPI.

Experiments

We considered deep learning based models for our initial set of experiments. This is primarily owing to their popularity and the promise which they have demonstrated in the area of object detection over the past several years.

There are many deep learning based models which have been used for object detection. The purpose of these experiments was to gain a deeper insight into the performance of deep neural networks for pedestrian detection. We experimented with Faster-RCNN [88] and SSD detector [78]. These were chosen owing to the fact that they are recent models (2015 for Faster-RCNN and 2016 for SSD Detector), and have displayed state-of-art performance in terms of detection speeds and accuracy across many object categories.

The results shown in table7 were obtained by fine-tuning VGG-16 with imagenet and MS-COCO datasets which did not involve any public dataset specific to pedestrian detection. Hence, we took the fine-tuned model and further fine-tuned it with different pedestrian datasets to study the effectiveness of fine-tuning with pedestrian-specific datasets.

Each row in the first column of table8 , reflects the dataset(s) which were used to fine-tune the model. For each row, the model was fine-tuned using the dataset indicated in its first column, as well as the datasets indicated in the first column of all rows preceding it. The model was then evaluated against the test set of each dataset and the miss-rates are indicated in the table.

Table 7. Performance of fine-tuned Faster RCNN on pedestrian detection datasets.Numbers indicate the miss-rate.

Performance of fine-tuned Faster RCNN		
Dataset	**Faster RCNN Performance**	**State of Art**
Inria	13.47%	13%
Daimler	37.7%	29%
ETH-Zurich	32.1%	
Caltech	26.7%	19%
TUD-Brussels	52.2%	45%

-

Table 8. Faster-RCNN performance after fine-tuning with pedestrian datasets. Numbers indicate the miss-rate.

	Image datasets				
Trained Model	**Inria**	**Daimler**	**TUD-Brussels**	**ETH-Zurich**	**Caltech**
+Inria	13.4%	36.9%	52%	32.1%	28.2%
+Daimler	13.6%	33.7%	51.1%	32.7%	29.1%
+ETH-Zurich	13.8%	34.6%	49.3%	32%	26%
+Caltech	16%	35.4%	48%	33.2%	25.2%

While the initial results as seen from table 7 are encouraging, they still need a lot of improvement especially with complex datasets such as TUD-Brussels and Caltech. We also see from table8 , that fine-tuning with pedestrian datasets tends to improve the performance but the magnitude of improvement varies depending upon the dataset(s) being fine-tuned with and the dataset(s) being tested upon. These observations indicate some important research directions. Data in computer vision applications are highly varied and it is not very easy to capture its complexity and variations with sufficient ease. It is important to proceed to work on better

dataset usage by clustering the datasets together based on traits such as viewpoint, resolution etc. Resolution is another important element which significantly affects deep learning based approaches. This is because deep learning involves automated feature extractions from the pixel level and low resolution appearance often makes that problem difficult.
We intend to work upon and cover these issues in subsequent efforts towards solving the pedestrian detection problem.

6.7. Automated Healthcare: Facial-expression-analysis for Alzheimer's patients in Musical Mnemotherapy

Participants: Antitza Dantcheva, Piotr Bilinski, Philippe Robert, François Brémond.

keywords: automated healthcare, healthcare monitoring, expression recognition

The elderly population has been growing dramatically and future predictions and estimations showcase that by 2050 the number of people over 65 years old will increase by 70%, the number of people over 80 years old will increase by 170%, outnumbering younger generations from 0-14 years. Other studies indicate that around half of the current population of over 75 year old suffer from physical and / or mental impairments and as a result are in need of high level of care. The loss of autonomy can be delayed by maintaining an active life style, which also would lead to reduced healthcare financial costs. With the expected increase of the world elderly population, and on the other hand limited available human resources for care a question arises as "How can we improve health care in an efficient and cost effective manner?".

Motivated by the above, we propose an approach for detecting facial expressions in Alzheimer's disease (AD) patients that can be a pertinent unit in an automated assisted living system for elderly subjects. Specifically, we have collected video-data of AD patients in musical therapy at the AD center Fondation G.S.F J. L. Noisiez in Biot, France from multiple therapy-sessions for validating our method. We note that in such sessions even AD patients suffering from apathy exhibit a number of emotions and expressions. We propose a spatio-temporal algorithm for facial expression recognition based on dense trajectories, Fisher Vectors and support vector machine classification. We compared the proposed algorithm to a facial-landmark-based algorithm concerning signal displacement of tracked points within the face.

Our algorithm differentiates between four different facial expressions: (i) neutral, (ii) smile, (iii) talking, and (iv) singing with an accuracy of 56%, outperforming the facial-landmark-based algorithm. Challenging for both algorithms has been the unconstrained setting involving different poses, changes in illumination and camera movement. One expected benefit for AD patients is that positive expressions and their cause could be determined and replicated in order to increase life standard for such patients, which also brings to the fore a delay in the development of AD (see figure 11).

This work is published in the Gerontolgy Journal.

6.8. Hybrid Approaches for Gender Estimation

Participants: Antitza Dantcheva, Piotr Bilinski.

keywords: gender estimation, soft biometrics, biometrics, visual attributes

Automated gender estimation has numerous applications including video surveillance, human computer-interaction, anonymous customized advertisement, and image retrieval. Most commonly, the underlying algorithms analyze facial appearance for clues of gender.
Can a smile reveal your gender? [28], [35]

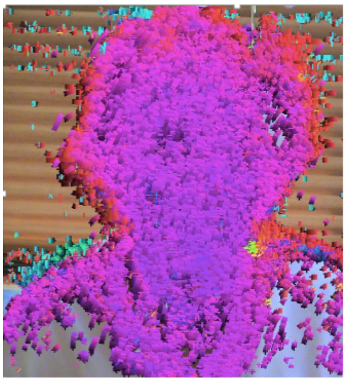


Figure 11. Expression recognition in AD patients based on dense trajectories and Fisher vectors. Dense trajectories visualization.

Deviating from such algorithms in [28] we proposed a novel method for gender estimation, exploiting dynamic features gleaned from smiles and show that (a) facial dynamics incorporate gender clues, and (b) that while for adults appearance features are more accurate than dynamic features, for subjects under 18, facial dynamics outperform appearance features. While it is known that sexual dimorphism concerning facial appearance is not pronounced in infants and teenagers, it is interesting to see that facial dynamics provide already related clues. The obtained results (see Table 9) show that smile-dynamic include pertinent and complementary to appearance gender information. Such an approach is instrumental in cases of (a) omitted appearance-information (*e.g.* low resolution due to poor acquisition), (b) gender spoofing (*e.g.* makeup-based face alteration), as well as can be utilized to (c) improve the performance of appearance-based algorithms, since it provides complementary information.

Table 9. True gender classification rates. Age given in years.

Age	< 20	> 19
Subj. amount	143	214
OpenBR	52.45%	78.04%
Dynamics (SVM+PCA) [28]	60.1%	69.2%
Dynamics (AdaBoost) [28]	59.4%	61.7%
OpenBR + Dynamics (Bagged Trees) [28]	60.8%	80.8%
Motion-based descriptors [35]	77.7%	80.11%
Improved dynamics [35]	$\mathbf{86.3}\%$	$\mathbf{91.01}\%$

We improve upon the above work by proposing a spatio-temporal features based on dense trajectories, represented by a set of descriptors encoded by Fisher Vectors [35]. Our results suggest that smile-based features include significant gender-clues. The designed algorithm obtains true gender classification rates of 86.3% for adolescents, significantly outperforming two state-of-the-art appearance-based algorithms (*OpenBR* and *how-old.net*), while for adults we obtain true gender classification rates of 91.01%, which is comparably discriminative to the better of these appearance-based algorithms (see Table 9).

Distance-based gender prediction: What works in different surveillance scenarios?

In this work [36] we studied gender estimation based on information deduced jointly from face and body, extracted from single-shot images. The approach addressed challenging settings such as low-resolution-images, as well as settings when faces were occluded. Specifically the face-based features included local binary patterns (LBP) and scale-invariant feature transform (SIFT) features, projected into a PCA space. The features of the novel body-based algorithm proposed in this work included continuous shape information extracted from body silhouettes and texture information retained by HOG descriptors. Support Vector Machines (SVMs) were used for classification for body and face features. We conduct experiments on images extracted from video-sequences of the Multi-Biometric Tunnel database, emphasizing on three distance-settings: close, medium and far, ranging from full body exposure (far setting) to head and shoulders exposure (close setting). The experiments suggested that while face-based gender estimation performs best in the close-distance-setting, body-based gender estimation performs best when a large part of the body is visible. Finally we presented two score-level-fusion schemes of face and body-based features, outperforming the two individual modalities in most cases (see Table10 and Table 11).

Table 10. Performance (%) of the Face Gender Estimation algorithm (FGE) and the Body Gender Estimation algorithm (BGE).

Distance	FGE			BGE		
Scenario	Male TPR	Fem. TPR	Acc.	Male TPR	Fem. TPR	Acc.
Far	94.28	20	57.14	87.14	88.57	87.85
Medium	71.42	90	80.71	85.71	87.14	86.42
Close	88.57	90	89.28	78.57	80	79.28

Table 11. Performance (%) of the Sum fusion and Smarter Sum Fusion of FGE and BGE in terms of True Positive Rate (TPR) for Male and Female (Fem.), overall Accuracy (Acc.). Best performance (in terms of Acc.) of each distance-setting is bolded.

Distance	Sum Fusion			Prop. Sum Fusion		
Scenario	Male TPR	Fem TPR	Acc.	Male TPR	Fem TPR	Acc.
Far	87.14	88.57	**87.85**	87.14	88.57	**87.85**
Medium	88.57	90	**89.28**	88.57	90	**89.28**
Close	87.14	88.57	87.85	92.85	94.28	**93.57**

6.9. Unsupervised Metric Learning for Multi-shot Person Re-identification

Participants: Furqan Khan, François Brémond.

keywords: re-identification, long term visual tracking, metric learning, unsupervised labeling

Automatic label generation for metric learning

Appearance based person re-identification is a challenging task, specially due to difficulty in capturing high intra-person appearance variance across cameras when inter-person similarity is also high. Metric learning is often used to address deficiency of low-level features by learning view specific re-identification models. The models are often acquired using a supervised algorithm. This is not practical for real-world surveillance systems because annotation effort is view dependent. Therefore, everytime a camera is replaced or added, a significant amount of data has to be annotated again. We propose a strategy to automatically generate labels for person tracks to learn similarity metric for multi-shot person re-identification task. Specifically, we use the fact that non-matching (negative) pairs far out-number matching (positive) pairs in any training set. Therefore, the true class conditional probability of distance given negative class can be estimated using the empirical marginal distribution of distance. This distribution can be used to sample non-matching person pairs for metric learning. A brief overview of the approach is presented below, please refer to [33] for details.

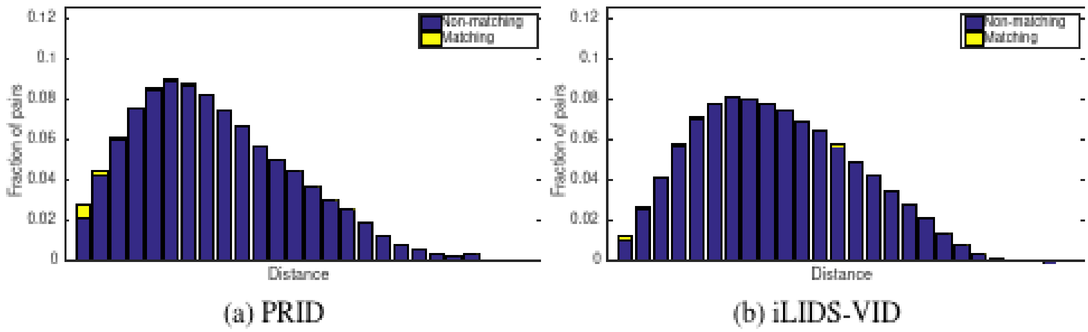


Figure 12. Distributions of distances between pairs of signature of randomly selected half of a) PRID, and b) iLIDS-VID datasets for MCM representation using Euclidean distance. The distributions are averaged for 10 trials.

In figure 12 , empirical distribution of Euclidean distance (using MCM [43] representation) is plotted for two publicly available datasets. It can be noted that the positive samples lie on one side of distribution mode. Therefore, negative pairs can be sampled according to the probability proportional to the signed distance from the mode. Practically, we only select sample pairs that are farthest away in the distribution as negative pairs. For positive pairs, we use the fact that each track has more than one image for a person. Thus we generate positive pairs using the persons selected for negative pairs. We evaluated our approach on three publicly available datasets in multi-shot settings: iLIDS-VID, PRID and iLIDS-AA. Performance comparison of different representations using recognition rates at rank r are detailed in table 12 , table 13 and table 14 . Our results validate the effectiveness of our approach by considerably reducing the performance gap between fully-supervised models using KISSME algorithm and Euclidean distance.

Table 12. PRID

Method	r=1	r=5	r=10	r=20
MCM+MPD	53.6	83.1	91.0	96.9
MCM+UnKISSME	59.2	81.7	90.6	96.1
MCM+KISSME	64.3	86.1	94.5	98.0

Table 13. iLIDA-VID

Method	r=1	r=5	r=10	r=20
MCM+MPD	34.3	61.5	74.4	83.3
MCM+UnKISSME	38.2	65.7	75.9	84.1
MCM+KISSME	40.3	69.9	79.0	87.5

6.10. Semi-supervised Understanding of Complex Activities in Large-scale Datasets

Participants: Carlos F. Crispim-Junior, Michal Koperski, Serhan Cosar, François Brémond.

keywords: Semi-supervised methods, activity understanding, probabilistic models, pairwise graphs

Table 14. iLIDS-AA

Method	r=1	r=5	r=10	r=20
MCM+MPD	56.5	79.7	90.9	95.2
MCM+UnKISSME	61.2	85.1	92.8	96.0
MCM+KISSME	62.9	84.7	93.4	97.0

Informations

Methods for action recognition have evolved considerably over the past years and can now automatically learn and recognize short term actions with satisfactory accuracy. Nonetheless, the recognition of complex activities - compositions of actions and scene objects - is still an open problem due to the complex temporal and composite structure of this category of events. Existing methods focus either on simple activities or oversimplify the modeling of complex activities by targeting only whole-part relations between its sub-parts (*e.g.*, actions). We study a semi-supervised approach (Fig. 13) that can learn complex activities from the temporal patterns of concept compositions in different arities (*e.g.*, "slicing-tomato" before "pouring_into-pan"). So far, our semi-supervised, probabilistic model using pairwise relations both in compositional and temporal axis outperforms prior work by 6 % (59% against 53%, mean Average precision, Fig. 14). Our method also stands out from the competition by its capability to handle relation learning in a setting with large number of video sequences (*e.g.*, 256) and distinct concept classes (Cooking Composite dataset, 218 classes, [90]), an ability that current state-of-the-art methods lack. Our initial achievements in this line of research has been published in [31]. Further work will focus on learning relations of higher arity.

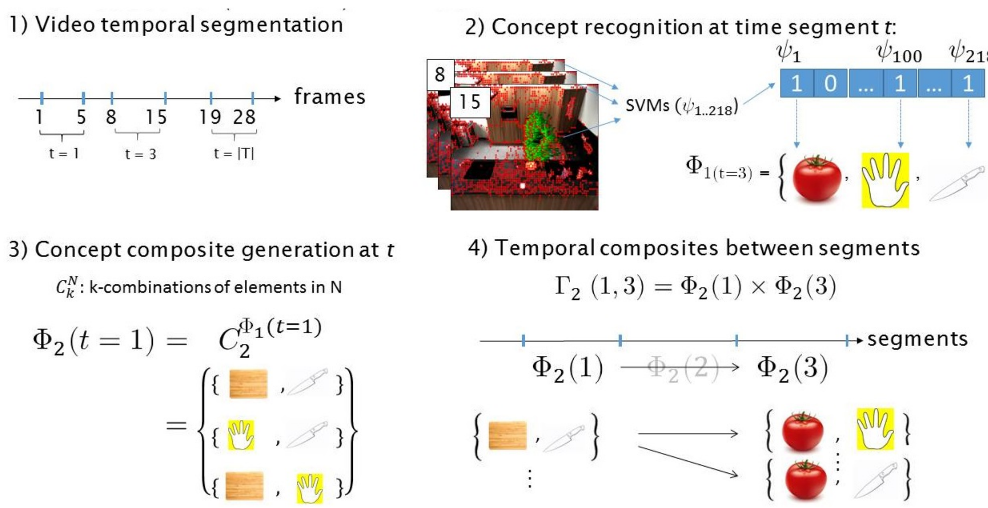


Figure 13. Semi-supervised learning of a video representation: 1) video temporal segmentation, 2) concept recognition 3) composite concept generation per time segment, 4) Temporal composite generation between segments.

6.11. On the Study of the Visual Behavioral Roots of Alzheimer's disease

Participants: Carlos F. Crispim-Junior, François Brémond.

Keywords: Activities of Daily Living, Dementia prediction, RGBD sensors, Activity Recognition, Cognitive Health

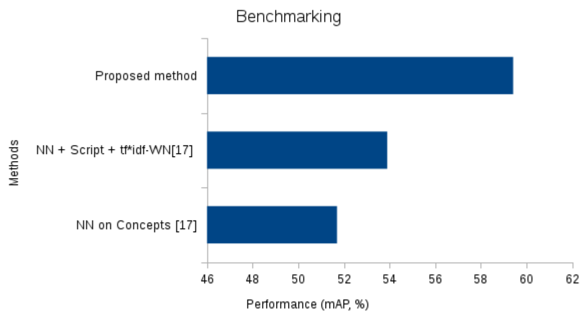


*Figure 14. Performance benchmarking of our approach against data set baselines: a) Nearest Neighbor classifier (NN) on concepts, script data, and tf*idf-WN, and b) NN only on concepts.*

Existing computer vision studies for the diagnosis of Dementia have focused on extracting discriminative patterns between healthy and people with dementia from neuroimagery exams, like functional MRI and PET scans. Nonetheless, the effects of dementia over human behaviors are a discriminative component that is barely explored by automatic vision-based methods. We studied a framework to automatically recognize the cognitive health of seniors from the visual observation of their activities of daily living (Fig.16). We employ a lightweight activity recognition system based on RGBD sensors to recognize the set of target activities (*e.g.*, prepare drink, prepare medication, make a payment transaction) performed by a person in a continuous video stream. Then, we summarize the absolute and relative activity patterns present in the video sequence using a novel probabilistic representation of activity patterns. Finally, this representation serves as input to Random Forest classifiers to predict the class of cognitive health that the person in question belongs to. We demonstrate that with the current framework can recognized the cognitive health status of seniors (*e.g.*, healthy, Mild Cognitive Impairment and Alzheimer's disease) with an average F_1-score of 69 % in real life scenarios.

6.12. Uncertainty Modeling with Ontological Models and Probabilistic Logic Programming

Participants: Carlos F. Crispim-Junior, François Brémond.

keywords: probabilistic logic programming, activities of daily living, senior monitoring, ontological models,

We have been investigating novel probabilistic, knowledge-driven formalisms that can join the representation expressiveness of an ontology-based language with the probabilistic reasoning of probabilistic graphical models, like probabilistic graphical models and probabilistic programming languages. The goal is to support the representation of events (entities, sub-events and constraints) and hierarchical structures (event, sub-events) and at the same time be capable of handling uncertainty related to both entity/sub-event detection and soft constraints. Prior work in probabilistic logic provides support to reasoning either about uncertainty related to entity recognition (probability of entity x in the scene defined in ProbLog2) or to soft-constraint (relevance of violation of constraint i to model y as defined in Markov Logic). In our current work in partnership with KU university of Leuven, we have extended the ontological models of our vision pipeline (Fig.17) with probabilistic logic formalism proposed by ProbLog (Fig.18), a probabilistic logic programming language. Current results on the recognition of daily activities of seniors are promising as they improved the precision of our prior method by 1%. Further work will focus on extending our uncertainty models to be robust to constraint violations.

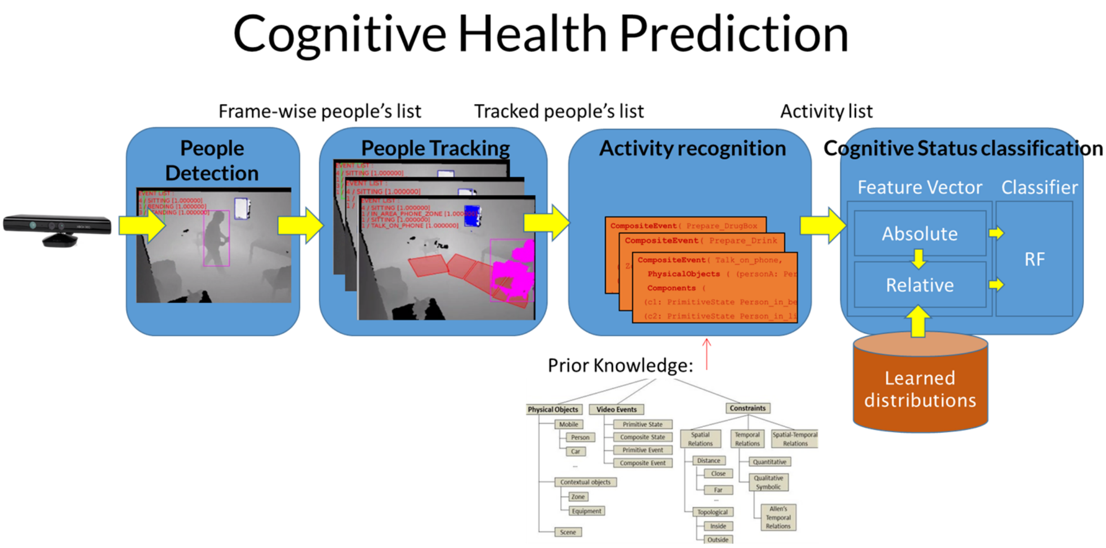


Figure 15. Automatic framework for visual recognition of cognitive health status: visual event recognition is responsible to detect and track people in the scene and recognize their events based on spatio-temporal relations with scene objects. Cognitive health classification represents absolute and relative information about the target classes.

Figure 16. Monitoring a senior performing at a gait-related event

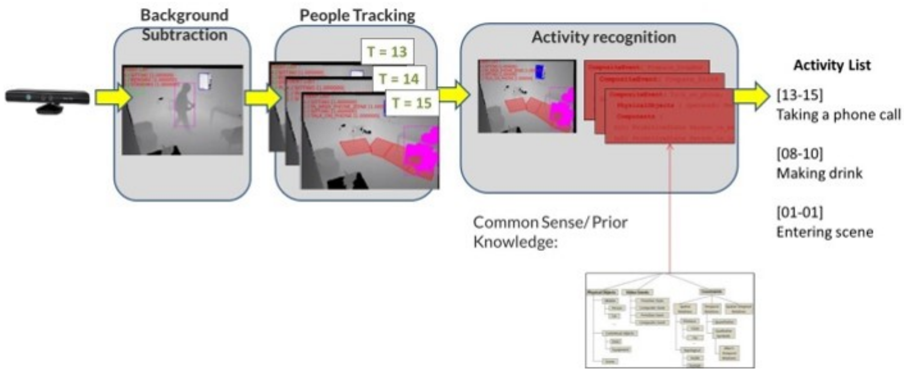


Figure 17. Pipeline for online activity recognition: given an acquisition camera (e.g. a Kinect), it firstly detects people using background subtraction algorithm, then it looks for appearance correspondence between people detected in the current frame with respect to past detections (past-present approach), and thirdly it recognizes the activities performed by each of the tracked people.

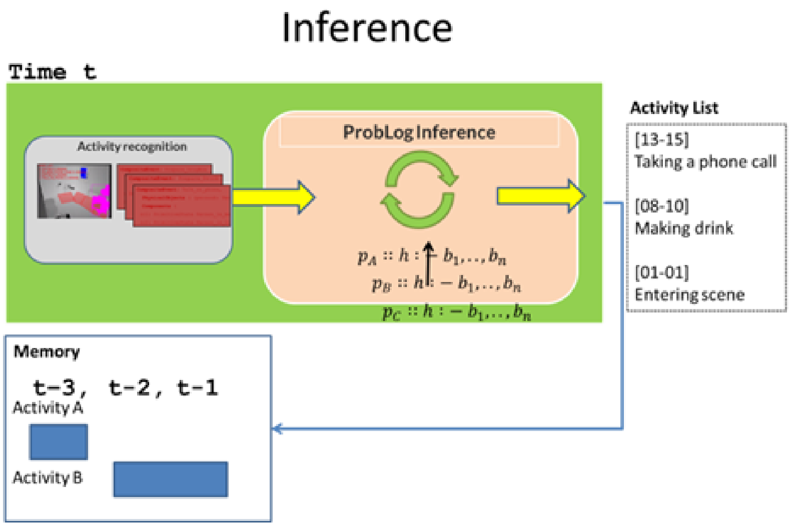


Figure 18. Temporal Inference using ProbLog engine. It takes as input deterministic observations and frame-wisely it recognizes the target events. Frame-events are aggregated into time intervals to create the time intervals of complex activities.

6.13. A Hybrid Framework for Online Recognition of Activities of Daily Living In Real-World Settings

Participants: Farhood Negin, Serhan Cosar, Michal Koperski, Carlos Crispim, Konstantinos Avgerinakis, François Brémond.

keywords: Supervised and Unsupervised Learning, Activity Recognition

State-of-the-art and Current Challenges

Recognizing human actions from videos has been an active research area for the last two decades. With many application areas, such as surveillance, smart environments and video games, human activity recognition is an important task involving computer vision and machine learning. Not only the problems related to image acquisition, e.g., camera view, lighting conditions, but also the complex structure of human activities makes activity recognition a very challenging problem. Traditionally, there are two variants of approach to cope with these challenges: supervised and unsupervised methods. Supervised approaches are suitable for recognizing short-term actions. For training, these approaches require a huge amount of user interaction to obtain well-clipped videos that only include a single action. However, Activities of Daily Living (ADL) consists of many simple actions which form a complex activity. Therefore, the representation in supervised approaches are insufficient to model these activities and a training set of clipped videos for ADL cannot cover all the variations. In addition, since these methods require manually clipped videos, they can only follow an offline recognition scheme. On the other hand, unsupervised approaches are strong in finding spatio-temporal patterns of motion. However, the global motion patterns are not enough to obtain a precise classification of ADL. For long-term activities, there are many unsupervised approaches that model global motion patterns and detect abnormal events by finding the trajectories that do not fit in the pattern [70], [83]. Many methods have been applied on traffic surveillance videos to learn the regular traffic dynamics (e.g. cars passing a cross road) and detect abnormal patterns (e.g. a pedestrian crossing the road) [71].

Proposed Method

We propose a hybrid method to exploit the benefits of both approaches. With limited user interaction our framework recognizes more precise activities compared to available approaches. We use the term precise to indicate that, unlike most of trajectory-based approaches which cannot distinguish between activities under same region, our approach can be more sensitive in the detection of activities thanks to local motion patterns. We can summarize the contributions of this work as following: i) online recognition of activities by automatic clipping of long-term videos and ii) obtaining a comprehensive representation of human activities with high discriminative power and localization capability.

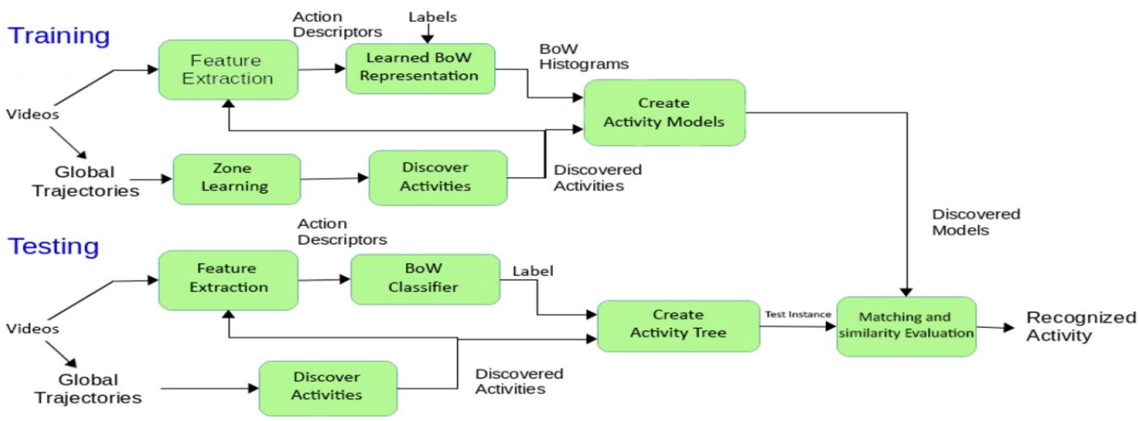


Figure 19. Architecture of the framework: Training and Testing phases

Figure 19 illustrates the flow of the training and testing phases in the proposed framework. For the training phase, the algorithm learns relevant zones in the scene and generates activity models for each zone by complementing the models with information such as duration distribution and BoW representations of discovered activities. At testing, the algorithm compares the test instances with the generated activity models and infers the most similar model.

The performance of the proposed approach has been tested on the public GAADRD dataset [73] and CHU dataset. Our approach always performs equally or better than online supervised approach in [99] (see Table15 and Table16). And even most of the time it outperforms totally supervised approach (manually clipped) of [99]. This reveals the effectiveness of our hybrid technique where combining information coming from both constituents could contribute to enhance recognition. The paper of this work was accepted in AVSS 2016 conference [30].

Table 15. The activity recognition results for CHU dataset. Bold values represent the best sensitivity and precision results for each class.

	Supervised (Manually Clipped) of [99]		Online Version of [99]		Unsupervised Using Global Motion [66]		Proposed Approach	
ADLs	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)
Answering Phone	57	78	**100**	**86**	**100**	60	**100**	81.82
P. Tea + W. Plant	89	**86.5**	76	38	84.21	80	**94.73**	81.81
Using Phar. Basket	**100**	83	**100**	43	90	**100**	**100**	**100**
Reading	35	**100**	92	36	81.82	**100**	**100**	91.67
Using Bus Map	90	**90**	**100**	50	**100**	54.54	**100**	83.34
AVERAGE	74.2	87.5	93.6	50.6	91.2	78.9	**98.94**	**87.72**

6.14. Praxis and Gesture Recognition

Participants: Farhood Negin, Jeremy Bourgeois, Emmanuelle Chapoulie, Philippe Robert, François Brémond.

keywords: Gesture Recognition, Dynamic and Static Gesture, Alzheimer Disease, Reaction Time, Motion Descriptors

Challenges and Proposed Method

Most of the developed societies are experiencing an aging trend of their population. Aging is correlated with cognitive impairment such as dementia and its most common type: Alzheimer's disease. So, there is an urgent need to develop technological tools to help doctors to do early and precise diagnoses of cognitive decline. Inability to correctly perform purposeful skilled movements with hands and other forelimbs most commonly is associated with Alzheimerâs disease [84]. These patients have difficulty to correctly imitate hand gestures and mime tool use, e.g. pretend to brush one's hair. They make spatial and temporal errors. We propose a gesture recognition and evaluation framework as a complementary tool to help doctors to spot symptoms of cognitive impairment at its early stages. It is also useful to assess one's cognitive status. First, the algorithm classifies the defined gestures in the gestures set and then it evaluates gestures of the same category to see how well they perform compared to correct gesture templates. Methods Shape and motion descriptors such as HOG (histogram of oriented gradient) [61] and HOF (histogram of optical flow) [62] are an efficient clue to characterize different gestures (Figure 20 Left). Extracted descriptors are utilized as input to train the

Table 16. The activity recognition results for GAADRD dataset. Bold values represent the best sensitivity and precision results for each class.

ADLs	Supervised (Manually Clipped) Approach [99]		Online Version of [99]		Classification by detection using SSBD [49]		Unsupervised Using Global Motion [66]		Proposed Approach	
	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)
Answering Phone	**100**	88	**100**	70	96	34.29	**100**	**100**	**100**	88
Establish Acc. Bal.	67	**100**	**100**	29	41.67	41.67	**100**	86	67	**100**
Preparing Drink	**100**	69	**100**	69	96	80	78	**100**	**100**	82
Prepare Drug Box	58.33	**100**	11	20	**86.96**	51.28	33.34	**100**	22.0	**100**
Watering Plant	54.54	**100**	0	0	**86.36**	86.36	44.45	57	44.45	**80**
Reading	**100**	**100**	88	37	**100**	31.88	**100**	**100**	**100**	**100**
Turn On Radio	60	86	**100**	75	96.55	19.86	89	89	89	89
AVERAGE	77.12	**91.85**	71.29	42.86	**86.22**	49.33	77.71	90.29	74.57	91.29

classifiers. We use bag-of-visual-words approach to characterize gestures with descriptors. The classification happens in two steps: first we train a classifier to distinguish different gestures and after, we train another classifier with correct and incorrect samples of the same class. This way, we could recognize which gesture is performed and whether it is performed accurately or not.

Experiments and Results

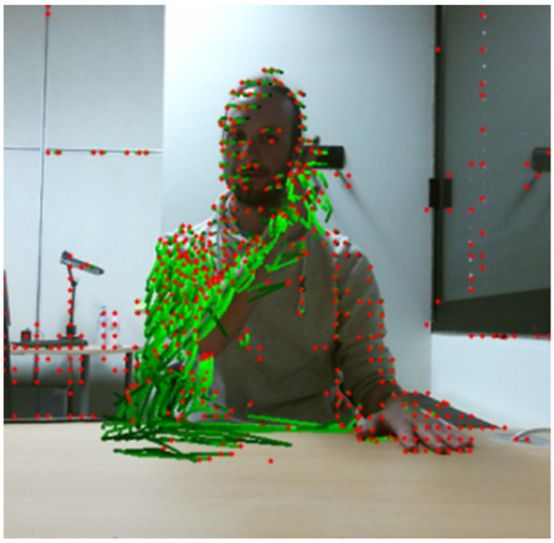
The framework is fed by input data which come from a depth sensor (Kinect, Microsoft). At first phase, the correct samples of gestures performed by clinicians, are recorded. We train the framework using correct instances of each gesture class. In the second phase, participants were asked to perform the gestures. We use virtual reality as modality to interact with subjects to make the experiments more immersive and realistic experience. First an avatar performs a specific gesture and then she asks the subject to repeat the same gesture (Figure 20 Right). In this work, we analyze two categories of gestures. First category is dynamic gestures where the whole motion of the hands is considered as a complete gesture. Second category of gestures is static gestures where only a static pose of hands is the desired gesture. For static gestures, we also need to detect this key frame. Moreover, reaction time which starts after avatar asked the subject to do the gesture, until subject really starts to perform the gesture, could be an important diagnostic factor. Our algorithm uses motion descriptors to detect key frames and reaction time. In the preliminary tests, our framework successfully recognized more than 80% of the dynamic gestures. It also detects key frames and reaction time with a high precision. Thus the proposed gesture recognition framework helps doctors by providing a complete assessment of gestures performed by subject.

This work is published in [30] and will appear in the Gerontechnology Journal.

6.15. Scenario Recognition

Participants: Inès Sarray, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault.

Keywords: Synchronous Modeling, Model checking, Mealy machine, Cognitive systems.



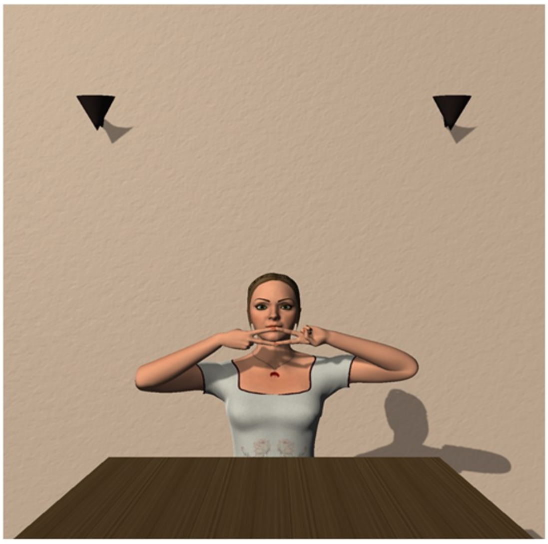


Figure 20. Left: Extracted motion descriptors while performing a gesture Right: virtual avatar guides patients in a virtual reality environment

Activity recognition systems aim at recognizing the intentions and activities of one or more persons in real life, by analyzing their actions and the evolution of the environment. This is done thanks to a pattern matching and clustering algorithms, combined with adequate knowledge representation (e.g scene topology, temporal constraints) at different abstraction levels (from raw signal to semantics).Stars has been working to ameliorate and facilitate the generation of these activity recognition systems. As we can use these systems in a big range of important fields, we propose a generic approach to design activity recognition engine. These engines should continuously and repeatedly interact with their environment and react to its stimuli. On the other hand, we should take into consideration the dependability of these engines which is very important to avoid possible safety issue, that's why we need also to rely on formal methods that allow us to verify these engines behavior. Synchronous modeling is a solution that allows us to create formal models that describe clearly the system behavior and its reactions when it detects different stimuli. Using these formal models, we can build effective recognition engines for each formal model and validate them easily using model checking. This year, we adapted this approach to create a new simple scenario language to express the scenario behaviors and to automatically generate its recognition automata at compile time. This automata will be embedded into the recognition engine at runtime.

Scenario description Language

As we work with non-computer-science end-users, we need a friendly description language that helps them to express easily their scenarios. To this aim, we collaborated with Ludotic ergonomists to define the easiest way for a simple user to deal with the new language. Using AxureRP tool, we defined two types of language:

1- Textual language:

For the textual language, we decided to use a simple language. Using 9 operators, and after the definition of the types, roles, and sub-scenarios, the user can describe a scenario in a simple way, such as in figure21 .

This year, we implemented this textual language and it is under testing.

2)- Graphical language:

```
Type Personne, Equipement,Zone;

Scenario coupTel :

role
        Patient:Personne;
        Tel: Equipement;
        table: Equipement;
        sejour: Zone;

Subscenarios
        entend(Personne, Equipement);
        decroche(Personne, Equipement);
        commence_a_parler(Personne);
        finit_de_parler(Personne);
        raccroche(Personne,Equipement);

EtatInitial : dans_Zone(Patient, sejour);

debut

   pres_de(Patient, table) parallele entend(Patient, Tel)
puis
   decroche(Patient, Tel)
puis
   commence_a_parler(Patient)
puis
   finit_de_parler(Patient)
puis
   raccroche(Patient, Tel)
puis
      Alert (fin_de_scenario)

fin
```

Figure 21. Example of the textual language

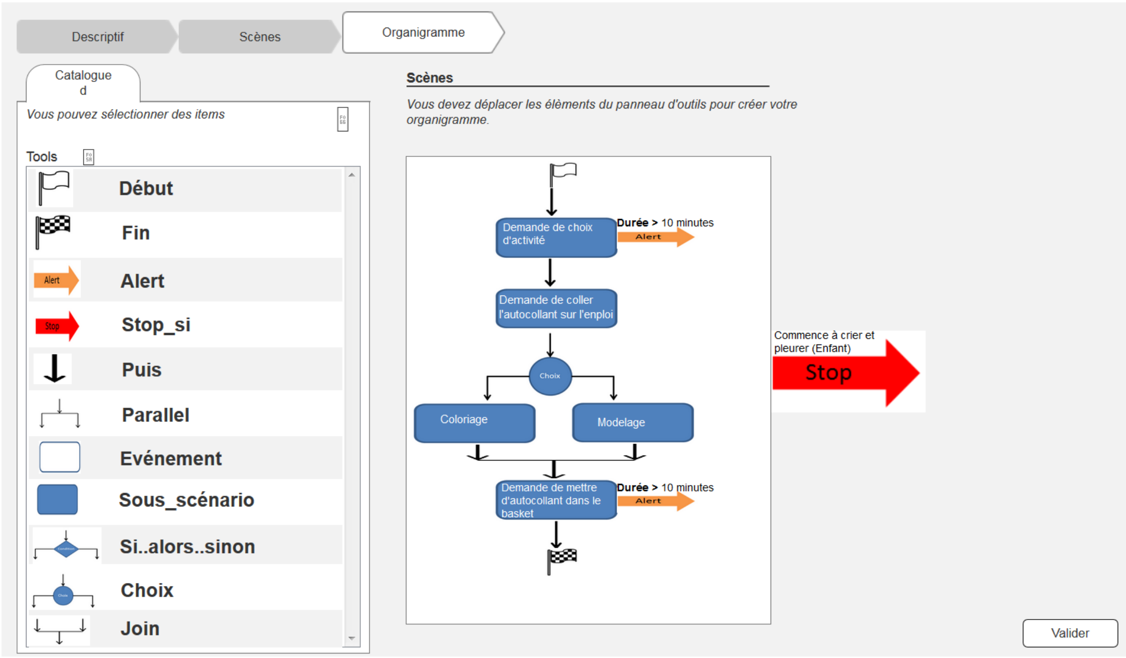


Figure 22. Generic flowchart

The graphical language model has 3 basic interfaces: The first interface allows the user to define the types, roles, and the initial state of the scenario. The second one is dedicated to describe the sub-scenarios and to express simple scenarios using a timeline. In case of complicated scenarios, the third interface offers users a tool panel that allows them to describe their scenarios in a hierarchical way using a flowchart-like representation (see figure 22).

Recognition Automata

This year, we worked also on recognition automata generation. We used the synchronous modeling and semantics to define these engines. The semantics consists in a set of formal rules that describe the behavior of a program. We specified first the language operators: we rely on a 4-valued algebra with a bilattice structure to define two semantics for the recognition engine: a behavioral and equational one. A behavioral semantics defines the behavior of a program and its operators and gives it a clear interpretation. Equational semantics allows us to make a modular compilation of our programs using rules that translate each program into an equation system. After defining these two semantics, we verified their equivalence for all operators, by proving that these semantics agree on both the set of emitted signals and the termination value for a program P. We implemented these semantics and we are now working on the automatic generation of the recognition automata.

6.16. The Clem Workflow

Participants: Annie Ressouche, Daniel Gaffé.

Keywords: Synchronous languages, Synchronous Modeling, Model checking, Mealy machine.

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit around the language (see Figure 23) to design, simulate, verify, and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality.

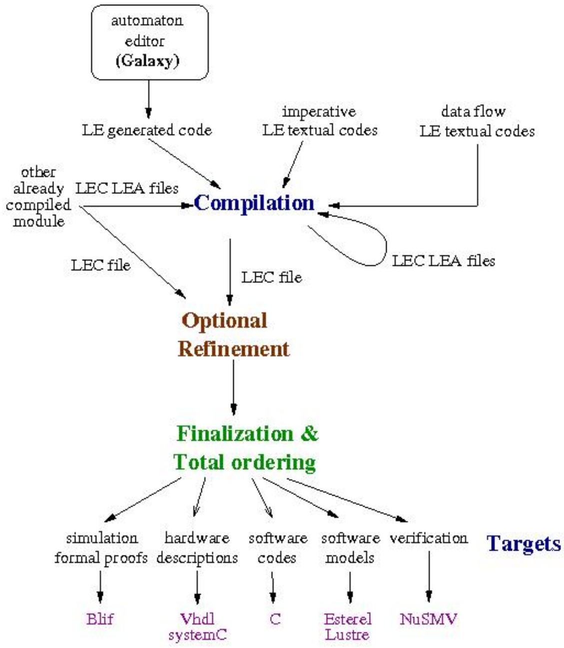


Figure 23. The Clem Toolkit

This year, we continued to focus on the improvement of both LE language and compiler concerning data handling and the generation of back-ends, required by other research axis of the team. We also designed a large application: a mechatronics system in CLEM and we have proved that its main safety properties hold in our modeling. Now, to complete the improvement done these two last years concerning data handling, we want to extend the verification side of CLEM. To this aim, this year we began to replace the fundamental representation of Boolean values as BDD (Binary Decision Diagrams) with LDD (Logical Decision Diagrams), which allow to encode integer values in a very efficient way. It turns out that the validation mechanism of CLEM could take into account properties over integer data. However, this is a first test and the integration of a model checking technique in CLEM remains a challenge.

6.17. Safe Composition in Middleware for Internet of Things

Participants: Annie Ressouche, Daniel Gaffé, Jean-Yves Tigli.

Keywords: Synchronous Modeling, Ubiquitous Computing, middleware, internet of things

The main concern of this research axis is the dependability of a component-based adaptive middleware which dynamically adapt and recompose assemblies of web components. Such a middleware plays an important role in the generation of event recognition engines we are currently building in Stars team (see section 6.15). One of the main challenge is how to guarantee and validate some safety and integrity properties throughout the system's evolution. These two last years, we have proposed to rely on synchronous models to represent component behavior and their composition and to verify that these compositions verify some constraints during the dynamic adaptation to appearance and disappearance of components. We defined a generic way to express these constraints and we proposed the Description Constraint Language (DCL) to express these constraints. Hence, we compile them into LE programs (see 6.16) and we benefit from CLEM model checking facilities to ensure that they are respected [93]. This year, we improved the DCL language in order to take into account both the dynamic variation of components and also applications which use these components and we are currently testing the efficiency of our method to add and remove components. Moreover, genericity is expressed by the notion of type and we aim at extending this notion to a thinner representation of knowledge about components.

6.18. Verification of Temporal Properties of Neuronal Archetypes

Participants: Annie Ressouche, Daniel Gaffé.

Keywords: Synchronous Modeling, model-checking, lustre, temporal logic, biologic archetypes

This year, we began a collaboration with with the I3S CNRS laboratory and Jean Dieudonné CNRS laboratory to verify temporal properties of neuronal archetypes. There exist many ways to connect two, three or more neurons together to form different graphs. We call archetypes only the graphs whose properties can be associated to specific classes of biologically relevant structures and behaviors. These archetypes are supposed to be the basis of typical instances of neuronal information processing. To model different representative archetypes and express their temporal properties, we use a synchronous programming language dedicated to reactive systems (Lustre). Then, we generate several back ends to interface different model checkers supporting data types and automatically validate these properties. We compare the respective results. They mainly depend on the underlying abstraction methods used in model checkers.

These results are published in [32]

6.19. Dynamic Reconfiguration of Feature Models

Participants: Sabine Moisan, Jean-Paul Rigault.

Keywords: feature models, model at run time, self-adaptive systems

In video understanding systems, context changes (detected by system sensors) are often unexpected and can combine in unpredictable ways, making it difficult to determine in advance (off line) the running configuration suitable for each context combination. To address this issue, we keep, at run time, a model of the system and its context together with its current running configuration. We adopted an enriched Feature Model approach to express the variability of the architecture as well as of the context. A context change is transformed into a set of feature modifications (selection/deselection of features) to be processed on the fly. This year we proposed a fully automatic mechanism to compute at run time the impact of the current selection/deselection requests. First, the modifications are checked for consistency; second, they are applied as a single atomic “transaction” to the current configuration to obtain a new configuration compliant with the model; finally, the running system architecture is updated accordingly. This year we implemented the reconfiguration step and its algorithms and heuristics and we evaluated its run time efficiency.

Our ultimate goal is to control the system through a feed back loop from video components and sensor events to feature model manipulation and back to video components modifications.

The fully automatic adaptation that we propose is similar to a Feature Model editor. That is the reason why our previous attempt was to embed a general purpose feature model editor at run time. This revealed two major differences between our mechanism and an editor. First, in a fully automatic process there is no human being to drive a series of edits, hence heuristics are required. Second, the editor operations are often elementary while we need a global “transaction-like” application of all the selections/deselections to avoid temporary unconsistencies.

In order to evaluate our algorithm performance, we randomly generated feature models (from 60 to 1400 features). We also randomly generated context changes. The results are shown on figure 24 : no processing time explosion is noticeable; in fact the time seems to grow rather linearly. Moreover, the computation time of a new initial partial configuration does not exceed 3ms for a rather big model. The algorithm and its evaluation are detailed in [41].

6.20. Setup and management of SafEE devices

Participants: Matias Marin, Etienne Corvée, François Brémond.

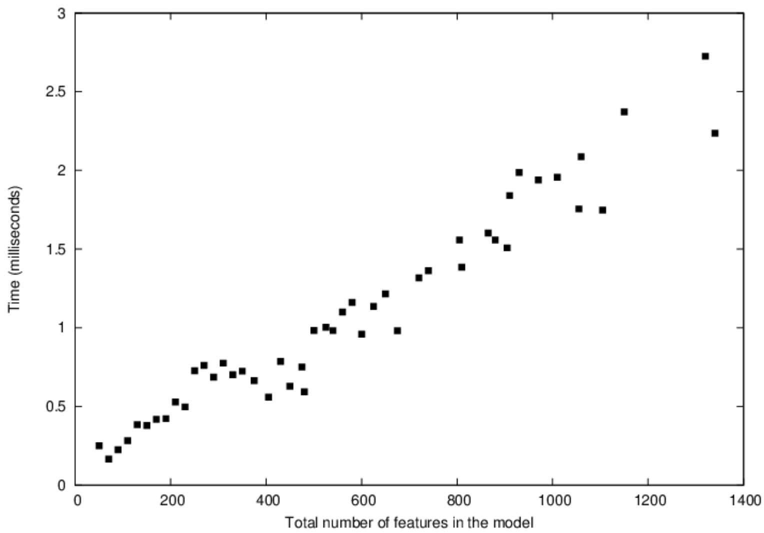


Figure 24. Computation time of initial models

The aim of the SafEE project (see section 8.1.1.2) is to provide assistance for the safety, autonomy and quality of life of elderly people at risk or already presenting Alzheimer's disease or related pathology.

Within EHPAD building (in Nice), 4 patients participated to our experiment and we plan to include more patients in the project throughout next years. Besides, 2 other patients have participated in the project at their own home.

More precisely, the SafEE project focuses on specific clinical targets: behavior, motricity, cognitive capabilities For this, the SafEE project includes:

- srvsafee(web server): a behavior analysis platform has been created to allow identification of certain daytime behavior disturbances (agitation, for example) and nocturnal disturbances (sleep disorders), locomotor capacities (walking and posture). It centralizes data saved in each local PC with Kinect2 sensor on the one hand, and postgresql database, on the other hand. About 30 Gb data are recorded for each patient in a day, which represents a huge amount to manage in the long run.
- Aroma diffuser (AromaCare): for sleep disturbances, using in particular an automated device for diffusing fragrances (aromatherapy) adapted to the perturbations detected by the analysis platform.
- Tablet (Serious game, MusicCare): for disturbances in spatial orientation, improved procedural memory and a sense of control and confidence in technological tools, using multimedia interfaces using an application for Android OS.
- Kinect2: motion detection for analysis linked to a PC, with a database to store recorded events.
- Bed sensor: able to track the sleep by analyzing the movements of the body, the breathing, and the beating of the heart.

Fig. 25 shows the Safee project environment.

6.21. Brick & Mortar Cookies

Participants: Julien Badie, Manikandan Bakthavatchalam, Vasanth Bathrinarayanan, Ghada Balhoul, Anais Ducoffe.

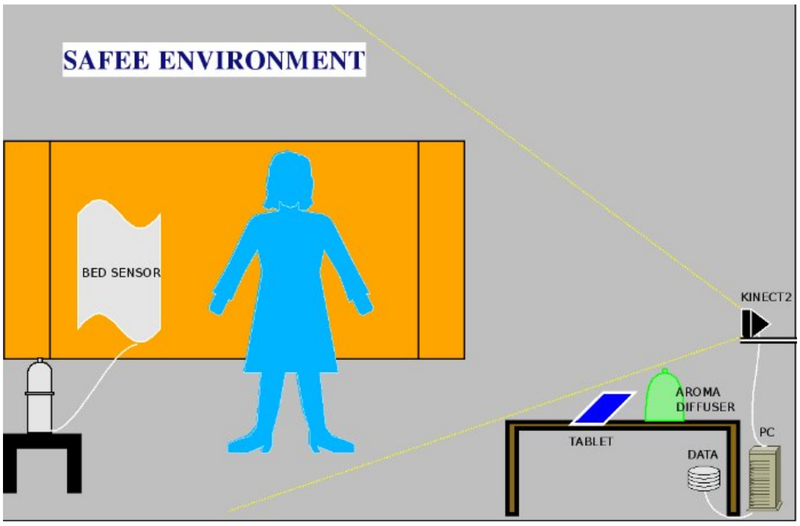


Figure 25. The Safee environment

The objective of the BMC project is to create a software that aims to present attendance and attractiveness of the customer in stores, based on automatic video analysis. The final system should be designed to be used without changing the current camera network of the customer store, dedicated to security purpose. Analysis should be given at different time and space resolutions. For instance, attendance of one particular day can be as interesting as attendance of the entire year. Moreover, shop owners want to be able to compare two given years or months, etc... As space resolution is concerned, the software should be able to give information about the global attractiveness of the store but should also analyze some specific zones.

IVA embedded on Bosch cameras

Intelligence Video Analysis (IVA) is embedded in some models of Bosch cameras. The algorithms are composed of human detection and tracking. They can be configured directly on the camera interface via *tasks*. We are using a live connection to get metadata directly from the camera stream using a RTSP connection. Thie year we improved the results of last year using calibration tool embedded in the camera : shape of people detected was better, feet were followed with more precision as bounding boxes were more stable. We also tested the new IVA developed by BOSCH which was built to better manage changes in scene brightness and crossing of people. In the former version people close to each other were often detected as one person. Our first tests in shop revealed that it reduces the number of false detection but people were detected later than in the previous version. The case of people crossing doesn't seem to be better managed than before.

Inria algorithms : people detection and tracking

The previously enumerated tasks use algorithms to detect people and get their trajectories. Stars team has developed similar algorithms and has adapted their parameters values to the specific needs of this software. To improve results after some tests made during summer, the people detection is now using a deep learning method. People are detected earlier than before with this new algorithm and people crossing and occlusions are far better managed. The performances and the reliability of those algorithms were tested using an annotation tool developed in Stars Team.

Annotation tool

Manual annotation of videos requires major human effort. It can take hours and hours of fastidious work to annotate a tiny set of data. That's why we propose a semi-automatic tool which reduces the time of the annotation. This new semi automatic annotation tool uses a simple input data format, XML file or XGTF file to describe the video contents and algorithms output. Users only have to correct false or missing detection and to fix some wrong object id of the algorithms results using the annotation tool interface.

Tests in real conditions

We tested our video acquisition tool and our people detection and people tracking algorithms during summer in a partner supermarket in Nice. We successfully acquire 2 weeks of the desired metadata. By the end of summer, our results were highly improved by using a deep learning method to detect people. Moreover we can get results in quasi real-time. Except for the video stream acquisition tool, which needs to be connected to the camera network, our system is now running on an independent and local network. In case there is a crash of our system, the supermarket network will not be affected. Moreover, sensitive data are protected. A test is starting soon in SuperU to run and evaluate this new prototype.

Metadata storage in database

Last year metadata outputs of our analysis were first stored in XML files. Now to manage the quasi real-time solution, metadata are stored directly in the database we designed last year. We improve architecture of this database to manage simultaneously several connections as the final solution is supposed to be composed of several servers which will manage several video streams at the same time.

Web interface (HIM)

The web graphic interface is in progress. User interactions were added and improved so that the interface should be more user-friendly. We also changed some charts and tables so that statistical results should be better understood by users.

THOTH Project-Team

7. New Results

7.1. Visual recognition in images

7.1.1. Convolutional Neural Fabrics

Participants: Shreyas Saxena, Jakob Verbeek.

Despite the success of CNNs, selecting the optimal architecture for a given task remains an open problem. Instead of aiming to select a single optimal architecture, in this work [20], we propose a "fabric" that embeds an exponentially large number of architectures. See 1 for a schematic illustration of how fabrics embed different architectures. The fabric consists of a 3D trellis that connects response maps at different layers, scales, and channels with a sparse homogeneous local connectivity pattern. The only hyper-parameters of a fabric are the number of channels and layers. While individual architectures can be recovered as paths, the fabric can in addition ensemble all embedded architectures together, sharing their weights where their paths overlap. Parameters can be learned using standard methods based on back-propagation, at a cost that scales linearly in the fabric size. We present benchmark results competitive with the state of the art for image classification on MNIST and CIFAR10, and for semantic segmentation on the Part Labels dataset.

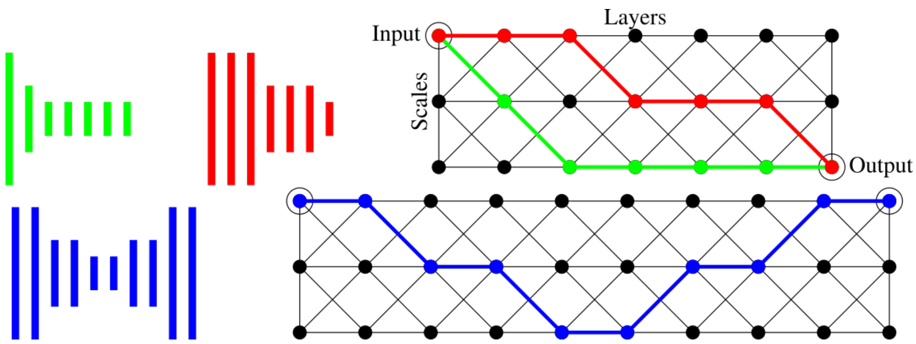


Figure 1. Fabrics embedding two seven-layer CNNs (red, green) and a ten-layer deconvolutional network (blue). Feature map size of the CNN layers are given by height. Fabric nodes receiving input and producing output are encircled. All edges are oriented to the right, down in the first layer, and towards the output in the last layer. The channel dimension of the 3D fabric is omitted for clarity.

7.1.2. Heterogeneous Face Recognition with CNNs

Participants: Shreyas Saxena, Jakob Verbeek.

Heterogeneous face recognition aims to recognize faces across different sensor modalities, see 2 for a schematic illustration. Typically, gallery images are normal visible spectrum images, and probe images are infrared images or sketches. Recently significant improvements in visible spectrum face recognition have been obtained by CNNs learned from very large training datasets. In this paper [21], we are interested in the question to what extent the features from a CNN pre-trained on visible spectrum face images can be used to perform heterogeneous face recognition. We explore different metric learning strategies to reduce the discrepancies between the different modalities. Experimental results show that we can use CNNs trained on visible spectrum images to obtain results that are on par or improve over the state-of-the-art for heterogeneous recognition with near-infrared images and sketches.

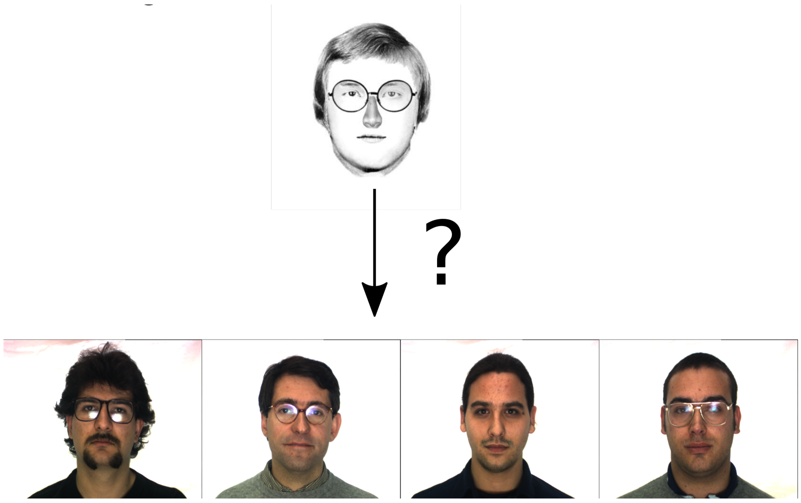


Figure 2. Schematic illustration for the task of heterogenous face recognition. The goal is to find the identity of the probe image (shown as a sketch) among one of the identities from the gallery set (shown in the bottom row). In contrast to standard face recognition, the probe and the gallery set do not share the same modality. In the illustartion, the probe image is a sketch and the galley images are normal visible spectrum images.

7.1.3. Mocap-guided Data Augmentation for 3D Pose Estimation in the Wild

Participants: Grégory Rogez, Cordelia Schmid.

In this paper [19], 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 3 . 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.

7.1.4. End-to-End Kernel Learning with Supervised Convolutional Kernel Networks

Participant: Julien Mairal.

In [16], we introduce a new image representation based on a multilayer kernel machine. Unlike traditional kernel methods where data representation is decoupled from the prediction task, we learn how to shape the kernel with supervision. We proceed by first proposing improvements of the recently-introduced convolutional kernel networks (CKNs) in the context of unsupervised learning; then, we derive backpropagation rules to take advantage of labeled training data. The resulting model is a new type of convolutional neural network, where optimizing the filters at each layer is equivalent to learning a linear subspace in a reproducing kernel Hilbert space (RKHS). We show that our method achieves reasonably competitive performance for image classification on some standard " deep learning " datasets such as CIFAR-10 and SVHN, and also for image

Figure 3. 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.

super-resolution, demonstrating the applicability of our approach to a large variety of image-related tasks. The model is illustrated in Figure 4 .

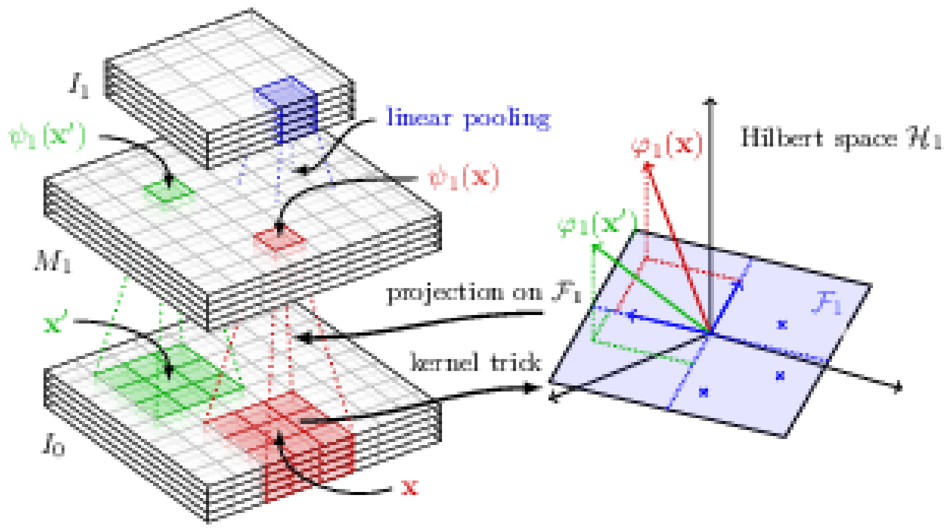


Figure 4. Our variant of convolutional kernel networks, illustrated between layers 0 and 1.

7.1.5. Semantic segmentation using Adversarial Networks

Participants: Pauline Luc, Camille Couprie [Facebook], Soumith Chintala [Facebook], Jakob Verbeek.

Adversarial training has been shown to produce state of the art results for generative image modeling. In [24], we propose an adversarial training approach to train semantic segmentation models. We train a convolutional semantic segmentation network along with an adversarial network that discriminates segmentation maps coming either from the ground truth or from the segmentation network, as shown in Figure 5 . The motivation for our approach is that it can detect and correct higher-order inconsistencies between ground truth segmentation maps and the ones produced by the segmentation net. Our experiments show that our adversarial training approach leads to improved accuracy on the Stanford Background and PASCAL VOC 2012 datasets.

7.1.6. Enhancing Energy Minimization Framework for Scene Text Recognition with Top-Down Cues

Participants: Anand Mishra [IIIT Hyderabad], Karteek Alahari, C. v. Jawahar [IIIT Hyderabad].

Recognizing scene text, i.e., text in images such as the one in Figure 6 , is a challenging problem, even more so than the recognition of scanned documents. This problem has gained significant attention from the computer vision community in recent years, and several methods based on energy minimization frameworks and deep learning approaches have been proposed. In our work presented in [8], we focus on the energy minimization framework and propose a model that exploits both bottom-up and top-down cues for recognizing cropped words extracted from street images. The bottom-up cues are derived from individual character detections from an image. We build a conditional random field model on these detections to jointly model the strength of the detections and the interactions between them. These interactions are top-down cues obtained from a lexicon-based prior, i.e., language statistics. The optimal word represented by the text image is obtained by minimizing the energy function corresponding to the random field model. We evaluate our proposed algorithm extensively on a number of cropped scene text benchmark datasets, namely Street View Text, ICDAR 2003, 2011 and 2013 datasets, and IIIT 5K-word, and show better performance than comparable methods. We perform a

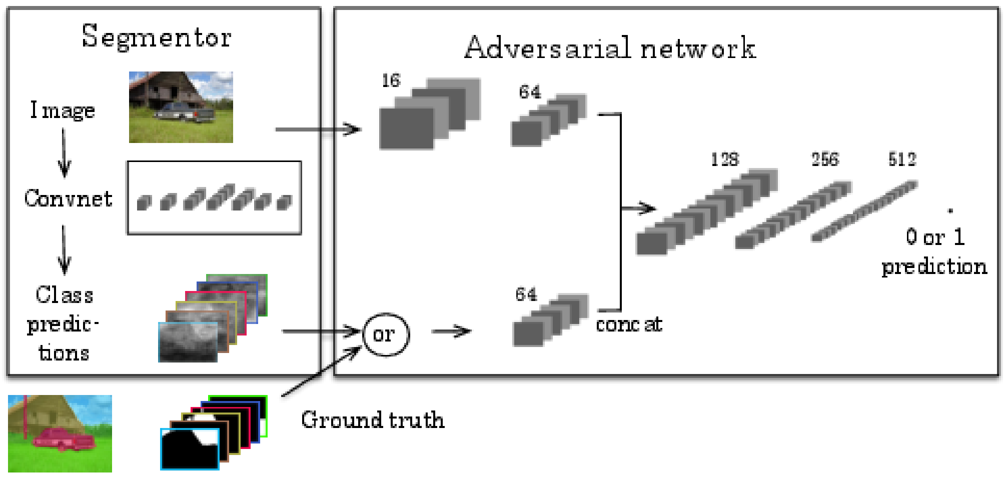


Figure 5. We use adversarial training to simultaneously learn a segmentation model (left) and a high order loss term to train it, given by the adversarial network (right). This encourages the segmentation model to output plausible segmentations, by enforcing forms of high order consistencies that are learned rather than manually designed.

rigorous analysis of all the steps in our approach and analyze the results. We also show that state-of-the-art convolutional neural network features can be integrated in our framework to further improve the recognition performance.

7.1.7. Local Convolutional Features with Unsupervised Training for Image Retrieval

Participants: Mattis Paulin, Matthijs Douze [Facebook], Zaid Harchaoui [University of Washington], Julien Mairal, Florent Perronnin [Xerox], Cordelia Schmid.

Patch-level descriptors underlie several important computer vision tasks, such as stereo-matching or content-based image retrieval. We introduce a deep convolutional architecture that yields patch-level descriptors, as an alternative to the popular SIFT descriptor for image retrieval. The proposed family of descriptors, called Patch-CKN[9], adapt the recently introduced Convolutional Kernel Network (CKN), an unsupervised framework to learn convolutional architectures. We present a comparison framework to benchmark current deep convolutional approaches along with Patch-CKN for both patch and image retrieval (see Fig. 7 for our pipeline), including our novel “RomePatches” dataset. Patch-CKN descriptors yield competitive results compared to supervised CNNs alternatives on patch and image retrieval.

7.2. Visual recognition in videos

7.2.1. Towards Weakly-Supervised Action Localization

Participants: Philippe Weinzaepfel, Xavier Martin, Cordelia Schmid.

In this paper [33], we present a novel approach for weakly-supervised action localization, i.e., that does not require per-frame spatial annotations for training. We first introduce an effective method for extracting human tubes by combining a state-of-the-art human detector with a tracking-by-detection approach. Our tube extraction leverages the large amount of annotated humans available today and outperforms the state of the art by an order of magnitude: with less than 5 tubes per video, we obtain a recall of 95% on the UCF-Sports and

Figure 6. A typical street scene image taken from Google Street View. It contains very prominent sign boards with text on the building and its windows. It also contains objects such as car, person, tree, and regions such as road, sky. Many scene understanding methods recognize these objects and regions in the image successfully, but overlook the text on the sign board, which contains rich, useful information. The goal of our work [8] is to address this gap in understanding scenes.

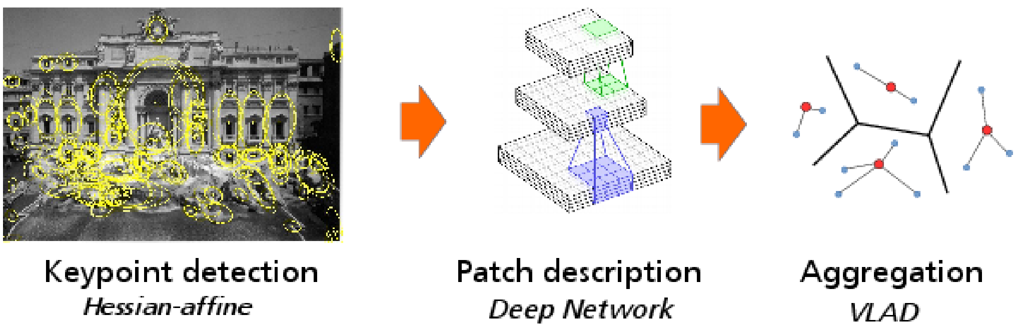


Figure 7. Image retrieval pipeline. Interest points are extracted with the Hessian-affine detector (left), encoded in descriptor space using convolutional features (middle), and aggregated into a compact representation using VLAD-pooling (right).

J-HMDB datasets. Given these human tubes, we perform weakly-supervised selection based on multi-fold Multiple Instance Learning (MIL) with improved dense trajectories and achieve excellent results. Figure 8 summarizes the approach. We obtain a mAP of 84% on UCF-Sports, 54% on J-HMDB and 45% on UCF-101, which outperforms the state of the art for weakly-supervised action localization and is close to the performance of the best fully-supervised approaches. The second contribution of this paper is a new realistic dataset for action localization, named DALY (Daily Action Localization in YouTube). It contains high quality temporal and spatial annotations for 10 actions in 31 hours of videos (3.3M frames), which is an order of magnitude larger than standard action localization datasets. On the DALY dataset, our tubes have a spatial recall of 82%, but the detection task is extremely challenging, we obtain 10.8% mAP.

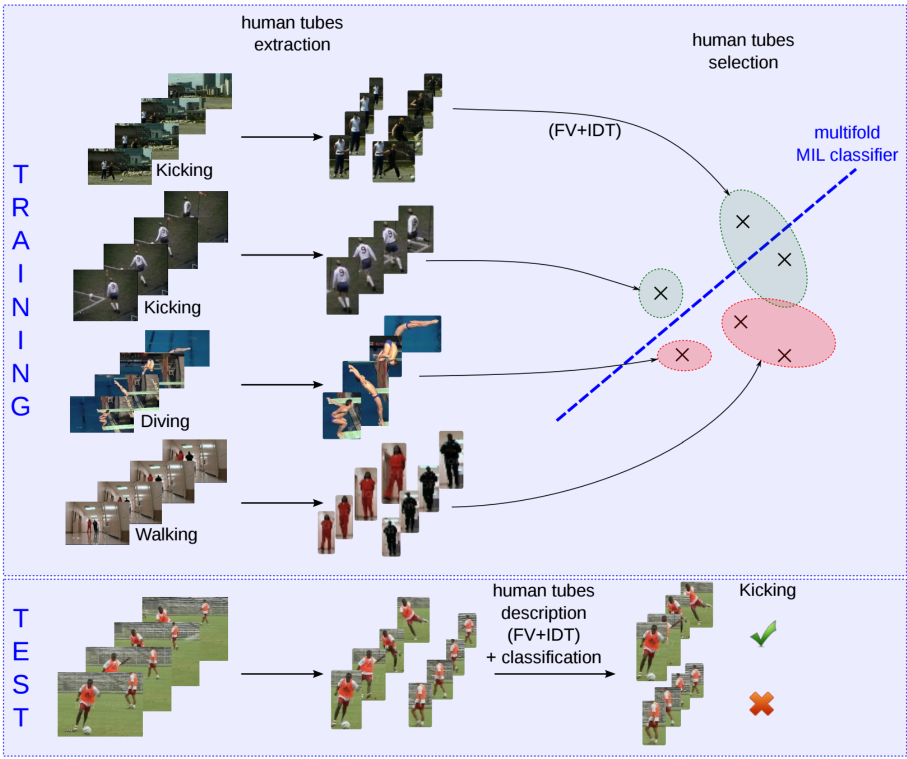


Figure 8. Overview of our approach for action localization without spatial supervision.

7.2.2. The DALY dataset

Participants: Philippe Weinzaepfel, Xavier Martin, Cordelia Schmid.

We introduce a new action localization dataset named DALY (Daily Action Localization in YouTube). DALY consists of more than 31 hours of videos (3.3M frames) from YouTube with 10 realistic daily actions, see Figure 9 , and 3.6k spatio-temporal instances. Annotations consist in the start and end time of each action instance, with high-quality spatial annotation for a sparse subset of frames. The task is to localize relatively short actions (8 seconds in average) in long untrimmed videos (3min 45 in average). Furthermore, it includes videos with multiple humans performing actions simultaneously. It overcomes the limitations of existing benchmarks that are limited to trimmed or almosttrimmed videos with specific action types, e.g. sports only, showing in most cases one human per video.

Figure 9. Overview of our approach for action localization without spatial supervision.

7.2.3. *Weakly-Supervised Semantic Segmentation using Motion Cues*

Participants: Pavel Tokmakov, Karteek Alahari, Cordelia Schmid.

Fully convolutional neural networks (FCNNs) trained on a large number of images with strong pixel-level annotations have become the new state of the art for the semantic segmentation task. While there have been recent attempts to learn FCNNs from image-level weak annotations, they need additional constraints, such as the size of an object, to obtain reasonable performance. To address this issue, in [23] we present motion-CNN (M-CNN), a novel FCNN framework which incorporates motion cues and is learned from video-level weak annotations. Our learning scheme to train the network uses motion segments as soft constraints, thereby handling noisy motion information, as shown in Figure 10 . When trained on weakly-annotated videos, our method outperforms the state-of-the-art EM-Adapt approach on the PASCAL VOC 2012 image segmentation benchmark. We also demonstrate that the performance of M-CNN learned with 150 weak video annotations is on par with state-of-the-art weakly-supervised methods trained with thousands of images. Finally, M-CNN substantially outperforms recent approaches in a related task of video co-localization on the YouTube-Objects dataset.

7.2.4. *Multi-region two-stream R-CNN for action detection*

Participants: Xiaojiang Peng, Cordelia Schmid.

This work [18] introduces a multi-region two-stream R-CNN model for action detection, see Figure 11 . It starts from frame-level action detection based on faster R-CNN and makes three contributions. The first one is the introduction of a motion region proposal network (RPN) complementary to a standard appearance RPN. The second is the stacking of optical flow over several frames, which significantly improves frame-level action detection. The third is the addition of a multi-region scheme to the faster R-CNN model, which adds complementary information on body parts. Frame-level detections are linked with the Viterbi algorithm, and action are temporally localized with the maximum subarray method. Experimental results on the UCF-Sports, J-HMDB and UCF101 action detection datasets show that the approach outperforms the state of the art with a significant margin in both frame-mAP and video-mAP.

7.2.5. *Analysing domain shift factors between videos and images for object detection*

Participants: Vicky Kalogeiton, Vittorio Ferrari [Univ. Edinburgh], Cordelia Schmid.

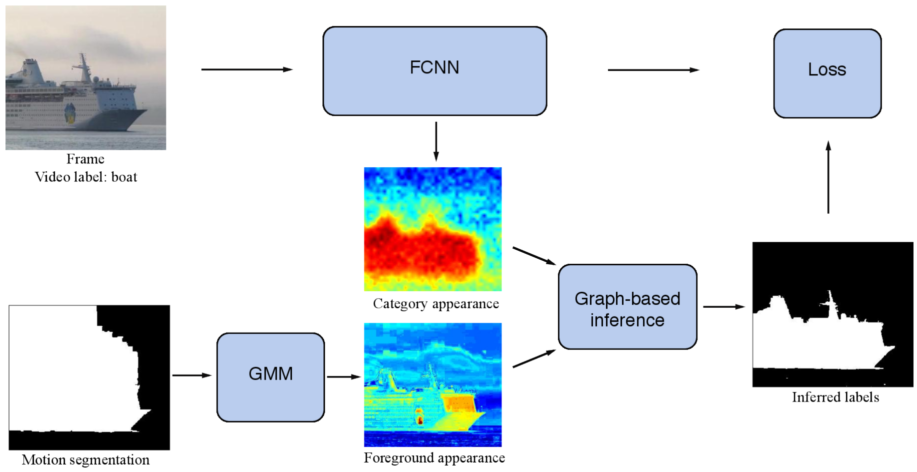


Figure 10. Overview of our M-CNN framework, where we show only one frame from a video example for clarity. The soft potentials (foreground appearance) computed from motion segmentation and the FCNN predictions (category appearance) jointly determine the latent segmentation (inferred labels) to compute the loss, and thus the network update.

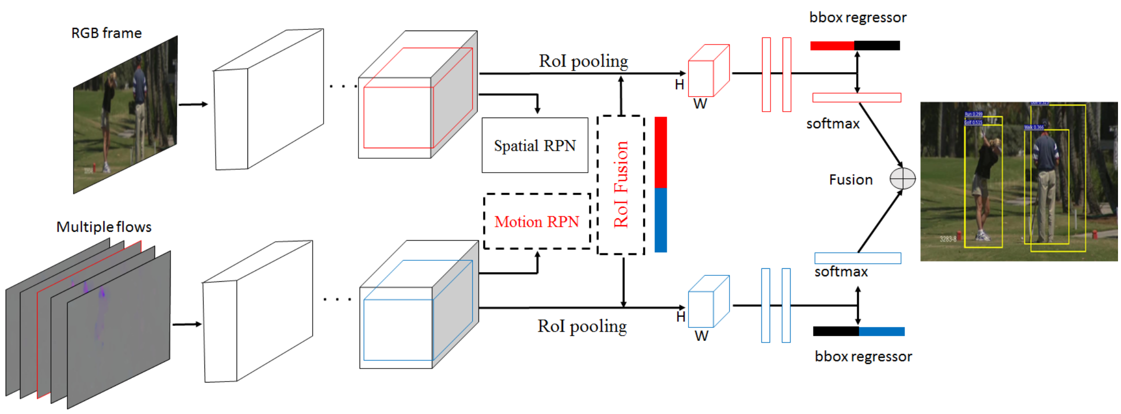


Figure 11. Two-stream faster R-CNN for spatio-temporal action detection.

Object detection is one of the most important challenges in computer vision. Object detectors are usually trained on bounding-boxes from still images. Recently, video has been used as an alternative source of data. Yet, for a given test domain (image or video), the performance of the detector depends on the domain it was trained on. In this paper [7], we examine the reasons behind this performance gap. We define and evaluate different domain shift factors (see Figure 12): spatial location accuracy, appearance diversity, image quality and aspect distribution. We examine the impact of these factors by comparing performance before and after factoring them out. The results show that all four factors affect the performance of the detectors and their combined effect explains nearly the whole performance gap.

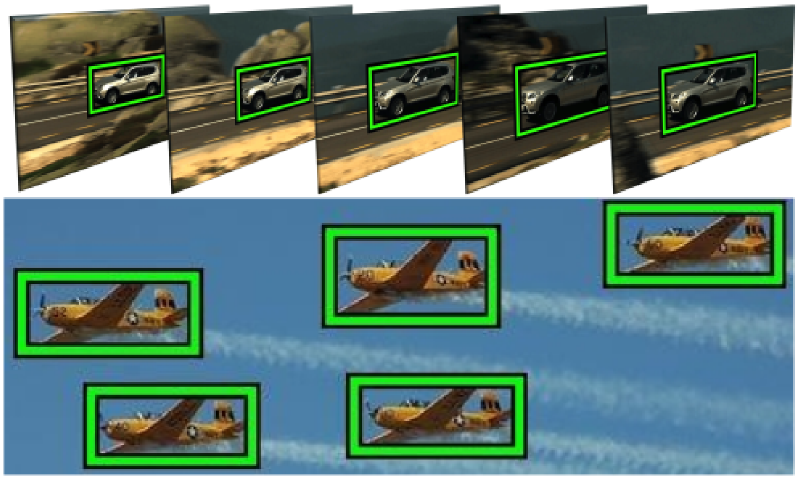


Figure 12. Example of apperance diversity domain shift factor. (top row): Frames in the same shot that contain near identical samples of an object. (bottom row): Example of near identical samples in the same image.

7.3. Large-scale statistical learning

7.3.1. Dictionary Learning for Massive Matrix Factorization

Participants: Julien Mairal, Arthur Mensch [Parietal], Gael Varoquaux [Parietal], Bertrand Thirion [Parietal].

Sparse matrix factorization is a popular tool to obtain interpretable data decompositions, which are also effective to perform data completion or denoising. Its applicability to large datasets has been addressed with online and randomized methods, that reduce the complexity in one of the matrix dimension, but not in both of them. In [25], [17], we tackle very large matrices in both dimensions. We propose a new factorization method that scales gracefully to terabyte-scale datasets. Those could not be processed by previous algorithms in a reasonable amount of time. We demonstrate the efficiency of our approach on massive functional Magnetic Resonance Imaging (fMRI) data, and on matrix completion problems for recommender systems, where we obtain significant speed-ups compared to state-of-the art coordinate descent methods. The main principle of the method is illustrated in Figure 13 .

7.3.2. Stochastic Optimization with Variance Reduction for Infinite Datasets with Finite-Sum Structure

Participants: Alberto Bietti, Julien Mairal.

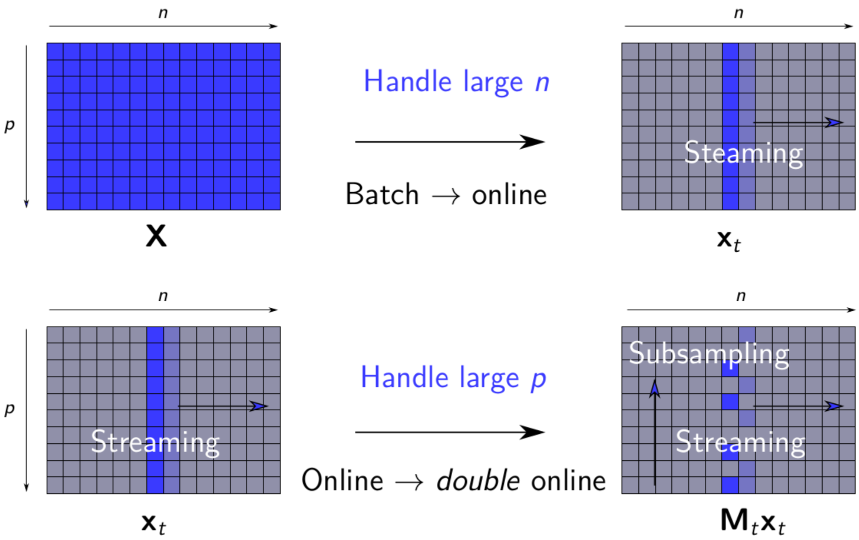


Figure 13. Illustration of the matrix factorization algorithm, which streams columns in one dimension while subsampling them.

Stochastic optimization algorithms with variance reduction have proven successful for minimizing large finite sums of functions. However, in the context of empirical risk minimization, it is often helpful to augment the training set by considering random perturbations of input examples. In this case, the objective is no longer a finite sum, and the main candidate for optimization is the stochastic gradient descent method (SGD). In this paper [26], we introduce a variance reduction approach for this setting when the objective is strongly convex. After an initial linearly convergent phase, the algorithm achieves a $O(1/t)$ convergence rate in expectation like SGD, but with a constant factor that is typically much smaller, depending on the variance of gradient estimates due to perturbations on a single example.

7.3.3. QuickeNing: A Generic Quasi-Newton Algorithm for Faster Gradient-Based Optimization

Participants: Hongzhou Lin, Julien Mairal, Zaid Harchaoui [University of Washington].

In this paper [28], we propose an approach to accelerate gradient-based optimization algorithms by giving them the ability to exploit curvature information using quasi-Newton update rules. The proposed scheme, called QuickeNing, is generic and can be applied to a large class of first-order methods such as incremental and block-coordinate algorithms; 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; with no line-search, it is also simple to use and to implement. Besides, it enjoys a worst-case linear convergence rate for strongly convex problems. We present experimental results, see Figure 14 , where QuickeNing gives significant improvements over competing methods for solving large-scale high-dimensional machine learning problems.

7.3.4. Dictionary Learning from Phaseless Measurements

Participants: Julien Mairal, Yonina Eldar [Technion], Andreas Tillmann [TU Darmstadt].

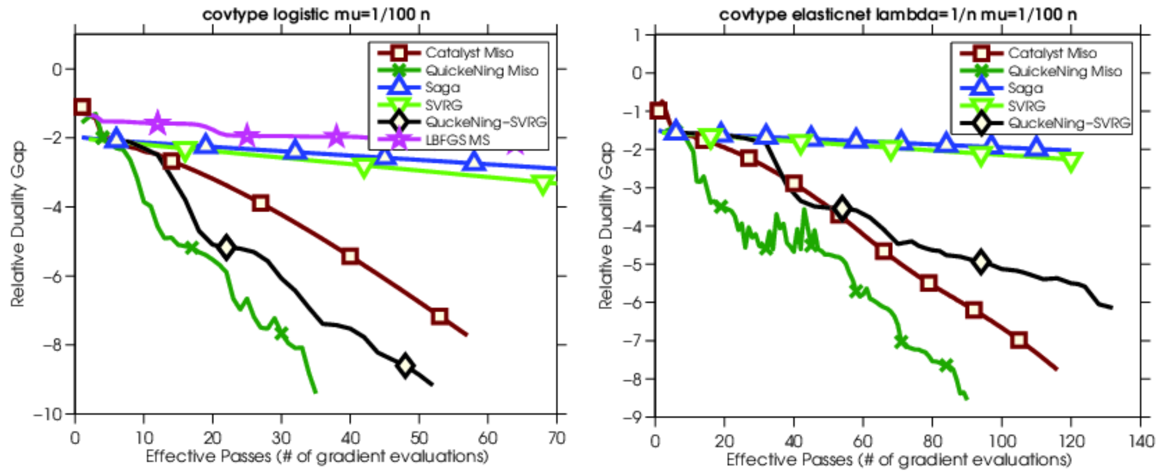


Figure 14. Relative duality gap for different number of passes performed over dataset covtype.

In [22], [12], we propose a new algorithm to learn a dictionary for reconstructing and sparsely encoding signals from measurements without phase. Specifically, we consider the task of estimating a two-dimensional image from squared-magnitude measurements of a complex-valued linear transformation of the original image. Several recent phase retrieval algorithms exploit underlying sparsity of the unknown signal in order to improve recovery performance. In this work, we consider such a sparse signal prior in the context of phase retrieval, when the sparsifying dictionary is not known in advance. Our algorithm jointly reconstructs the unknown signal—possibly corrupted by noise—and learns a dictionary such that each patch of the estimated image can be sparsely represented. Numerical experiments demonstrate that our approach can obtain significantly better reconstructions for phase retrieval problems with noise than methods that cannot exploit such "hidden" sparsity. Moreover, on the theoretical side, we provide a convergence result for our method.

WILLOW Project-Team

7. New Results

7.1. 3D object and scene modeling, analysis, and retrieval

7.1.1. Trinocular Geometry Revisited

Participants: Jean Ponce, Martial Hebert, Matthew Trager.

When do the visual rays associated with triplets of point correspondences converge, that is, intersect in a common point? Classical models of trinocular geometry based on the fundamental matrices and trifocal tensor associated with the corresponding cameras only provide partial answers to this fundamental question, in large part because of underlying, but seldom explicit, general configuration assumptions. In this project, we use elementary tools from projective line geometry to provide necessary and sufficient geometric and analytical conditions for convergence in terms of transversals to triplets of visual rays, without any such assumptions. In turn, this yields a novel and simple minimal parameterization of trinocular geometry for cameras with non-collinear or collinear pinholes, which can be used to construct a practical and efficient method for trinocular geometry parameter estimation. This work has been published at CVPR 2014, and a revised version that includes numerical experiments using synthetic and real data has been published in IJCV [7] and example results are shown in figure 1 .

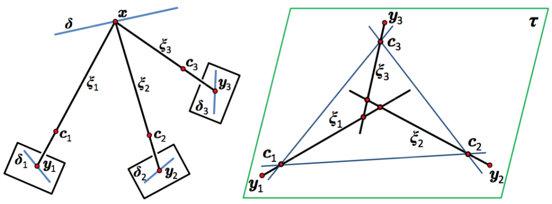


Figure 1. Left: Visual rays associated with three (correct) correspondences. Right: Degenerate epipolar constraints associated with three coplanar, but non-intersecting rays lying in the trifocal plane.

7.1.2. Consistency of silhouettes and their duals

Participants: Matthew Trager, Martial Hebert, Jean Ponce.

Silhouettes provide rich information on three-dimensional shape, since the intersection of the associated visual cones generates the "visual hull", which encloses and approximates the original shape. However, not all silhouettes can actually be projections of the same object in space: this simple observation has implications in object recognition and multi-view segmentation, and has been (often implicitly) used as a basis for camera calibration. In this paper, we investigate the conditions for multiple silhouettes, or more generally arbitrary closed image sets, to be geometrically "consistent". We present this notion as a natural generalization of traditional multi-view geometry, which deals with consistency for points. After discussing some general results, we present a "dual" formulation for consistency, that gives conditions for a family of planar sets to be sections of the same object. Finally, we introduce a more general notion of silhouette "compatibility" under partial knowledge of the camera projections, and point out some possible directions for future research. This work has been published in [16] and example results are shown in 2 .

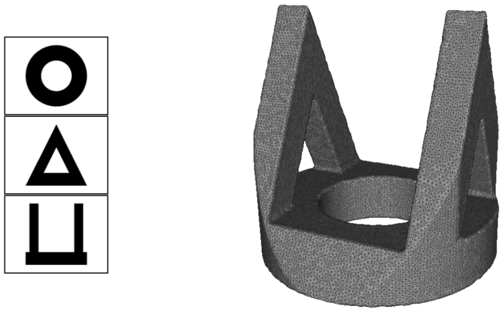


Figure 2. Geometrically consistent silhouettes are feasible projections of a single object.

7.1.3. Congruences and Concurrent Lines in Multi-View Geometry

Participants: Jean Ponce, Bernd Sturmfels, Matthew Trager.

We present a new framework for multi-view geometry in computer vision. A camera is a mapping between P^3 and a line congruence. This model, which ignores image planes and measurements, is a natural abstraction of traditional pinhole cameras. It includes two-slit cameras, pushbroom cameras, catadioptric cameras, and many more. We study the concurrent lines variety, which consists of n-tuples of lines in P^3 that intersect at a point. Combining its equations with those of various congruences, we derive constraints for corresponding images in multiple views. We also study photographic cameras which use image measurements and are modeled as rational maps from P^3 to P^2 or $P^1 \times P^1$. This work has been accepted for publication in [19] and example results are shown in 3 .

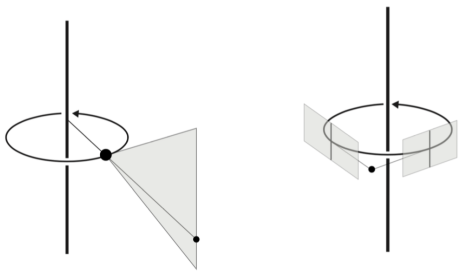


Figure 3. Non-central panoramic (left) and stereo panoramic cameras (right) are examples of non-linear cameras that can be modeled using line congruences.

7.1.4. NetVLAD: CNN architecture for weakly supervised place recognition

Participants: Relja Arandjelović, Petr Gronat, Akihiko Torii, Tomas Pajdla, Josef Sivic.

In [9], we tackle the problem of large scale visual place recognition, where the task is to quickly and accurately recognize the location of a given query photograph. We present the following three principal contributions. First, we develop a convolutional neural network (CNN) architecture that is trainable in an end-to-end manner directly for the place recognition task. The main component of this architecture, NetVLAD, is a new generalized VLAD layer, inspired by the "Vector of Locally Aggregated Descriptors" image representation commonly used in image retrieval. The layer is readily pluggable into any CNN architecture and amenable to training via backpropagation. Second, we develop a training procedure, based on a new weakly supervised ranking loss, to learn parameters of the architecture in an end-to-end manner from images depicting the same places over time downloaded from Google Street View Time Machine. Finally, we show that the proposed architecture obtains a large improvement in performance over non-learnt image representations as well as significantly outperforms off-the-shelf CNN descriptors on two challenging place recognition benchmarks. This work has been published at CVPR 2016 [9]. Figure 4 shows some qualitative results.

(a) Mobile phone query (b) Retrieved image of same place

Figure 4. Our trained NetVLAD descriptor correctly recognizes the location (b) of the query photograph (a) despite the large amount of clutter (people, cars), changes in viewpoint and completely different illumination (night vs daytime).

7.1.5. Pairwise Quantization

Participants: Artem Babenko, Relja Arandjelović, Victor Lempitsky.

We consider the task of lossy compression of high-dimensional vectors through quantization. We propose the approach that learns quantization parameters by minimizing the distortion of scalar products and squared distances between pairs of points. This is in contrast to previous works that obtain these parameters through the minimization of the reconstruction error of individual points. The proposed approach proceeds by finding a linear transformation of the data that effectively reduces the minimization of the pairwise distortions to the minimization of individual reconstruction errors. After such transformation, any of the previously-proposed quantization approaches can be used. Despite the simplicity of this transformation, the experiments demonstrate that it achieves considerable reduction of the pairwise distortions compared to applying quantization directly to the untransformed data. This work has been published on arXiv [18] and submitted to Neurocomputing journal.

7.1.5.1. Learning and Calibrating Per-Location Classifiers for Visual Place Recognition

Participants: Petr Gronat, Josef Sivic, Guillaume Obozinski [ENPC / Inria SIERRA], Tomáš Pajdla [CTU in Prague].

The aim of this work is to localize a query photograph by finding other images depicting the same place in a large geotagged image database. This is a challenging task due to changes in viewpoint, imaging conditions and the large size of the image database. The contribution of this work is two-fold. First, we cast the place recognition problem as a classification task and use the available geotags to train a classifier for each location in the database in a similar manner to per-exemplar SVMs in object recognition. Second, as only few positive training examples are available for each location, we propose a new approach to calibrate all the per-location SVM classifiers using *only* the negative examples. The calibration we propose relies on a significance measure essentially equivalent to the p-values classically used in statistical hypothesis testing. Experiments are performed on a database of 25,000 geotagged street view images of Pittsburgh and demonstrate improved place recognition accuracy of the proposed approach over the previous work. This work has been published at CVPR 2013, and a revised version that includes additional experimental results has been published at IJCV [3].

7.2. Category-level object and scene recognition

7.2.1. Proposal Flow

Participants: Bumsub Ham, Minsu Cho, Cordelia Schmid, Jean Ponce.

Finding image correspondences remains a challenging problem in the presence of intra-class variations and large changes in scene layout, typical in scene flow computation. In [10], we introduce a novel approach to this problem, dubbed proposal flow, that establishes reliable correspondences using object proposals. Unlike prevailing scene flow approaches that operate on pixels or regularly sampled local regions, proposal flow benefits from the characteristics of modern object proposals, that exhibit high repeatability at multiple scales, and can take advantage of both local and geometric consistency constraints among proposals. We also show that proposal flow can effectively be transformed into a conventional dense flow field. We introduce a new dataset that can be used to evaluate both general scene flow techniques and region-based approaches such as proposal flow. We use this benchmark to compare different matching algorithms, object proposals, and region features within proposal flow with the state of the art in scene flow. This comparison, along with experiments on standard datasets, demonstrates that proposal flow significantly outperforms existing scene flow methods in various settings. This work has been published at CVPR 2016 [10]. The proposed method and its qualitative result are illustrated in Figure 5 .

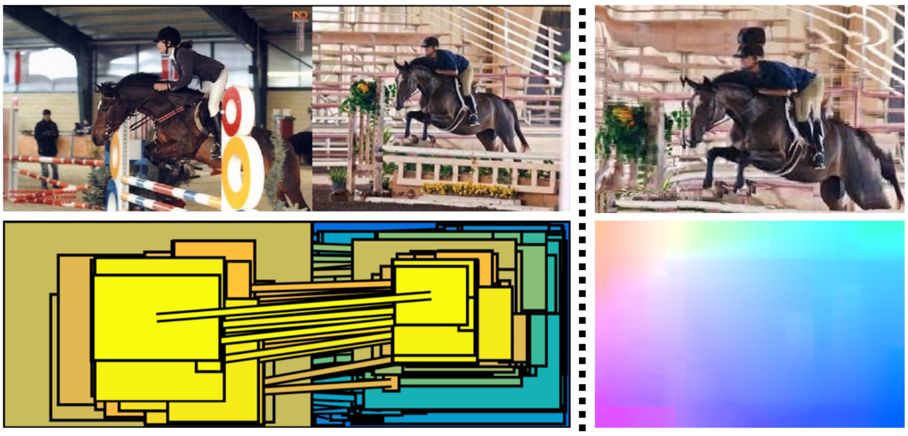


Figure 5. Proposal flow generates a reliable scene flow between similar images by establishing geometrically consistent correspondences between object proposals. (Left) Region-based scene flow by matching object proposals. (Right) Color-coded dense flow field generated from the region matches, and image warping using the flow.

7.2.1.1. Learning Discriminative Part Detectors for Image Classification and Cosegmentation

Participants: Jian Sun, Jean Ponce.

In this work, we address the problem of learning discriminative part detectors from image sets with category labels. We propose a novel latent SVM model regularized by group sparsity to learn these part detectors. Starting from a large set of initial parts, the group sparsity regularizer forces the model to jointly select and optimize a set of discriminative part detectors in a max-margin framework. We propose a stochastic version of a proximal algorithm to solve the corresponding optimization problem. We apply the proposed method to image classification and cosegmentation, and quantitative experiments with standard bench- marks show that it matches or improves upon the state of the art. The first version of this work has appeared at CVPR 2013. An extended version has been published at IJCV [6].

7.2.2. ContextLocNet: Context-aware deep network models for weakly supervised localization

Participants: Vadim Kantorov, Maxime Oquab, Minsu Cho, Ivan Laptev.

In [11] we aim to localize objects in images using image-level supervision only. Previous approaches to this problem mainly focus on discriminative object regions and often fail to locate precise object boundaries. In [11] we address this problem by introducing two types of context-aware guidance models, additive and contrastive models, that leverage their surrounding context regions to improve localization. The additive model encourages the predicted object region to be supported by its surrounding context region. The contrastive model encourages the predicted object region to be outstanding from its surrounding context region. Our approach benefits from the recent success of convolutional neural networks for object recognition and extends Fast R-CNN to weakly supervised object localization. Extensive experimental evaluation on the PASCAL VOC 2007 and 2012 benchmarks shows hat our context-aware approach significantly improves weakly supervised localization and detection. A high-level architecture of our model is presented in Figure 6 , the project webpage is at http://www.di.ens.fr/willow/research/contextlocnet/.

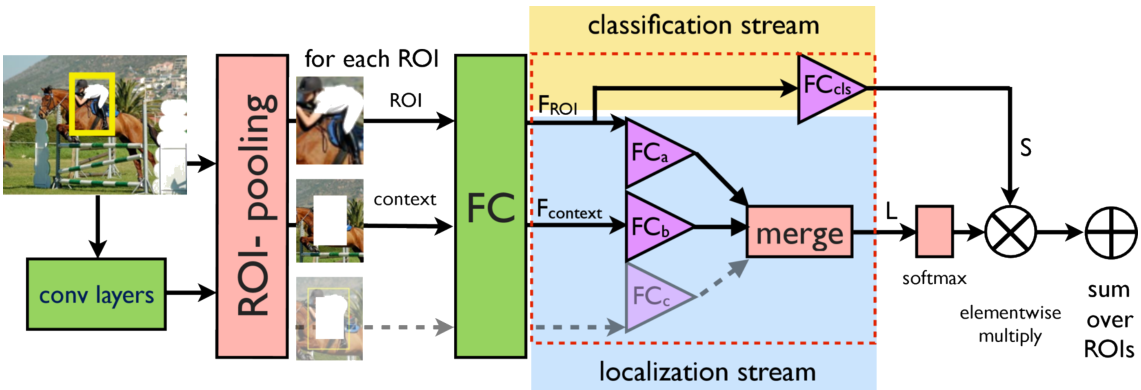


Figure 6. ContextLocNet improves localization by comparing an object score between a proposal and its context.

7.2.3. Faces In Places: Compound query retrieval

Participants: Yujie Zhong, Relja Arandjelović, Andrew Zisserman.

The goal of this work is to retrieve images containing both a target person and a target scene type from a large dataset of images. At run time this compound query is handled using a face classifier trained for the person, and an image classifier trained for the scene type. We make three contributions: first, we propose a hybrid convolutional neural network architecture that produces place-descriptors that are aware of faces and their corresponding descriptors. The network is trained to correctly classify a combination of face and scene classifier scores. Second, we propose an image synthesis system to render high quality fully-labelled face-and-place images, and train the network only from these synthetic images. Last, but not least, we collect and annotate a dataset of real images containing celebrities in different places, and use this dataset to evaluate the retrieval system. We demonstrate significantly improved retrieval performance for compound queries using the new face-aware place-descriptors. This work has been published at BMVC 2016 [17]. Figure 7 shows some qualitative results.

Figure 7. Examples of the top two retrieved images for various compound queries.

7.3. Image restoration, manipulation and enhancement

7.3.1. Robust Guided Image Filtering Using Nonconvex Potentials

Participants: Bumsub Ham, Minsu Cho, Jean Ponce.

Filtering images using a guidance signal, a process called joint or guided image filtering, has been used in various tasks in computer vision and computational photography, particularly for noise reduction and joint upsampling. The aim is to transfer the structure of the guidance signal to an input image, restoring noisy or altered image structure. The main drawbacks of such a data-dependent framework are that it does not consider differences in structure between guidance and input images, and it is not robust to outliers. We propose a novel SD (for static/dynamic) filter to address these problems in a unified framework by jointly leveraging structural information of guidance and input images. Joint image filtering is formulated as a nonconvex optimization problem, which is solved by the majorization-minimization algorithm. The proposed algorithm converges quickly while guaranteeing a local minimum. The SD filter effectively controls the underlying image structure at different scales and can handle a variety of types of data from different sensors. It is robust to outliers and other artifacts such as gradient reversal and global intensity shifting, and has good edge-preserving smoothing properties. We demonstrate the flexibility and effectiveness of the SD filter in a great variety of applications including depth upsampling, scale-space filtering, texture removal, flash/non-flash denoising, and RGB/NIR denoising. This has been published at CVPR 2015. A new revised version is currently in submission [4]. The SD filter is illustrated in Figure 8 .

7.4. Human activity capture and classification

7.4.1. Hollywood in Homes: Crowdsourcing Data Collection for Activity Understanding

Participants: Gunnar A. Sigurdsson, Gül Varol, Xiaolong Wang, Ali Farhadi, Ivan Laptev, Abhinav Gupta.

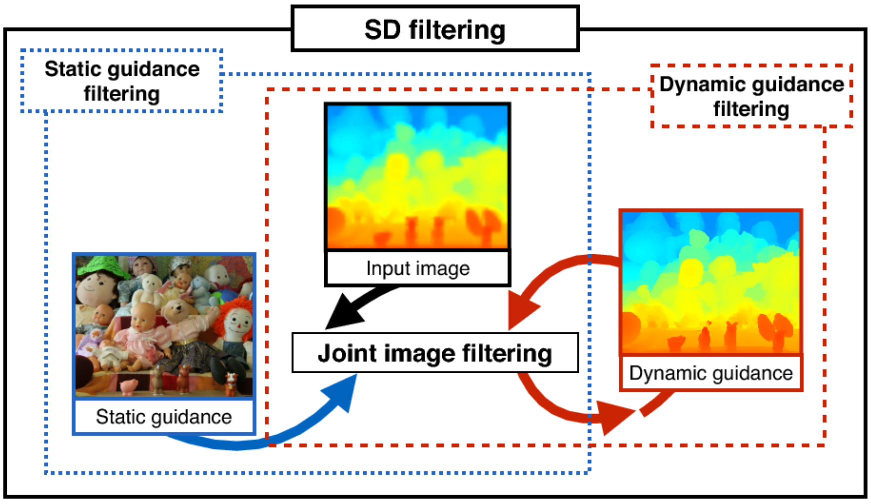


Figure 8. Sketch of joint image filtering and SD filtering: Static guidance filtering convolves an input image with a weight function computed from static guidance, as in the dotted blue box. Dynamic guidance filtering uses weight functions that are repeatedly obtained from regularized input images, as in the dotted red box. We have observed that static and dynamic guidance complement each other, and exploiting only one of them is problematic, especially in the case of data from different sensors (e.g., depth and color images). The SD filter takes advantage of both, and addresses the problems of current joint image filtering.

Computer vision has a great potential to help our daily lives by searching for lost keys, watering flowers or reminding us to take a pill. To succeed with such tasks, computer vision methods need to be trained from real and diverse examples of our daily dynamic scenes. While most of such scenes are not particularly exciting, they typically do not appear on YouTube, in movies or TV broadcasts. So how do we collect sufficiently many diverse but boring samples representing our lives? We propose a novel Hollywood in Homes approach to collect such data. Instead of shooting videos in the lab, we ensure diversity by distributing and crowdsourcing the whole process of video creation from script writing to video recording and annotation. Following this procedure we collect a new dataset, *Charades*, with hundreds of people recording videos in their own homes, acting out casual everyday activities (see Figure 9). The dataset is composed of 9,848 annotated videos with an average length of 30 seconds, showing activities of 267 people from three continents. Each video is annotated by multiple free-text descriptions, action labels, action intervals and classes of interacted objects. In total, Charades provides 27,847 video descriptions, 66,500 temporally localized intervals for 157 action classes and 41,104 labels for 46 object classes. Using this rich data, we evaluate and provide baseline results for several tasks including action recognition and automatic description generation. We believe that the realism, diversity, and casual nature of this dataset will present unique challenges and new opportunities for computer vision community. This work has been published at ECCV 2016 [15].

Figure 9. Comparison of actions in the Charades dataset and on YouTube: Reading a book, Opening a refrigerator, Drinking from a cup. YouTube returns entertaining and often atypical videos, while Charades contains typical everyday videos.

7.4.2. Unsupervised learning from narrated instruction videos

Participants: Jean-Baptiste Alayrac, Piotr Bojanowski, Nishant Agrawal, Josef Sivic, Ivan Laptev, Simon Lacoste-Julien.

In [8], we address the problem of automatically learning the main steps to complete a certain task, such as changing a car tire, from a set of narrated instruction videos. The contributions of this paper are three-fold. First, we develop a new unsupervised learning approach that takes advantage of the complementary nature of the input video and the associated narration. The method solves two clustering problems, one in text and one in video, applied one after each other and linked by joint constraints to obtain a single coherent sequence of steps in both modalities. Second, we collect and annotate a new challenging dataset of real-world instruction videos from the Internet. The dataset contains about 800,000 frames for five different tasks that include complex interactions between people and objects, and are captured in a variety of indoor and outdoor settings. Third, we experimentally demonstrate that the proposed method can automatically discover, in an unsupervised manner, the main steps to achieve the task and locate the steps in the input videos. This work has been published at CVPR 2016 [8].

7.4.3. Long-term Temporal Convolutions for Action Recognition

Participants: Gul Varol, Ivan Laptev, Cordelia Schmid.

Typical human actions such as hand-shaking and drinking last several seconds and exhibit characteristic spatio-temporal structure. Recent methods attempt to capture this structure and learn action representations with convolutional neural networks. Such representations, however, are typically learned at the level of single frames or short video clips and fail to model actions at their full temporal scale. In [20], we learn video representations using neural networks with long-term temporal convolutions. We demonstrate that CNN models with increased temporal extents improve the accuracy of action recognition despite reduced spatial resolution. We also study the impact of different low-level representations, such as raw values of video pixels and optical flow vector fields and demonstrate the importance of high-quality optical flow estimation for learning accurate action models. We report state-of-the-art results on two challenging benchmarks for human action recognition UCF101 and HMDB51. This work is under review. The results for the proposed method are illustrated in Figure 10 .

7.4.4. Thin-Slicing forPose: Learning to Understand Pose without Explicit Pose Estimation

Participants: Suha Kwak, Minsu Cho, Ivan Laptev.

In [12], we address the problem of learning a pose-aware, compact embedding that projects images with similar human poses to be placed close-by in the embedding space (Figure 11). The embedding function is built on a deep convolutional network, and trained with a triplet-based rank constraint on real image data. This architecture allows us to learn a robust representation that captures differences in human poses by effectively factoring out variations in clothing, background, and imaging conditions in the wild. For a variety of pose-related tasks, the proposed pose embedding provides a cost-efficient and natural alternative to explicit pose estimation, circumventing challenges of localizing body joints. We demonstrate the efficacy of the embedding on pose-based image retrieval and action recognition problems. This work has been published at CVPR 2016 [12].

7.4.5. Instance-level video segmentation from object tracks

Participants: Guillaume Seguin, Piotr Bojanowski, Rémi Lajugie, Ivan Laptev.

In [14], we address the problem of segmenting multiple object instances in complex videos. Our method does not require manual pixel-level annotation for training, and relies instead on readily-available object detectors or visual object tracking only. Given object bounding boxes at input as shown in Figure 12 , we cast video segmentation as a weakly-supervised learning problem. Our proposed objective combines (a) a discriminative clustering term for background segmentation, (b) a spectral clustering one for grouping pixels of same object instances, and (c) linear constraints enabling instance-level segmentation. We propose a convex relaxation of this problem and solve it efficiently using the Frank-Wolfe algorithm. We report results and compare our method to several baselines on a new video dataset for multi-instance person segmentation. This work has been published at CVPR 2016.

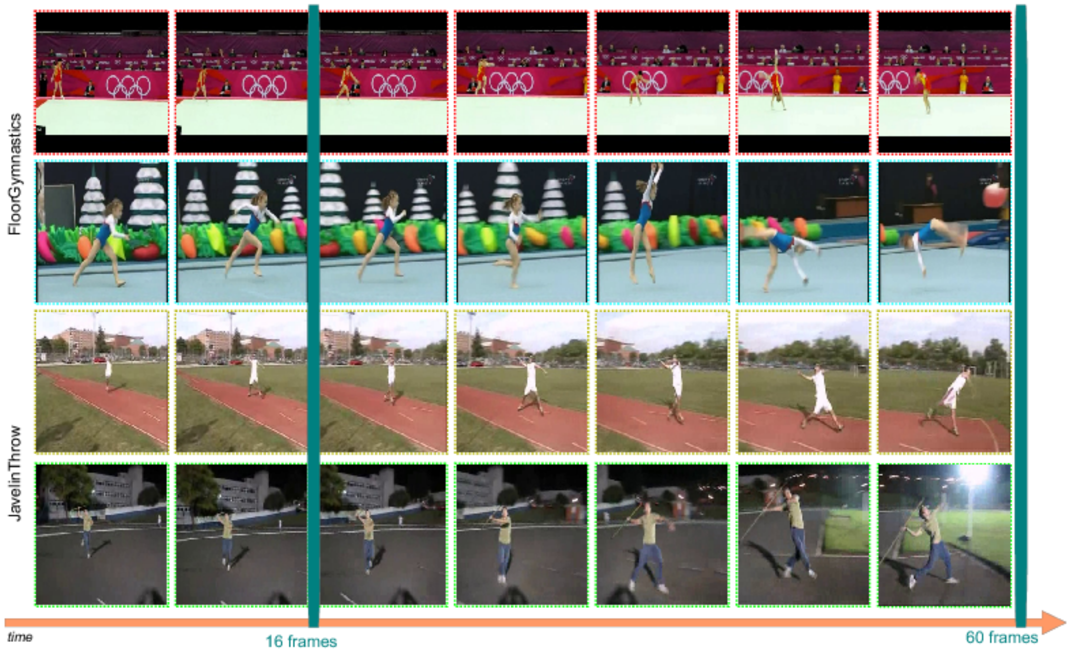


Figure 10. The highest improvement of long-term temporal convolutions in terms of class accuracy is for "JavelinThrow". For 16-frame network, it is mostly confused with "FloorGymnastics" class. We visualize sample videos with 7 frames extracted at every 8 frames. The intuitive explanation is that both classes start by running for a few seconds and then the actual action takes place. Long-term temporal convolutions with 60 frames can capture this interval, whereas 16-frame networks fail to recognize such long-term activities.

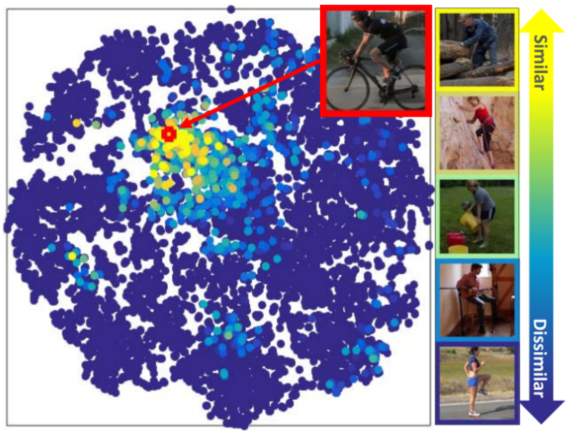


Figure 11. The manifold of our pose embedding visualized using t-SNE. Each point represents a human pose image. To better show correlation between the pose embedding and annotated pose, we color-code pose similarities in annotation between an arbitrary target image (red box) and all the other images. Selected examples of color-coded images are illustrated in the right-hand side. Images similar with the target in annotated pose are colored in yellow, otherwise in blue. As can be seen, yellow images lie closer by the target in general, which indicates that a position on the embedding space implicitly represents a human pose.

Figure 12. Results of our method applied to multi-person segmentation in a sample video from our database. Given an input video together with the tracks of object bounding boxes (left), our method finds pixel-wise segmentation for each object instance across video frames (right).