RESEARCH CENTER

FIELD
Perception, Cognition and Interaction

Activity Report 2015

Section New Results

Edition: 2016-03-21

VISION, PERCEPTION AND MULTIMEDIA INTERPRETATION

DAHU Project-Team

6. New Results

6.1. Specification and verification of data-driven systems

Process-centric views of data-driven workflows. Declarative, data-aware workflow models are becoming increasingly pervasive. While these have numerous benefits, views describing valid sequences of tasks are also useful to provide stake-holders with high-level descriptions of the workflow. In [23], we study the problem of recovering process-centric views from declarative, data-aware workflow specifications. The views are most naturally specified by finite-state transition systems describing regular languages. The results characterize when process-centric views of artifact systems are regular, with both linear and branching-time semantics.

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. We have fundamental results on the computational complexity of these problems, including the first explicit upper bound for reachability in VAS [24] and the best known lower bound for reachability in branching VAS [17] (where it is currently unknown whether reachability is decidable at all). We have furthermore defined a first sequent calculus for a modal data logic [29] as preliminary groundwork for the ANR PRODAQ project on proof systems for data queries.

6.2. Query processing for the Web

Query languages for graph databases. Graph-structured data on the Web can be found in emerging applications such as *RDF* and *linked data*, or *social networks*. Classical database languages are not suitable to query such data, essentially because they do not allow to (easily) express simple connectivity queries, which are the basic building block in graph navigation. We use Regular Path Queries, computing pairs of nodes reachable via a path satisfying a regular expression. We have tackled the problem of answering queries over graph databases which are available only trough a given set of views. We have shown that in the "asymptotic case", i.e. when the query is large enough relative to the view definition, it is decidable whether the view determines the query [22].

6.3. Distributed knowledge base.

Webdamlog The Webdamlog language is an extension of datalog to the distributed context, with *delegation* as the main novelty. A summary of the project was presented in [20].

We introduced an access control mechanism based on provenance in [26]. This access control is designed for a distributed setting. Peers getting data are also willing to enforce access control on that data, so that the owner of the data keeps some control over it when the data is passed around in the network. A second version of Webdamlog was developed in 2015 at Drexel, primarily by Vera Moffit also as part of her thesis (in collaboration with S. Abiteboul). It includes access control mechanism.

DREAM Project-Team

7. New Results

7.1. Simulator-based decision support

Participants: Philippe Besnard, Marie-Odile Cordier, Anne-Isabelle Graux, Christine Largouët, Véronique Masson, Laurence Rozé.

7.1.1. Ecosystem model-checking for decision-aid

Former studies of ecosystem modelling have concentrated on temporal modelling. In recent studies we have focussed on the formalization of spatial diffusion of a prey-predator trophic network composed of weeds and ground beetle. For this purpose, an approach coupling landscape representation and population models has been used. A reaction-diffusion model was developped through the synchronization ability of timed-automata. The agronomical rules of beetle migration and weeds diffusion have been translated into communications between timed automata. Landscapes have been simulated and can be evaluated thanks to landscape-metrics distance. The optimization aims to maximize the ground beetle abundance while minimizing the use of pesticides. The model obtained in this first study is quite complex but preliminary results are beeing studied.

7.1.2. Controler synthesis for optimal strategy search

Similarly to previous work, this approach relies on a qualitative model of a dynamical system. The problem consists in finding a strategy in order to help the user achieveing a specific goal. The model is now considered as a timed game automata expressing controllable and uncontrollable actions. The strategy represents the sequence of actions that can be performed by a user to reach a particular state (in case of a reachability problem for instance). A first approach based on a "generate and test" method has been developped for the marine ecosystem example [86].

Recently, we generalized the work of Yulong Zhao applied in the context of a dairy production system [87] to the planning domain. The planning task 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 on interacting agents such that the state to reach is expressed using temporal logic. We have chosen to represent each agent using the formalism of Priced Timed Game Automata (PTGA). PTGA is an extension of Timed Automata that allows the representation of cost on actions and uncontrollable actions. Relying on this domain description, we define a planning algorithm that computes the best strategy to achieve the goal. This algorithm is based on recognized model-checking and synthesis tools from the UPPAAL suite. The expressivity of this approach is evaluated 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.

7.1.3. A datawarehouse for simulation data

In previous work we have proposed a datawarehouse architecture to store the huge data produced by deep agricultural simulation models [50]. This year, we have worked on hierarchical skyline queries to introduce skyline queries in a datawarehouse framework. Conventional skyline queries retrieve the skyline points in a context of dimensions with a single hierarchical level. However, in some applications with multidimensional and hierarchical data structure (e.g. data warehouses), skyline points may be associated with dimensions having multiple hierarchical levels. Thus, we have proposed an efficient approach reproducing the effect of the OLAP operators "drill-down" and "roll-up" on the computation of skyline queries [52]. It provides the user with navigation operators along the dimensions hierarchies (i.e. specialize / generalize) while ensuring an online calculation of the associated skyline.

Anne-Isabelle Graux, on leave from INRA (National Institute for Agronomical Research), is working on an adaptation and extension of this method for storing the simulation results of a comprehensive farm model named MELODIE [53]. The new datawarehouse will enable the analysis of simulation results within dynamic preferences, related to grassland management for instance, for identifying the data satisfying the best compromises with respect to possibly inconsistent criteria.

7.1.4. Post-mining classification rules

We consider sets of classification rules with quantitative and qualitative attributes inferred by supervised machine learning, as in the framework of the Sacadeau project. Our aim is to improve the human understanding of such sets of rules. First, we consider quantitative attributes in rules that often contain too many intervals which are difficult to intepret. We propose two algorithms to merge some of these intervals in order to get more understandable rules. These algorithms take into account the final rule quality. We are also working on formalizing what could be the quality of a set of rules. There are lots of studies about the quality of one rule but very few about the quality of the whole set of rules and this is still an issue.

7.2. Data Mining

Participants: Marie-Odile Cordier, Yann Dauxais, Serge Vladimir Emteu Tchagou, Clément Gautrais, Thomas Guyet, Yves Moinard, Benjamin Negrevergne, René Quiniou, Laurence Rozé, Alexandre Termier.

7.2.1. Sequential pattern mining with intervals

In previous work, we developed a framework for sequential pattern mining with intervals [3]. It has been applied in various application (care-pathways, customer relationship management databases [35], etc.).

This year we explored chronicle mining algorithms for mining care-pathways (see section 9.1.1 , for an applicative context). Chronicles are alternative patterns for representing temporal behaviors [58]. A chronicle can be briefly defined as a set of events linked by constraints indicating the minimum and maximum time elapsed between two events. A care-pathway contains point-based events (e.g. surgery) and interval-based events (e.g. drug exposures). A chronicle can express such a complex temporal behaviour, for instance: *The patient was exposed to a drug X between 1 and 2 years, he met his doctor between 400 to 600 days after the beginning of the exposure and, finally, he was hospitalized.*

The first algorithm we worked on [23] is an adaptation of existing chronicle mining algorithms [55], [63] to mine the complete set of frequent chronicles from a collection of care-pathways. This algorithm uses the search-space browsing strategy of HDCA [55] and the support evaluation of CCP-Miner [63]. As the complete set of chronicle is huge, we also proposed an incomplete algorithms based on the original simplifications of [58]. These algorithms were implemented and evaluated on real and simulated datasets.

We also investigated discriminant chronicles mining which consists in extracting the chronicles that are α times more frequent in a database $\mathcal{D}_+$ than in a database $\mathcal{D}_-$. Mining discriminant chronicles is very useful to discover the features of care-pathways that are related, for instance, to a specific disease. Our approch has been implemented and is under evaluation.

7.2.2. Multiscale segmentation of satellite image time series

Satellite images enable the acquisition of large-scale ground vegetation information. Images have been recorded for several years with a high acquisition frequency (one image every two weeks). Such data are called satellite image time series (SITS). Several articles were published this year and they correspond to past work on algorithms and method to analyse SITS.

In [11], we presented a method to segment an image through the characterization of the evolution of a vegetation index (NDVI) on two scales: annual and multi-year. The main issue of this approach was the required computation resources (time and memory).

We also explored the supervised classification of SITS using classification trees for time-series [27] by implementing a parallelized version of this algorithm. Next, we explored the adaptation of the object-oriented segmentation to SITS. The object-oriented segmentation is able to segment images based on segment uniformity. We proposed a measure for time-series uniformity and applied the adapted algorithm on large multivariate SITS of Senegal [10].

Third, we presented an supervised approach to extract features from classified satellite images to analyse urban sprawl [28]. In this work, we have satellite images at only two dates, and the objective is to identify characteristics that can foster or prevent changes.

Our satellite images analysis approaches are used in two applicative contexts: understanding urban sprawl and analyzing drought in Senegal. Analysis of urban sprawl was a collaborative work with collegues in remote sensing, in landscapes analysis and in economical modelling. Our collective contribution was published in a book of the PDD2 [0] program [38]. Analysis of drought in Senegal is a long term collaboration with H. Nicolas (INRA/SAS) that we would like to continue in a collaboration with A. Fall (Université of Dakar) to confront our results with ground observations.

7.2.3. Analysis and simulation of landscape based on spatial patterns

Researchers in agro-environment need a great variety of landscapes to test their scientific hypotheses using agro-ecological models. Real landscapes are difficult to acquire and do not enable the agronomists to test all their hypotheses. Working with simulated landscapes is then an alternative to get a sufficient variety of experimental data. Our objective is to develop an original scheme to generate landscapes that reproduce realistic interface properties between parcels. This approach consists of the extraction of spatial patterns from a real geographic area and the use of these patterns to generate new "realistic" landscapes. It is based on a spatial representation of landscapes by a graph expressing the spatial relationships between the agricultural parcels (as well as the roads, the rivers, the buildings, etc.), in a specific geographic area.

In past years, we worked on the exploration of graph mining techniques, such as gSPAN [85], to discover the relevant spatial patterns present in a spatial-graph. We assume that the set of the frequent graph patterns are the characterisation of the landscape. Our remaining challenge was to simulate new realistic landscapes that reproduce the same patterns.

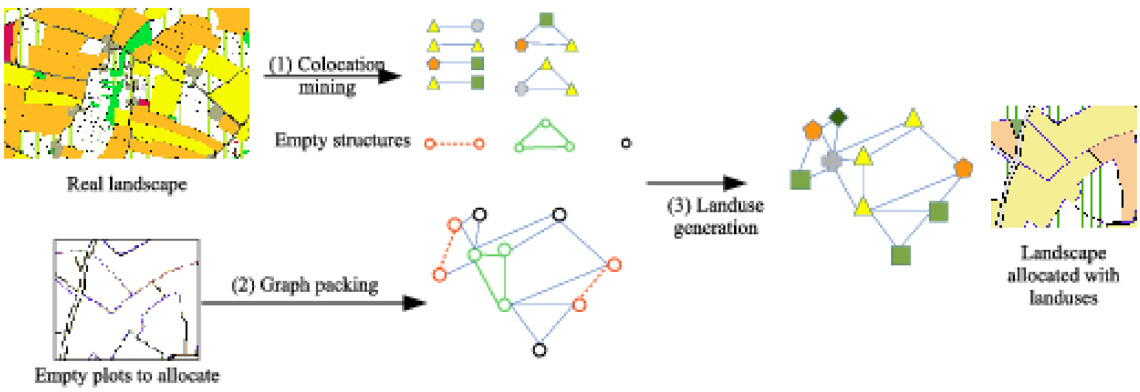


Figure 1. Simulation process in three steps: 1) characteristic graph-patterns mining, 2) graph packing of the cadastral landscape and 3) crop allocation.

[0]PDD2: Paysage Developpement Durable/Landscape Sustainable Developement

We have formalized the simulation process as a graph packing problem [66]. The process is illustrated by Figure 1 . Solving instances of the general graph packing problem has a high combinatorics and no efficient algorithm can solve it. We proposed an ASP program to tackle the combinatorics of the graph packing and to assign the land use considering some expert knowledge. Our approach combines the efficiency of ASP to solve the packing issue and the simplicity of the declarative programming to take into account expert contraints on the land use. Contraints about the minimum surface of crops or about the impossibility of some crops colocation can be easily defined. This work have been presented at the conference RFIA and an extended version has been published in the Revue d'Intelligence Artificielle (RIA) [13].

In addition, we are collaborating with J. Nicolas (EPI Dyliss) to improve the efficiency of our first programs. The improvements are based on symmetry breaking of ASP programs. To this end, we proposed a simplified encoding of the graph patterns using spanning trees and used automorphism detection in graph patterns to automatically encodes symmetry breakings. Intensive evaluation of our encoding shown that this improvement enable to tackle significantly larger graphs than early programs did. This work will be soon submitted to a high ranking conference.

7.2.4. Mining with ASP

In pattern mining, a pattern is considered interesting if it occurs frequently in the data, i.e. the number of its occurrences is greater than a fixed given threshold. As non informed mining methods tend to generate massive results, there is more and more interest in pattern mining algorithms able to mine data considering some expert knowledge. Though a generic pattern mining tool that could be tailored to the specific task of a data-scientist is still a holy grail for pattern mining software designers, some recent attempts have proposed generic pattern mining tools [61] for itemset mining tasks. In collaboration with Torsten Schaub, we explore the ability of a declarative language, such as Answer Set Programming (ASP), to solve pattern mining tasks efficiently. In 2011, Jarvisälo proposed a first attempt devoted to itemset mining [64]. In Dream, we are working on sequential pattern mining, which is known to be more challenging than itemset mining and which has been also recently considered by constraint programming approaches [76].

We have worked on encoding in ASP most of sequential pattern mining tasks: sequences with constraints (gaps, maximum length, etc.), closed/maximal patterns, emergent sequences. Our first result is to show that ASP is suitable for encoding such complex pattern mining tasks. The experimental results show that our purely declarative approach is less efficient than constraint programming approaches [36]. Nonetheless, it is suitable to be blend with intensive knowledge. The challenge is now to show that our ASP framework can extract the meaningful patterns that other approaches loose in the overwhelming amount of sequential patterns.

A first attempt has been done in this direction in collaboration with J. Romero from the University of Potsdam. We used the system ASPRIN to define preferences on patterns. Defining preferences on patterns is also a classical approach to select the most interesting patterns. Some classical preferences on sequential patterns have been defined and the ASPRIN system is used to extract the preferred patterns according to one preference or a combination of preferences (skypatterns [81])

This work will be soon submitted to a high ranking international conference.

7.2.5. Mining time series

Monitoring cattle. Following the lines of a previous work [79], we are working on a method for detecting Bovine Respiratory Diseases (BRD) from behavioral (walking, lying, feeding and drinking activity) and physiological (rumen temperature) data recorded on feedlot cattle being fattened up in big farms in Alberta (Canada). This year, we have especially worked on multivariate sensor data analysis, especially on the evaluation of different combinations of sensors for determining the best configuration and parameter setting. This work was part of Afra Verena Mang's master thesis defended in september 2015 [73]. Two papers are in preparation.

SIFT-based time-series symbolisation Time series classification is an application of particular interest with the increase of such data. Computing the distance between time-series is time consuming. An abstract representation of time-series that accurately approximates distances between time-series and makes easier their comparison is highly expected. In [17], we proposed a time series classification scheme grounded on the SIFT framework [70] adapted to time series. The SIFTs feed a Bag-of-Words representation of time-series. We have shown that this framework efficiently and accurately classifies time series, despite the fact that BoW representation ignores temporal order.

Mining sequential patterns from multimedia data Analyzing multimedia data to extract knowledge is a challenging problem due to the quantity and complexity of such data. Finding recurrent patterns is one method to structure and segment the data. In a collaboration with the EPI LinkMedia, we have proposed audio data symbolization and sequential pattern mining methods to extract patterns from audio streams. Experiments show this the task is hard and that the symbolization is a critical step for extracting relevant audio patterns [29].

7.2.6. Mining customer data for predicting and explaining attrition

Predicting customer defection in a retail context is difficult because, in most situations, the customer does not leave the store totally (there is no contract break as with banks or phone operators). We have proposed a new pattern model for representing the evolution of an individual customer purchase behavior that enables to early detect and to explain customer attrition. In particular, this model enables the analyst to determine which important kinds of product receives less and less attention from the customer. Thus, this model provides actionable knowledge at an individual scale that lets the retailer trigger targeted marketing actions to counter attrition. A poster has been submitted to the EBDT conference. This work has been performed during Clément Gautrais's master [59] and will be further investigated and extended during his PhD.

7.2.7. Mining energy consumption data

Machine tools in companies consume a lot of energy (before, during and after producing worked pieces). This year, we are beginning to work, with the start-up Energiency, on mining machine tool energy consumption data in order to propose energy savings to the companies. Firstly, we try to determine, according to the analyzed company, which data-mining algorithm should be used and which is the best configuration and parameter setting. Then, we aim to extract actions rules from patterns to help companies to consume less energy.

7.2.8. Trace reduction

One problem of execution trace of applications on embedded systems is that they can grow very large, typically several Gigabytes for 5 minutes of audio/video playback. Some endurance tests require continuous playback for 96 hours, which would lead to hundreds of Gigabytes of traces, that current techniques cannot analyze. We have proposed TraceSquiz, an online approach to monitor the trace output during endurance test, in order to record only suspicious portions of the trace and discard regular ones. This approach is based on anomaly detection techniques. Our detailed experiments have shown that our approach has a good anomaly detection performance, and can reduce the size of an output trace by an order of magnitude [24]. Serge Emteu successfully defended his PhD about this work on the 15/12/2015 [5].

7.3. Causal reasoning and argumentation

Participants: Philippe Besnard, Louis Bonneau de Beaufort, Marie-Odile Cordier, Yves Moinard.

7.3.1. Searching for explanations from causal relations and ontology for argumentation

We have continued our work on reasoning (precisely search for explanations) from causal relations and ontology [48]. We resort to a well-known model [49] in computational argumentation in order to provide some structure to the collection of potential explanations given by our causal formalism. We have developed a case study, namely the Xynthia storm case, (February 2010, western France, trial September 2014) for which there exists a huge amount of data from various official reports. We have implemented an ASP program which thereby provides another application, besides those already mentioned: mining and landscape simulation, for ASP.

7.3.2. Cognitive maps and Bayesian causal maps

Cognitive map is a qualitative decision model which is frequently used in social science and decision making applications. This model allows to easily organize individuals' judgments, thinking or beliefs about a given problem in a graphical representation containing different concepts and influences between them. However, cognitive maps cannot model uncertainty within the variables and provides only deductive reasoning (predicting an effect given a cause). In [37], we show how to translate the knowledge represented in cognitive maps in the form of arguments and attack relations among them. Given a decision problem, we propose to build, first, a cognitive map by eliciting knowledge from experts and then to transform it into a weighted argumentation framework (WAF for short) for ensuring efficient reasoning. Another contribution concerns enriching the WAF obtained from a given cognitive map for dealing with dynamics through the consideration of a varying set of observations.

Cognitive maps and Bayesian networks are useful formalisms to address knowledge representation. Cognitive maps are powerful graphical models for gathering or displaying knowledge but while offering an easy means to express individuals judgments, drawing inferences remains a difficult task. Bayesian networks are widely used for decision making processes that face uncertain information or diagnosis but are difficult to elicitate. To take advantage of both formalisms and to overcome their drawbacks, Bayesian causal maps (BCM) were developed [75]. In [6], we propose to start from a causal map to construct the model and then set the conditional probabilities. Once the common causal map (CM) is built we can transform it into a BCM which combines causal modeling techniques and bayesian probability theory. We have developed a complete framework and applied it on a real problem in an environmental context. The implemented decision facilitating tool enables the representation of different shellfish dredgers views about their activity as well as the test of different fishery management scenarios.

EXMO Project-Team

7. New Results

7.1. Ontology matching and alignments

We pursue our work on ontology matching and alignment support [4] with contributions to evaluation and the use of algebras of relations within alignments.

7.1.1. 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 also handed out the organisation of OAEI 2015 to Ernesto Jiménez Ruiz (University of Oxford). We used again our generator for generating 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 10th Ontology Matching workshop [13], held Bethleem (PA US). More information on OAEI can be found at http://oaei.ontologymatching.org/.

7.1.2. Algebras of alignment relations

Participants: Armen Inants [Correspondent], Jérôme Euzenat.

Qualitative calculi are central in qualitative binary constraint satisfaction problems. All formalisms developed so far are homogeneous – they assume a single universe. We had previously shown the advantages of using a homogeneous qualitative calculus for expressing ontology alignment relations between concepts.

They make it possible to aggregate alignments disjunctively or conjunctively and to propagate alignments within a network of ontologies. The previously considered algebra of relations contains taxonomical relations between classes only. We have tackled the problem of combining two or more calculi over disjoint universes into a single calculus [9]. The problem is important because ontology matching deals with various kinds of ontological entities: concepts, individuals, properties. We have designed an algorithm for combining two homogeneous calculi with different universes into a single calculus. This has been applied to alignment relations [9] combining algebras for relations between concepts and individuals. It is, first, able to deal with empty classes, and, second, incorporates all qualitative taxonomical relations that occur between individuals and concepts, including the relations “is a” and “is not”. We have proved that this algebra is coherent with respect to the simple semantics of alignments.

The proposed algebras of relations and others have been integrated within the Alignment API (§6.1).

This work is part of the PhD of Armen Inants.

7.2. 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, both for the web of data and for the semantic web that it contributes to feed. We consider this problem from different perspectives.

7.2.1. Interlinking cross-lingual RDF data sets

Participants: Tatiana Lesnikova [Correspondent], Jérôme David, 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 investigated the second approach [10], finding that results were not as good as the translation approach. We also conducted evaluations on TheSoz, Agrovoc and Eurovoc thesauri.

This work is part of the PhD of Tatiana Lesnikova developed in the LINDICLE project (§9.1.1).

7.2.2. *An iterative import-by-query approach to data interlinking*

Participant: Manuel Atencia Arcas [Correspondent].

We modelled the problem of data interlinking as a reasoning problem on possibly decentralised data. We described an import-by-query algorithm that alternates steps of sub-query rewriting and of tailored querying of data sources [11]. It only imports data as specific as possible for inferring or contradicting target owl:sameAs assertions. Experiments conducted on a real-world dataset have demonstrated in practice the feasibility and usefulness of this approach for data interlinking and disambiguation purposes.

Additionally, and in line with the problem of dealing with uncertainty in linked data, we have proposed a probabilistic mechanism of trust that allow peers in a semantic peer-to-peer network to select the peers that are better suited to answer their queries, when query reformulation based on alignments may be unsatisfactory due to unsoundness or incompleteness of alignments [5].

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

7.2.3. *Link key extraction*

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

Ontologies do not necessarily come with key descriptions, and never with link key assertions (§3.3). Keys can be extracted from data by assuming that keys holding for specific data sets, may hold universally.

Following the work of last year on link key extraction [1] and the characterisation of the approach in formal concept analysis, we have fully characterised the results of our algorithm as formal concepts. We have also plans for extending both the approach and its formal concept analysis description through (i) applying it to full link keys as described in §3.3 , (ii) applying it to join and hierarchical key extraction, and (iii) applying it to hierarchical key extraction.

This work has been developed partly in the LINDICLE project (§9.1.1). Formal concept analysis aspects are considered with Amedeo Napoli (Orpailleur, LORIA).

7.3. Dynamic aspects of networks of ontologies

Huge quantities of data described by ontologies and linked together are made available. These are generated in an independent manner by autonomous providers such as individuals or companies. They are heterogeneous and their joint exploitation requires connecting them.

However, data and knowledge have to evolve facing changes in what they represent, changes in the context in which they are used and connections to new data and knowledge sources. As their production and exchange are growing larger and more connected, their evolution is not anymore compatible with manual curation and maintenance. We work towards their continuous evolution as it is critical to their sustainability.

Two different approaches are currently explored.

7.3.1. *Evolution of ontology networks and linked data*

Participants: Adam Sanchez Ayte [Correspondent], Jérôme David, Jérôme Euzenat.

We are considering the global evolution of knowledge represented by interdependent ontologies, data, alignments and links. Our goal is to be able to maintain such a structure with respect to the processes which are involved in its construction: logical inference, ontology matching, link key extraction, link generation, etc.

Our initial work is focused on how data and ontology changes cause alignment evolution, in particular when the alignment have been produced through instance-based matching using links between data. In this regard, we are developing techniques for circumscribing the elements and relationships affected by the change as well as evaluating the need for change propagation, i.e, most of the time a simple change will not trigger link key recomputation (§7.2.3).

This work is part of the PhD thesis of Adam Sanchez Ayte developed in the LINDICLE project (§9.1.1).

7.3.2. Revision in networks of ontologies

Participant: Jérôme Euzenat [Correspondent].

We reconsidered the belief revision problem in the context of networks of ontologies (§3.2): given a set of ontologies connected by alignments, how to evolve them such that they account for new information. In networks of ontologies, inconsistency may come from two different sources: local inconsistency in a particular ontology or alignment, and global inconsistency between them. Belief revision is well-defined for dealing with ontologies; we have investigated how it can apply to networks of ontologies. We formulated revision postulates for alignments and networks of ontologies based on an abstraction of existing semantics of networks of ontologies. We showed that revision operators cannot be simply based on local revision operators on both ontologies and alignments. We adapted the partial meet revision framework to networks of ontologies and show that it indeed satisfies the revision postulates [7]. Finally, we considered strategies based on network characteristics for designing concrete revision operators.

GRAPHIK Project-Team

7. New Results

7.1. Ontology-Based Query Answering with Existential Rules

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

Since Meghyn Bienvenu joined the team very recently (September 2015), we only include here the work she did in collaboration with GraphIK members.

Ontolology-based query answering (and more generally *Ontology-Based Data Access*, *OBDA*) is a new 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. In this context, *existential rules* (also called Datalog+) have been proposed to represent the ontological component [42], [41]. This expressive formalism generalizes both description logics used in OBDA (such as $\mathcal{EL}$ and DL-Lite), which form the cores of so-called tractable profiles of the Semantic Web ontological language OWL2) and Datalog, the language of deductive databases. Since about five years, we have been studying the theoretical foundations of this framework (mainly concerning decidability and complexity) and developing associated algorithmic techniques. We have started the development of a platform dedicated to OBDA with existential rules (see section 6.3).

Before presenting this year' results, we recall the two classical ways of processing rules, namely forward chaining and backward chaining, also known as "materialization" and "query rewriting" in the OBDA setting. In forward chaining, the rules are applied to enrich the initial data and query answering can then be solved solved by evaluating the query against the "saturate" database (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 database (again, as in a classical database system). Since entailment is not decidable with general existential rules, both forward and backwards processes may not halt.

7.1.1. Embedding transitivity rules.

In recent years, many classes of existential rules have been exhibited for which CQ entailment is decidable. However, most of these classes cannot express transitivity of binary relations, a frequently used modelling construct. We began to investigate the issue of whether transitivity can be safely combined with decidable classes of existential rules. On the one hand, we obtained negative results, proving that transitivity is incompatible with many classes having finite chase, and with UCQ-reducible classes in general. Second, we showed that transitivity can be safely added to linear rules (a subclass of guarded rules, which generalizes the description logic DL-Lite$_R$) in the case of atomic CQs, and also for general CQs if we place a minor syntactic restriction on the rule set (only needed when predicate arity is strictly greater than 2). Finally, we pinpointed the combined and data complexities of query entailment over linear rules + transitivity.

▷ *IJCAI 2015* [22]

7.1.2. A generic algorithm for query reformulation.

We first designed and implemented a query reformulation algorithm that takes as input any set of existential rules and a UCQ q, and outputs a sound, minimal and complete UCQ-reformulation of q, whenever such a reformulation exists (i.e., when the set of existential rules is UCQ-reducible). The core operation, unification, relies on a special technique that we first developed for conceptual graphs ("piece-unification"). A noteworthy feature of the implemented unification is that it is able to process rules without decomposing their head into single atoms. Experiments showed that this feature has a very high impact on the efficiency of query reformulation in terms of running time.

This algorithm can be seen as an instantiation of a generic reformulation algorithm, parametrized by a reformulation operator. As a complementary work, we studied the properties that should be fulfilled by any reformulation operator to ensure the correctness and the termination of this generic algorithm and analyzed some known operators with respect to these properties.

▷ *Semantic Web Journal 2015* [15]

7.1.3. Optimization of query reformulation algorithms

Query reformulation techniques have the advantage of being independent from the data. However, a main bottleneck is that the size of the obtained query can be exponential in the size of the original query, hence the produced reformulation maybe not usable in practice (and the corresponding SQL query may not even be accepted by the RDBMS). To overcome this combinatorial explosion in practice, we made two proposals, which have in common to consider other forms of reformulation, while staying equivalent to UCQs in terms of expressivity.

We defined semi-conjunctive queries (SCQs), which are a syntactical extension of conjunctive queries allowing for internal disjunctions. Briefly, a union of SCQs can be encoded in a more compact form than a UCQ. We designed and implemented an algorithm called Compact, which computes a sound and complete reformulation of a UCQ in the form of a union of SCQs (USCQ). First experiments showed that USCQs are both very efficiently computable and (often) more efficiently evaluable than their equivalent UCQs.

We developed another solution, which starts from a simple observation: in practice, combinatorial explosion is mainly due to some very simple rules, which form the core of any ontology. These rules typically express concept and relation hierarchies, concept properties and relation signatures. We proposed a technique that consists in compiling these rules into a preorder on atoms and embedding this preorder into the reformulation process. This allows us to compute compact reformulations that can be considered as “pivotal” representations, in the sense that they can be easily translated into different kinds of queries that can be evaluated by different kinds of database systems (e.g., unfolded into a classical UCQ or a USCQ, processed as such on data saturated by the compilable rules, or transformed into a Datalog program). Experiments show that this technique leads to substantial gains in the query reformulation process, in terms of size and runtime, it scales on very large ontologies (several ten thousands of rules), and it is competitive w.r.t. other existing tools, including those tailored for more specific rules corresponding to DL-Lite ontologies. This technique has been implemented in the sofware platform Graal.

▷ *IJCAI 2015*[28], *RuleML 2015*[23]

7.1.4. Ontology-based query answering with Semantic Web languages

On the one hand, we proposed Deductive RDF Triplestores, which are RDF knowledge bases equipped with Datalog inference rules. This work was developed in the context of the tool MyCorporisFabrica http://www.mycorporisfabrica.org/, an ontology-based tool for querying complex anatomical models.

In particular, we studied how to extract modules from deductive RDF triplestores. Indeed, many ontologies are extremely large, while users often need to reuse only a small part of resources in their work. A module is a Deductive RDF Triplestore entailed from the reference knowledge base, which is defined upon a restricted vocabulary. We proposed a new semantics for bounded-level modules allowing one to control their size, and then presented extraction algorithms compliant with the novel semantics.

▷ *AAAI 2015*[30] and *Journal of Biomedical Semantics* [16]. In collaboration with Marie-Christine Rousset (U. of Grenoble) and MyCorporisFabrica’s team.

On the other hand, in the context of the Graal platform, we defined a translation from the Semantic Web Ontological Language OWL 2 to our existential rule format. This gave rise to the definition of the “existential rule” OWL 2 profile, which covers the so-called tractable profiles of OWL 2 (see Section 6.3).

▷ *RuleML challenge*[33] (this paper obtained the RuleML 2015 challenge award)

7.2. Reasoning with Imperfect Information and Priorities

Participants: Abdallah Arioua, Patrice Buche, Madalina Croitoru, Jérôme Fortin, Nouredine Tamani, Rallou Thomopoulos.

This year, we mainly explored the use of argumentation frameworks in practical applications. Indeed, we have been involved in three main projects that have employed argumentation techniques. The projects were all in the context of agronomy where the nature of the problem studies fits well the use of argumentation: (1) the knowledge bases considered to model the domain are inconsistent, (2) the reasoning / decision making has to take into account the inconsistency, (3) the end user is a non-computing expert thus explanation facilities are needed.

We enumerate below the three projects and explain our results:

- Bread Making project (financed by the Food and Bioproducts department at INRA) investigates the possibility of using wholemeal flour in bread as opposed to classic white flour. The main theoretical result that we exploited here was the instantiation work into the existential rule framework done with Srdjan Vesic [43]. We used reverse engineering and the subsequent logic-based argumentation in order to provide the experts with a cartography of possible pros and cons of using one type of flour vs the other.

  ▷ *Ecological Informatics 2015* [18]
- EcoBioCap (FP7 EU project led by INRA Montpellier, see Section 9.1) investigates the conception of biodegradable packaging for fruits and vegetables. The main theoretical result used here concerns the fuzzy aspects of argumentation but the modeling of the problem using argumentation and subsequent argument elicitation was also a very challenging process.

  ▷ *Computers and Electronics in Agriculture 2015* (2 papers) [14] [17]
- DURDUR (ANR project led by INRA Montpellier, see Section 9.1) investigates the technological itineraries to grow durum wheat for subsequent pasta and couscous making. This ongoing project investigates the use of argumentation for explanation facilities.

  ▷ Initial results have been published in *SUM 2015*[19] and *DEXA 2015* [20].

7.3. Quality and interoperability of large document catalogues

Participants: Michel Chein, Madalina Croitoru, Alain Gutierrez, Michel Leclère, Rallou Thomopoulos.

The work in this research line takes place in the ANR project Qualinca, devoted to methods and tools to repair linkage errors in bibliographical databases (see Qualinca in Section 9.1). 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).

This year, we devoted most of our research effort to the following aspects in collaboration with ABES:

1. the finalization of a conceptual model of ABES librarian expertise in their linkage activity, and its formalization in our theoretical framework; the formalized model is both logical (the knowledge is expressed by facts, rules and constraints in first-order logic) and numerical (some predicates, which correspond to qualitative criteria, are computed by numerical functions, which themselves take as input the result of logical queries to the knowledge base).
2. the development of a diagnosis prototype, called SudoQual, which implements this model; in brief, SudoQual takes as input a given apellation (i.e., family name and first name), retrieves all references potentially associated with this appellation and outputs *sameAs* and *Different* links between these references. To develop SudoQual, we built an API on top of our tool Cogui.
3. first experiments with SudoQual on the Sudoc base, with the results being checked manually by ABES librarians.

▷ *Research report* [37]

This work required a tight collaboration with ABES (materialized by bimonthly meetings and numerous ponctual exchanges). The first experiments yield extremely satisfactory results, hence ABES is now considering turning SudoQual into a production tool used by librarians in their daily work to validate/correct autority links in the Sudoc catalogue. This requires to define a suitable user-interface, which is an issue we are currently discussing with ABES. We are also preparing experiments at a larger scale on a sample provided by ABES.

Besides, in collaboration with Qualinca partner LRI, we developed a method and a tool to fusion data linked by "same-as" links. More precisely, given an RDF dataset, our tool allows to merge "same as" data, which are often conflictual, into a unified and consistant representation using a multi-criteria decision method. The tool was evaluated on a dataset provided by INA and LIG, two other partners of Qualinca.

▷ *EGC 2015* [32]

Still with the LRI partner, who developed a logic-based decision tool that statuates on the validity of same-as links in RDF data, we investigated the use of argumentation techniques to explain why "same-as" links are invalidated by this tool.

▷ *SUM 2015* [25]

LINKS Team

7. New Results

7.1. Querying Heterogeneous Linked Data

7.1.1. Recursive queries

P. Bourhis published a paper at IJCAI [17] in cooperation with the University of Dresden in Germany. There he developed a highly expressive Web query language of the Datalog family, for which static analysis problems such as query containment remain decidable.

In cooperation with Links' associated team in Oxford, P. Bourhis obtained an article at ACM TODS [5], where he studies the access of hidden data by recursive queries.

V. Hugot, A. Boiret, and J. Niehren study monadic second-order logic for unordered trees with data constraints on siblings. This language can be used to define recursive queries and schemas on unordered data trees [13]. They study restrictions of the logics, for which the usual static analysis problems become decidable, and study the complexity of the decidable cases. This work was done in cooperation with Paris 7.

7.1.2. Schemas

I. Boneva and S. Staworko contribute at ICDT the RDF schema language SheX [22], which they developed in cooperation with members of the W3C. The usual open world approach of RDF is schemaless in that the alphabets of RDF data are left open, so that data from different sources and with different alphabets can be unified. This raises serious problems for query writing and thus for linked data integration, since a query may become invalid when the alphabet changes. A SheX schema allows to express constraints on the alphabets, node labels and edge labels of RDF graphs, so that database queries become safe with respect to future changes without closing the alphabet. In a previous work the studied the case of XML data trees instead of RDF graphs [6].

A. Lemay and J. Niehren propose sublinear algorithms in the style of probabilistic property testing for validating XML data trees with respect to DTD [20].

P. Bourhis studies streaming bounded repair with respect to schema violations [8]. This work is done in a cooperation with the University of Bordeaux and the University of Santiago in Chile.

7.1.3. Provenance

P. Bourhis obtained an ICALP paper [11] in cooperation with Télécom ParisTech. They show how to propagate provenance information for monadic second-order logics on trees or tree like structures with polynomial data complexity. In their provenance framework, they can show how to generalize various aggregation tasks for monadic second-order logics, that were known to be solvable with polynomal data complexity before.

In a cooperation with Tel Aviv, P. Bourhis obtained a ACM CIKM paper [18], where they show how to approximately summarize data provenance.

7.1.4. Data integration

In a cooperation with the University of Toronto, R. Ciucanu obtained a paper at PVLDB [4] on how to gain control over data integration evaluations. I. Boneva, A. Bonifati and R. Ciucaniu presented a paper on graph data exchange with target constraints [14] in the GraphQ workshop, and proved that query answering is intractable in this context.

7.2. Managing Dynamic Linked Data

7.2.1. Complex event processing

T. Sebastian, J. Niehren and D. Debarbieux propose early nested word automata for evaluating navigational XPath queries on XML streams [9]. They show how to approximate earliest query answering for such queries in a highly efficient manner and with very good precision in practice, while exact earliest query answering is known to untractable for XPath. This work was done in an industrial cooparation with Innovimax from Paris and in cooperation with the University of Bordeaux. In a follow-up work [21] they show that the `XPath` streaming algorithm for early nested word automata can be speed up considerably, when combining it with projection algorithms for nested word automata that they developed.

J. Niehren developed X-Fun [19] a uniform programming language for implementing XML standards, and showed how to implement XSLT, XProc, and XSLT in this manner. This work, that is fully implemented, was done in cooperation with the University of Bratislava.

7.2.2. Data-centered workflows

P. Bourhis presents highly expressive query languages as needed for data-centric workflows in the context of Active XML [3] in cooperation with the Dahu project from Inria Saclay.

J. Niehren presents a general framework for the reasoning with observational program semantics [10] in a cooperation with the Universities of Frankfurt and Saarbrücken in Germany.

7.3. Linking Data Graphs

S. Staworko obtained his HDR for his work on symbolic inference methods for databases [2]. R. Ciucanu obtained his PhD for his work on cross-model query inference [1] supervised bu A. Bonifati.

7.3.1. Learning path queries

A. Lemay, R. Ciucanu, and A. Bonifat have a paper and a demo at EDBT showing how to learn simple path queries on graph databases based on automata techniques [16], [15], [25], [24]. This is a very interesting starting point for using automata inference techniques in the context of graph databases.

S. Staworko obtained a paper at ICDT where he shows how to infer XML Twig queries from examples [23]. This work is done in cooperation with the University of Wrazlaw.

7.3.2. Learning join queries

R. Ciucanu, A. Boneva, and S. Staworko published an ACM TODS article [7] showing how to learn join queries for relational databases from examples. This is the first query learning algorithm satisfying Gold's learning model, that relies on equalities of data values rather than on structural information.

MAGNET Team

7. New Results

7.1. Decentralized Estimation in Networks

In [3], we studied the problem of decentralized estimation in networks, where each node of the network holds a data point and the goal is to estimate some statistics on the entire data under communication constraints imposed by the graph topology of the network. This generic problem has many applications in Internet of Things as well as for extracting knowledge from massive information graphs such as interlinked Web documents and online social media. In this work, we focused on estimating pairwise mean statistics. Popular examples of such statistics include the sample variance, the average distance and the Area Under the ROC Curve, among others. We proposed new synchronous and asynchronous randomized gossip algorithms which simultaneously propagate data across the network and maintain local estimates of the quantity of interest. We establish convergence rate bounds of $O(1/t)$ and $O(\log t/t)$ for the synchronous and asynchronous cases respectively, where t is the number of iterations, with explicit data and network dependent terms. Beyond favorable comparisons in terms of rate analysis, numerical experiments provide empirical evidence the proposed algorithms surpasses the previously introduced approach.

7.2. Large-Scale Learning with Higher-Order Risk Functionals

In [6], 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. In this work, we focused on how to best implement a stochastic approximation approach to solve such risk minimization problems in the large-scale setting. We argue that gradient estimates should be obtained by sampling tuples of data points with replacement (incomplete U-statistics) rather than sampling data points without replacement (complete U-statistics based on subsamples). We develop a theoretical framework accounting for the substantial impact of this strategy on the generalization ability of the prediction model returned by the Stochastic Gradient Descent (SGD) algorithm. It reveals that the method we promote achieves a much better trade-off between statistical accuracy and computational cost. Beyond the rate bound analysis, we provide strong empirical evidence of the superiority of the proposed approach on metric learning and ranking problems.

7.3. Natural Language Processing

In [4], we introduce a new structured model for learning anaphoricity detection and coreference resolution in a joint fashion. Specifically, we use a latent tree to represent the full coreference and anaphoric structure of a document at a global level, and we jointly learn the parameters of the two models using a version of the structured perceptron algorithm. Our joint structured model is further refined by the use of pairwise constraints which help the model to capture accurately certain patterns of coreference. Our experiments on the CoNLL-2012 English datasets show large improvements in both coreference resolution and anaphoricity detection, compared to various competing architectures. Our best coreference system obtains a CoNLL score of 81.97 on gold mentions, which is to date the best score reported on this setting.

In [2], we present a detailed comparative framework for assessing the usefulness of unsupervised word representations for identifying so-called implicit discourse relations. Specifically, we compare standard one-hot word pair representations against low-dimensional ones based on Brown clusters and word embeddings. We also consider various word vector combination schemes for deriving discourse segment representations from word vectors, and compare representations based either on all words or limited to head words. Our main finding is that denser representations systematically outperform sparser ones and give state-of-the-art performance or above without the need for additional hand-crafted features.

7.4. Some Ongoing Work

7.4.1. Metric Learning for Graph-based Label Propagation

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 helps define the weight of the connection between entities. The typical choice for this metric is usually a distance 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 most appropriate to solve the task efficiently.

In a paper currently under review, we proposed an algorithm that aims at learning the most appropriate vectorial representation for building a graph on which label propagation is solved efficiently, with theoretical guarantees on the classification performance.

7.4.2. Link Classification in Signed Graphs

We worked on active link classification in signed graphs. Namely, the idea is to build a spanning tree of the graph and query all its edge signs. In the two clusters case, this allows to predict the sign of an edge between nodes u and v as the product of the signs of edge along the path in the spanning tree from u to v. It turns out that ensuring low error rate amounts to minimizing the stretch, a long open standing problem known as Low Stretch Spanning Tree [11]. While we are still working on the theoretical analysis, experimental results showed that our construction is generally competitive with a simple yet efficient baseline and outperforms it for specific graph geometry like grid graphs.

Moreover, based on experimental observations, we will also analyze a heuristic which exhibits good performance at a very low computational cost and is therefore well suited for large-scale graphs. In a nutshell, it predicts the sign of an edge from u to v based on the fraction of u negative outgoing edges and v negative incoming edges, exploiting a behavioral consistency bias from signed social network users.

Going further in link classification, we believe that the notion of sign can be extended, going from one binary label per edge to a more holistic approach where the similarity between two nodes is measured across different contexts. These contexts are represented by vectors whose dimension matches the dimension of unknown feature vectors associated with each node. The goal is to answer queries of the form: how similar are nodes u and v along a specific context? We first plan to validate the relevance of this modeling on real-world problems, then test baseline methods on synthetic and real data before looking for a more effective, online prediction method.

7.4.3. Graph-based Learning for Dependency Parsing

We are investigating the use of different graph-based learning techniques such as k-nearest neighbors classification and label propagation for the problem of dependency parsing. While most of current approaches rely on learning a single scoring model (through SVM, MIRA, neural networks) from a large set of hand annotated training data (usually thousands of sentences), we are interested in using the sentence space geometry (approximated via a similarity graph over some labeled and unlabeled sentences) to tune the model to better fit a given sentence. This amounts to learning a slightly different model for each unlabeled sentence.

In order to successfully parse sentences in this setting, we need to propagate parsing information from labeled sentences to unlabeled ones through the graph. In order to build a similarity graph well suited to dependency parsing, we worked on learning a similarity function between pairs of sentences, based on the idea that two sentences are similar if they have similar parse trees. We will then investigate how to propagate the trees (which may be of varying sizes) through the graph and consider several propagation schemes.

OAK Project-Team

7. New Results

7.1. Scalable and Expressive Techniques for the Semantic Web

On the topic of efficient query answering methods for semantic-rich RDF data, we have obtained new fundamental results for the RDF Schema ontology language [25] and for a simple DL-Lite dialect [23], [34]; we presented our results in a tutorial at IEEE ICDE [10] and in an invited keynote at SEBD, the Italian Database conference [4]. A demonstration issued from this work was presented at VLDB [26] and at BDA, the French database conference [27].

To help users get acquainted with large and complex RDF graphs, we have started to work on an approach for RDF graph summarization: a graph summary is a smaller RDF graph, often by several orders of magnitude, which preserves the core structural information of the original graph and thus allows to reason about several important graph property on a much more manageable structure. Our first results were presented in [17] and demonstrated at [29] and [30]. These results were also presented in the keynote of the Data Engineering and the Semantic Web workshop [5].

On the related topic of analytical RDF schemas, we have published novel techniques for incrementally computing the result of an RDF analytical query (also known as “RDF cube”) out of the result of a previously computed RDF cube [31]. Such computations, commonly known as roll-up, drill-down etc. in the classical relational database setting, require novel solutions for RDF due to the heterogeneity of the graph structure.

7.2. Massively Distributed Data Management Systems

One of the main results of the year is the publication of the full paper [15] and demonstration [14] on CliqueSquare in the highly prestigious IEEE Conference on Data Engineering (ICDE). CliqueSquare has also been released in open source in 2014 (see the Software section). Its main advantage is a novel technique for optimizing conjunctive queries in a massively parallel setting, using n-ary join operators; this allow the optimization algorithm to build plans which are as flat as possible. These results apply beyond the RDF conjunctive query evaluation to the general setting of relational conjunctive query processing in a massively parallel context.

Another crucial result of the year is the publication of the PAXQuery framework for massively processing XML queries based on the Stratosphere (now Apache Flink) platform [3]. We show that our algebra-based approach allows to capture the expressive processing performed by an XQuery query and to compile it efficiently into massively distributed plans which are then evaluated by the Flink platform; this outperforms a set of state-of-the-art approaches for evaluating XQuery queries in a parallel environment. The system was also demonstrated at SIGMOD [11].

7.3. Advanced Algorithms for Data Querying and Transformation

We focused on explaining why some data, so-called missing answers, are not part of the result of a query, even though a developer expects them to be there.

The query-based explanations we return during query analysis serve as the starting point for our query rewriting process. Indeed, knowing the condition combinations pruning data relevant to the missing answers significantly narrows the search space for eligible query rewritings as we can first focus on finding solutions that only affect these query conditions. To further prune the search space, our current solution applies a cost model for rewritings based on several criteria, including edit distance to the original query, or the number of side-effects (tuples additionally appearing in the result of the rewritten query that are not our original missing answers). To select the best solutions w.r.t. the different dimensions of our cost model, we compute and return the skyline over these. We have demonstrated a preliminary version of the proposed algorithn in [8]. This work is reported [7] and in the PhD thesis of K. Tzompanaki [1].

7.4. Social Data Management and Crowdsourcing

Some particular tasks such as annotating data or matching entities have traditionnally been outsourced to human workers for many years. But the last few years have seen the rise of a new research field called crowdsourcing that aims at delegating a wide range of tasks to human workers. Crowd workers tend to make mistakes, so that redundant tasks are typically submitted to mitigate errors. As the crowd is a relatively expansive resource, we have worked on building formal frameworks to improve the efficiency of these processes.

Our research has been focused on two kinds of queries: boolean queries (asking the crowd to identify relevant items in a list, e.g., meals containing a specific ingredient), and ranking queries (asking the crowd to retrieve one or a few preferred items; e.g., ski resorts). We proposed new algorithms and heuristics improving the state of the art for boolean queries, and claimed the first algorithms for ranking queries (more specifically, for top-k and skyline queries) in the comparison framework [16].

We considered top-k query answering in social tagging systems, also known as folksonomies, a problem that requires a significant departure from existing, socially agnostic techniques. In a network-aware context, one can and should exploit the social links, which can indicate how users relate to the seeker and how much weight their tagging actions should have in the result build-up. Beyond explicit social links, we also focus uncovering implicit, potentially richer relationships from user interactions and exploiting them to improve core functionality such as search. Specifically we considered as-you-type search in a social network, where results socially close to the user asking the query are more relevant, and proposed an efficient algorithm presenting, for any (increasingly longer) prefix of the query as the user types it, the k most relevant results [28].

ORPAILLEUR Project-Team

7. New Results

7.1. The Mining of Complex Data

Participants: Mehwish Alam, Aleksey Buzmakov, Victor Codocedo, Miguel Couceiro, Adrien Coulet, Esther Galbrun, Nicolas Jay, Florence Le Ber, Luis-Felipe Melo, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, My Thao Tang, Yannick Toussaint.

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

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 [92]) allow to build 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 [101] and can play an important role in text classification and text mining. Following this way, and regarding itemset and association rule discovery, we improved standard algorithms for building lattices from large data and for completing the algorithm collection of the Coron platform [103].

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 where the concept lattice is considered as an index space for answering disjunctive queries [54]. We developed also a whole line of work on pattern structures for the discovery of functional dependencies [80], text classification and heterogeneous pattern structures [83], and pattern structures for structured attribute sets [46]. FCA can also be considered as a clustering method and we adapted pattern structures to clustering for analyzing numerical datatables supporting recommendation problems [13]. Projections can be associated with pattern structures for leveraging the volume and the complexity of the computation [53]. We designed also a quasi-polynomial algorithm for mining top patterns w.r.t. measures satisfying special properties in a FCA framework [52]. We also proposed new visualization techniques and tools able to display important and useful information (e.g. stable concepts) from large concept lattices [49].

Still considering complex data, we worked on the analysis of molecular structures (or molecular graphs) [34]. The mining of molecular graphs is an important task for many reasons, among which the challenges it represents regarding knowledge discovery, life sciences and healthcare, and, as well, the industrial needs that can be met whenever substantial results are obtained (especially in pharmacology).

7.1.2. Text Mining

Ontologies help software and human agents to communicate by providing shared and common domain knowledge, and by supporting various tasks, e.g. problem-solving and information retrieval. In practice, building an ontology or at least “ontological concept definitions” depends on a number of ontological resources having different types: thesaurus, dictionaries, texts, databases, and ontologies themselves. We are currently working on the design of a methodology based on FCA and RCA for ontology engineering from heterogeneous ontological resources. This methodology is based on both FCA and RCA, and was previously successfully applied in domains such as astronomy and biology.

In the framework of the ANR Hybride project (see 8.2.1.2), an engineer is implementing a robust system based on these previous research results, for preparing the way to new research directions involving trees and graphs. Moreover, we led a first successful experiment on extracting drug-drug interactions applying “lazy pattern structure classification” to syntactic trees [66]. In addition, in his thesis work, Mohsen Sayed focused on extracting relations between named entities using graph mining methods applied to dependency graphs. We are currently investigating how this approach can be generalized, i.e. how to detect a relation between complex expressions which are not previously recognized as named entities [64].

The notion of “Jumping Emerging Patterns” (JEP) previously used in chemistry [12], was updated and adapted to the context of text mining within the ANR Termith project. The objective is to design a learning method for filtering candidate terms within a full text and to decide whether an occurrence should be tagged as a term, i.e. as a positive example, or as a simple word, i.e. as a negative example. The method extracts from a training set all JEPs which are considered as hypotheses [7]. To reduce the number of JEPs and to only retain the most significant from a linguistic point of view, JEPs are weighted and a constraint solver is used to check the maximal coverage of the positive examples. Results are currently under evaluation.

7.1.3. Mining Sequences and Trajectories

Sequence data is widely used in many applications. Computing the similarity between sequences is a very important challenge for many different data mining tasks. There is a plethora of similarity measures for sequences in the literature, most of them being designed for sequences of items. In a recent work with Elias Egho, we study the problem of measuring the similarity between sequences of itemsets [32]. We focus on the notion of common subsequences as a way to measure similarity between a pair of sequences composed of a list of itemsets. In this work, we present new combinatorial results for efficiently counting distinct and common subsequences. These theoretical results are the cornerstone of an effective dynamic programming approach to deal with this problem. In addition, we develop an approximate method to speed up the computation process for long sequences. We have applied the method to various data sets: healthcare trajectories, on-line handwritten characters and synthetic data. The results confirm that the current similarity measure produces competitive scores and indicate that the method is relevant for large scale sequential data analysis.

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. In a recent work, we focused on the analysis of “complex sequential data” by means of interesting sequential patterns [19]. 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. We show how pattern structures along with projections, i.e. a data reduction of sequential structures, are able to enumerate more meaningful patterns and increase the computing efficiency of the approach. Finally, we demonstrate the applicability of the method for discovering and analyzing patient patterns from a French healthcare data set on cancer. The quantitative and qualitative results –with annotations and analysis from a physician– are reported in this use case which is one main motivation for this work.

7.1.4. Mining with Preferences

In the last decade, the pattern mining community has witnessed a sharp shift from efficiency-based approaches to methods which can extract more meaningful patterns. Recently, new methods adapting results from studies of economic efficiency and multi-criteria decision analysis such as Pareto efficiency, or skylines, have been studied. Within pattern mining, this novel line of research allows the easy expression of preferences according to a dominance relation. We have developed approaches that are useful from a user-preference point of view, tending to promote the use of pattern mining algorithms for non-experts. These approaches are based on the discovery of skyline patterns, or “skypatterns”, in relation with condensed representations of patterns. This last relationship facilitates the computation of skypatterns, providing a flexible and efficient approach to mine skypatterns reusing a dynamic constraint satisfaction problems (CSP) framework [8].

7.1.5. Aggregation

Aggregation or consensus theory studies any process dealing the merging of several objects (numerical values, qualitative data, preferences, etc.) into a single (or several) object of similar type and that, in some way, is the best representation. The need to aggregate objects in a meaningful way has become more and more present in an increasing number of areas not only of mathematics, statistics or physics, but especially in applied fields such as engineering, computer science, social sciences and biology. In social choice and multicriteria decision aid, objects are preferences that are expressed by users, voters or criteria, and are modeled by order relations or utility functions. In cluster analysis, the objects to merge are classifications (such as partitions, hierarchies or trees) or related functions (such as similarity/dissimilarity measures).

With the proliferation of massive databases and new fields such as computational advertising, search engines and recommender systems, the need for information retrieval and knowledge discovery processes became emergent as well as the construction of user preference models for classification and prediction purposes. Also in biology and phylogenetics, aggregation is used to find consensus patterns among DNA sequences or finding consensus trees within taxonomies. As algorithms are often heuristic in such large datasets, they rarely produce the same output, highlighting the importance of finding means of aggregation to produce consensus structures. The difficulty in extracting such consensus structures comes down to define appropriate aggregation rules (e.g., counting and median procedures), and their impossibility is many times revealed by Arrowian results. A way to avoid such impossibility results is the consideration of alternative aggregation rules or the weakening of underlying structures, for instance weak hierarchies that allow overlapping clusters while keeping desirable tree-like properties.

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. Moreover, Arrowian results are also envisioned since they constitute an important tool in the identification of reasonable algebraic/relational structures for representing data as well as in the identification of meaningful aggregation processes.

Direct applications of this theory are preference learning and cluster analysis. In the first case, preferences are represented by global utility functions and alternatives with higher utilities are preferred. Moreover, simplified versions of this model will be explored in the context of feature selection for both dimension reduction of data as well as classifier design. In the second case, we consider median structures that include several ordered/relational structures (trees, graphs, orders) and that allow several consensus procedures. This is particularly useful in a context of classification that takes into account evolutionary relations between classes, for instance, in taxonomical biology and phylogenetics.

7.1.6. Video Game Analytics

The video game industry has enormously grown over the last twenty years, bringing new challenges to the artificial intelligence and data analysis communities. We are studying the automatic discovery of strategies in real-time strategy games through pattern mining. Such patterns are the basic units for many tasks such as automated agent design, but also to build tools for the professionally played video games in the electronic sports scene. Continuing our joint collaboration with researchers from the MIT GameLab we successfully extended our previous work to a journal paper that will be published in 2016.

7.2. Knowledge Discovery in Healthcare and Life Sciences

Participants: Miguel Couceiro, Adrien Coulet, 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 the use of bio-ontologies to improve KDDK, and as well information retrieval, access to "Linked Open Data" (LOD) and data integration.

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 [99]. 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. 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 used only molecular networks databases or biomedical literature. We are studying and working on a novel method that take advantage of an eclectic set of linked data sources to validate uncertain drug–gene relationships, i.e. pharmacogenes [3]. 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 proposed an initial selection of linked data sources relevant to pharmacogenomics. We formatted these data to train a random forest algorithm, producing a model that classify drug–gene pairs as related or not, thus validating candidate pharmacogenes.

With this same motivation of validating state-of-the-art knowledge in pharmacogenomics, a new ANR project called "PractiKPharma" will be initiated in 2016 and will rely on similar ideas. The originality of "PractiKPharma" is to use "Electronic Health Records" to constitute cohorts of patients that are then mined for validating extracted pharmacogenomics knowledge units (http://practikpharma.loria.fr/).

7.2.3. Biological Data Aggregation for Knowledge Discovery

During this year, in collaboration with the Capsid Team, we contributed to write up two multi-disciplinary projects with a group of clinicians from the Regional University Hospital (CHU Nancy) and bio-statisticians from the Maths Lab (IECL). The first project, entitled ITM2P [0] lying in the so-called CPER 2015–2020 framework, was accepted and granted. The funding is mainly intended for medical and computing equipments and will be used to set up four scientific platforms. We are involved in 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) and was accepted as a so-called "investissement d'avenir" and granted. We are in charge of a workpackage which will give us the opportunity of exploring important research questions. Among these questions, one is to define "data aggregation" mechanisms with a twofold objective: (i) the definition of pairwise patient similarity given that patients are described by complex dimensions involving relations and time and (ii) the efficient clustering of patients based

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

on this similarity measure. 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. For doing that, we are currently investigating consensus theories [95] and their applicability to a bio-medical context, and as well aggregation operators as defined in various contexts, e.g. databases, data-warehouses, web of data, and graph theory. The idea is to consider relational and temporal data aggregation as a first class citizen in the data preparation phase of the knowledge discovery. This allows to assess the contribution of aggregation for such a task and in this context.

Another question is related to the construction of a prediction model for each bioprofile/subgroup –once validated by the clinicians– to be used in a decision support system. This will likely require the combination of symbolic and numerical methods for the classification task.

7.2.4. Analysis of biomedical data annotated with ontologies

Annotating data with concepts of an ontology is a common practice in the biomedical domain. Resulting annotations define links between data and ontologies that are key for data exchange, data integration and data analysis. Since 2011, we collaborate with the National Center for Biomedical Ontologies (NCBO) to develop a large repository of annotations named the NCBO Resource Index. This repository contains annotations of 36 biomedical databases annotated with concepts of more than 200 ontologies of the BioPortal (http://bioportal.bioontology.org/). In the preceding years, we compared the annotations of a database of biomedical publications (Medline) with two databases of scientific funding (Crisp and ResearchCrossroads) to profile disease research. One main challenge is to mine these annotations.

As a first attempt, we adapted pattern structures to analyze the annotations of biomedical databases [85]. We considered annotated biomedical documents as objects and the corresponding annotations were classified according to various dimensions, i.e. a particular aspect of domain knowledge. The resulting classification of annotations allowed not only to discover correlations between annotations but also incomplete annotations that could be fixed afterward. This adaptation of pattern structures opens many perspectives in term of ontology reengineering and knowledge discovery.

7.3. Knowledge Engineering and Web of Data

Participants: Mehwish Alam, Aleksey Buzmakov, Victor Codocedo, Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer.

Keywords: knowledge engineering, web of data, 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) [84] (http://intoweb.loria.fr/taaable3ccc/). 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, minimal 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 to use 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 [61].

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 [62].

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 [4].

FCA allows to organize 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 [63].

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 engines included in the Revisor library (see § 6.4.5). This year, a mixed linear optimization has implemented to produce human easy understandable quantities. For example, when the ingredient is a lemon, its quantity will take the form of a quarter, a half, etc., instead of *54 g* (which corresponds to a half lemon) [63].

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 [1]. We have proposed a novel technique based on Formal Concept Analysis which organizes subsets of RDF data into a concept lattice. This allows data exploration as well as the discovery of implication rules which are used to automatically detect 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.

Other important aspects are concerned with data access, data visualization w.r.t. the SPARQL query language [46], [49]. SPARQL queries over the web of data usually produce lists of tuples as answers that may be voluminous and hard to interpret. We introduced Lattice-Based View Access (LBVA), a framework based on FCA, which provides a classification of the answers of SPARQL queries based on a concept lattice. This concept lattice can be considered as a materialized view of the data resulting from a SPARQL query and can be navigated for retrieving or mining specific patterns. We associate a VIEW-BY clause to SPARQL for facilitating the interaction between analysts and LOD. The organization of answers is based on an original proposition on pattern structures for structured sets of attributes, which appears to be quite efficient and very well-adapted to the classification and analysis of RDF data. The visualization and the navigation of the concept lattice are guided by RV-Xplorer (i.e. RDF View eXplorer), an adapted interactive visualization system. Experiments show that the approach is well-founded and that it opens many new perspectives in the domain.

7.4. Advances in Graph Theory

Participants: Miguel Couceiro, Amedeo Napoli, Chedy Raïssi, Jean-Sébastien Sereni, Mario Valencia.

Keywords: graph theory, extremal graph theory, chromatic number, triangle-free graph, planar graph, graph coloring

We announced in the last report that we started to work on a conjecture by Heckman and Thomas from 1999. We managed to confirm the conjecture and the demonstration was published in January 2014. A classical result by Staton, from 1979, states that every triangle-free graph G with maximum degree at most 3 contains an independent set of order at least $5n/14$, where n is the number of vertices of G. Heckman and Thomas conjectured a stronger fact: the fractional chromatic number of such a graph is at most $14/5$. We confirmed their conjecture by establishing the following stronger assertion: for any assignment of weights (i.e., real numbers) to the vertices of such a graph G, there exists an independent set I such that the weights of the vertices in I is at least $5/14$ times the total weight of the G.

Exploring further the methods we introduced to solve this conjecture, we obtained new results concerning the fractional chromatic number of planar triangle-free graphs. While the fractional chromatic number of such graphs is at most 3 (because their chromatic number is), a construction of Jones proved the existence of triangle-free planar graphs with fractional chromatic number arbitrarily close to 3. Thus one wonders whether there could be such graphs with fractional chromatic number exactly 3. We demonstrated this not to be the case, by proving a general upper bound of $\frac{9n}{3n+1} = 3(1 - \frac{1}{3n+1})$ for every triangle-free planar graph G with n vertices. This bound is qualitatively the best possible: Jones's construction yields graphs with fractional chromatic number $3 - \frac{c}{n}$ for some constant c. In addition, a tight bound was obtained if the graphs considered are furthermore required to have maximum degree at most 4. In this case, the bound becomes $\frac{3n}{3n+1}$.

Motivated by frequency assignment in office blocks, we study the chromatic number of the adjacency graph of a 3-dimensional parallelepiped arrangement. In the case each parallelepiped is within one floor, a direct application of the Four-Colour Theorem yields that the adjacency graph has chromatic number at most 8. We provide an example of such an arrangement needing exactly 8 colors. We also discuss bounds on the chromatic number of the adjacency graph of general arrangements of 3-dimensional parallelepipeds according to geometrical measures of the parallelepipeds (side length, total surface area or volume).

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. By combining this new index and the corresponding evaluation techniques, our embedded search engine is capable of reconciling high insert/delete/update rate and query scalability. We have demonstrated the search engine on a secure USB token in the context of a personal cloud, and have conducted in depth performance evaluations on a development board representative for different smart objects characteristics. The experimental results demonstrate the scalability of the approach and its superiority compared to state of the art methods. This work was published at VLDB'15 [21] and demonstrated at SIGMOD'15 [24]. It constitutes the main contribution of the PhD thesis of Saliha Lallali

Spatio-temporal indexing in Flash storage: The convergence of mobile computing, wireless communications and sensors has raised the development of many applications exploiting massive flows of spatio-temporal data such as in location-based services, participatory sensing, or traffic management [15]. Spatio-temporal data indexing is among the most active research topics in this area. Nevertheless, since a few years a new fundamental parameter has made its entry on the database scene: the NAND flash storage. The peculiar characteristics of flash memory require redesigning the existing data storage and indexing techniques that were devised for magnetic hard-disks. TRIFL, proposed in [16] is an efficient and generic TRajectory Index for FLash, designed around the key requirements of both trajectory indexing and flash storage. TRIFL is generic in the sense that it is efficient for both simple flash storage devices such as the SD cards and more powerful devices such as the solid state drives. In addition, TRIFL includes an online self tuning algorithm that allows adapting the index structure to the workload and the technical specifications of the flash storage device to maximize the index performance. Moreover, TRIFL achieves good performance with relatively low memory requirements, making it appropriate for many application scenarios. The experimental evaluation shows that TRIFL outperforms the representative indexing methods on flash disks but also on magnetic disks. This work [15] [16] is part of Dai Hai Ton That's Ph.D. thesis, co-supervised by Iulian Sandu Popa.

6.2. Secure Global Computing on Asymmetric Architecture

Participants: Benjamin Nguyen [correspondent], Philippe Pucheral, Quoc Cuong To.

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: 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 the Asymmetric Architecture, thus achieving secure and private computation. Our first contribution in this area was to study the execution of Privacy Preserving Data Publishing (PPDP) algorithms on such an architecture, and provided generic protocols to deal with all kinds of PPDP algorithms, which are robust against honest-but-curious and malicious adversaries [2][3]. Our second contribution was to study 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 [9]. This work was part of Quoc-Cuong To's Ph.D defended in 2015 [13]. We are extending this general framework through a collaboration with INSA Centre Val de Loire, LIFO Lab and University of Paris Nord, LIPN lab, to study the secure execution of Map/Reduce on the Asymmetric Architecture. Computing MapReduce processes on the Asymmetric Architecture means maintaining the flexibility and efficiency of MapReduce, while adding security into the mix. We have shown in [25] that it is possible to achieve seamless integration of distributed MapReduce processing using trusted cells, while maintaining reasonable performance.

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. 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, co-supervised by Iulian Sandu Popa. The system implementation was demonstrated in [26], and a paper describes the technical details of the system [31].

6.3. Personal Cloud

Participants: Nicolas Anciaux [correspondent], Luc Bouganim, Athanasia Katsouraki, Benjamin Nguyen, Philippe Pucheral, Iulian Sandu Popa, 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 [4]. 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 [1]. 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 and at ADBIS'15 [18]. 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 the PhD of Paul Tran Van (CIFRE SMIS-CozyCloud) has started to explore new data sharing techniques which could be enforced in the secure personal cloud model. A second PhD CIFRE SMIS-CozyCloud is being submitted to explore privacy-preserving distributed computations over personal clouds. Athanasia Katsouraki is working on privacy issues and on adoption of the secure data engine [29] in cooperation with the economists (CERDI) in the context of the Digital Society Institute (DSI). A paper written by jurists, economists and computer scientists from DSI has been invited for publication in Legicom'2016 to present our common vision of Privacy-by-Design principles in the context of Open Data and Internet of Things.

6.4. Applications

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

In 2014, we proposed a new paradigm, that we call Folk-enabled Information System (Folk-IS), based on a fully decentralized and participatory approach, where each individual implements a small subset of a complete information system without the need for a shared networked infrastructure [5]. Folk-IS builds upon the emergence of highly secure, portable and low-cost storage and computing devices, called hereafter Smart Tokens. Here however, the focus is on low-cost of ownership, deployment and maintenance, and on the absence of a networked infrastructure. With Folk-IS and thanks to their smart tokens, people will transparently and opportunistically perform data management and networking tasks as they physically move, so that IT services are truly delivered by the crowd. Following this work, we collaborate with researchers and doctors from Cameroon to study the specific case of diabete follow-up. Indeed, there are currently more than half a million diabetes cases in Cameroon and the deaths caused by diabetes complications will double before 2030. Diabetes complications mostly occur due to a bad follow-up of patients. Based on an analysis of the current situation, we proposed a new IT architecture for diabetes follow-up and introduce the bases of a new distributed computation protocol for this architecture. Our approach does not require any preexisting support communication infrastructure, can be deployed at low cost, and provides strong privacy and security guarantees. This work, published in AFRICOM [20] envisions an experiment in the field we plan to conduct under the authority of the Cameroonian National Center for Diabetes and Hypertension, with a potential for generalization to other diseases.

TYREX Project-Team

6. New Results

6.1. Expressive Logical Combinators

A popular technique for the analysis of web query languages relies on the translation of queries into logical formulas. These formulas are then solved for satisfiability using an off-the-shelf satisfiability solver. A critical aspect in this approach is the size of the obtained logical formula, since it constitutes a factor that affects the combined complexity of the global approach. In this work [21], we present logical combinators whose benefit is to provide an exponential gain in succinctness in terms of the size of the logical representation. This opens the way for solving a wide range of problems such as satisfiability and containment for expressive query languages in exponential-time, even though their direct formulation into the underlying logic results in an exponential blowup of the formula size, yielding an incorrectly presumed two-exponential time complexity. We illustrate this from a practical point of view on a few examples such as numerical occurrence constraints and tree frontier properties which are concrete problems found with semi-structured data [21].

6.2. Behavioural Types

Behavioural type systems ensure more than the usual safety guarantees of static analysis. They are based on the idea of “types-as-processes”, providing dedicated type algebras for particular properties, ranging from protocol compatibility to race-freedom, lock-freedom, or even responsiveness.

Two successful, although rather different, approaches, are session types and process types. The former allows to specify and verify (distributed) communication protocols using specific type (proof) systems; the latter allows to infer from a system specification a process abstraction on which it is simpler to verify properties, using a generic type (proof) system. What is the relationship between these approaches? Can the generic one subsume the specific one? At what price? And can the former be used as a compiler for the latter?

In [15], we showed how communication protocols can be integrated into an object-oriented type system supporting *non-uniform objects*, i. e. objects where the sequences of method calls are restricted, such as a `File` where `read()` cannot be called after `close()`. In such a system, communication protocols can be enforced by giving appropriate non-uniform types to the socket objects. We defined a sound and complete type checking algorithm for a small distributed class-based object-oriented language with structural subtyping. Static typing guarantees that both sequences of messages on channels, and sequences of method calls on objects, conform to type-theoretic specifications, thus ensuring type-safety.

6.3. SPARQL Queries

Static analysis is a core task in query optimization and knowledge base verification. In [14], [24], we study static analysis techniques for SPARQL, the standard language for querying Semantic Web data. We are interested in developing techniques through reductions to the validity problem in logic

In [22], we investigate techniques for detecting SPARQL query update independence. A query is independent of an update when the execution of the update does not affect the result of the query. Determining independence is especially useful in the context of huge RDF repositories, where it permits to avoid expensive yet useless re-evaluation of queries. While this problem has been intensively studied for fragments of relational calculus, very few works exist for the standard query language for the semantic web. We report on our investigations on how a notion of independence can be defined in the SPARQL context.

6.4. Semantic Subtyping

In a programming language, subtyping represents a notion of safe substitutability (it is always safe to replace a value of some type with a value of a subtype). There are several ways such a relation can be formally defined. Semantic subtyping consists of giving a set-theoretic denotation to types and using set inclusion to define subtyping. Works by Benzaken, Castagna, Frisch and Xu have described how to define such relations for complex type algebras comprising recursive, product, function, intersection, union, and complement types together with type variables. In [17], we showed how to formalise such a relation in logic and decide it in EXPTIME, answering an open question, and discussed experiments made with the full implementation of the system in our solver (5.3).

6.5. Spatio-temporal validation of multimedia documents

A multimedia document authoring system should provide analysis and validation tools that help authors find and correct mistakes before document deployment. Although very useful, multimedia validation tools are not often provided. Spatial validation of multimedia documents may be performed over the initial position of media items before presentation starts. However, such an approach does not lead to good results when media item placement changes over time. Some document authoring languages allow the definition of spatio-temporal relationships among media items and they can be moved or resized during runtime. Current validation approaches do not verify dynamic spatio-temporal relationships. In [19], we present a novel approach for spatio-temporal validation of multimedia documents. We model the document state, extending the Simple Hypermedia Model (SHM), comprising media item positioning during the whole document presentation. Mappings between document states represent time lapse or user interaction. We also define a set of atomic formulas upon which the author's expectations related to the spatio-temporal layout can be described and analyzed.

6.6. XQuery and Static Typing

XQuery is a functional language dedicated to XML data querying and manipulation. As opposed to other W3C-standardized languages for XML (e.g. XSLT), it has been intended to feature strong static typing. Currently, however, some expressions of the language cannot be statically typed with any precision.

In [20], we argue that this is due to a discrepancy between the semantics of the language and its type algebra. We discuss how to handle this discrepancy by improving the type system. We describe a logic-based language of extended types able to represent inner tree nodes and show how it can dramatically increase the precision of typing for navigation expressions. We describe how inclusion between these extended types and the classical regular tree types can be decided, allowing a hybrid system combining both type languages. The result is a net increase in precision of typing.

In a previous work, we aimed at bridging the gap between path-based XML processing languages like XQuery and pattern-based such languages like CDuce. We extend the language CDuce into a succinct core λ-calculus that captures XQuery 3.0. The extensions we consider essentially allow CDuce to implement XPath-like navigational expressions by pattern matching and precisely type them. The elaboration of XQuery 3.0 into the extended CDuce provides a formal semantics and a sound static type system for XQuery 3.0 programs.

6.7. Efficiently Deciding μ-calculus with Converse over Finite Trees

In [16], we present a sound and complete satisfiability-testing algorithm and its effective implementation for an alternation-free modal μ-calculus with converse, where formulas are cycle-free and are interpreted over finite ordered trees. The time complexity of the satisfiability-testing algorithm is $2^{O}(n)$ in terms of formula size n. The algorithm is implemented using symbolic techniques (BDD). We present crucial implementation techniques and heuristics that we used to make the algorithm as fast as possible in practice. Our implementation is detailed in 5.3 .

6.8. Reasoning with Style

The Cascading Style Sheets (CSS) language constitutes a key component of web applications. It offers a series of sophisticated features to stylize web pages. Its apparent simplicity and power are however counterbalanced by the difficulty of debugging and maintaining style sheets, tasks for which developers still lack appropriate tools. In particular, significant portions of CSS code become either useless or redundant, and tend to accumulate over time. The situation becomes even worse as more complex features are added to the CSS language (e.g. CSS3 powerful selectors). A direct consequence is a waste of CPU that is required to display web pages, as well as the significant amount of useless traffic at web scale. Style sheets are designed to operate on a set of documents (possibly generated). However, existing techniques consist in syntax validators, optimizers and runtime debuggers that operate in one particular document instance. As such, they do not provide guarantees concerning all web pages in CSS refactoring, such as preservation of the formatting. This is partly because they are essentially syntactic and do not take advantage of CSS semantics to detect redundancies. In [18], we propose a set of automated refactoring techniques aimed at removing redundant and inaccessible declarations and rules, without affecting the layout of any document to which the style sheet is applied. We implemented a prototype that has been extensively tested with popular web sites (such as Google Sites, CNN, Apple, etc.). We show that significant size reduction can be obtained while preserving the code readability and improving maintainability.

6.9. A Comparative Analysis of Attitude Estimation

We investigate the precision of attitude estimation techniques in the context of pedestrian dead-reckoning with commodity smartphones. We propose a comparative analysis of state-of-the-art algorithms for attitude estimation in this setting. We provide an experimental setup with a precise ground truth obtained with a motion capture system. We precisely quantify the error in attitude estimation obtained with each technique. We discuss the obtained results and analyse advantages and limitations of current technology for further PDR research.

WIMMICS Project-Team

7. New Results

7.1. Users modeling and designing interaction

7.1.1. Exploratory search

Participants: Emilie Palagi, Alain Giboin.

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. Exploratory search (ES) tasks are open-ended, multi-faceted, and iterative like learning or topic investigation [59]. 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. Two research actions were undertaken this year to contribute to achieve this goal.

7.1.1.1. Design of an exploratory-search-oriented protocol for testing an image search algorithm based on user's eye movements

Participants: Emilie Palagi, Alain Giboin.

This work was undertaken in the context of the VISIIR ANR project [0], leaded by the MinD team (I3S, UNS), with Stéphanie Lopez and Frédéric Precioso. One of the objectives of VISIIR is to design an interactive image search system based on user's eye movements. Detected by an eye-tracker, these movements allow the system to infer the image that the user is going to select; VISIIR's aim is to replace user's mouse clicks as a selection mode by implicit *eye-clicks*. Since ES behaviors can be observed in image search tasks, we designed an eye-tracking user test protocol on Discovery Hub [0] in order to: 1) increase our understanding of the ES process (at the cognitive and perceptual-motor levels); 2) verify if identified characteristics of gaze trajectories allow to infer the images that the user is going to select in ES tasks (as opposed to lookup search tasks).

7.1.1.2. 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.

(with Raphaël Troncy, Eurecom)

This work was undertaken in the context of the PhD of Emilie Palagi. In [41] we introduced our approach for designing a user-centered evaluation method for ESSs. Our method takes into account users's ES behavior and is based on a cognitive model of an ES task. We will specially work on Discovery Hub (Wimmics project – Inria) and 3cixty [0] (EURECOM project) ESSs.

7.1.2. Sentiment Analysis

Participant: Andrea Tettamanzi.

Together with Célia da Costa Pereira (I3S, UNS) and Mauro Dragoni of FBK, Trento, who visited our team for three months from April to June 2014, we have further refined our approach to concept-level sentiment analysis based on fuzzy logic [12].

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

Participants: Oscar Rodriguez Rocha, Catherine Faron-Zucker.

[0] http://www.agence-nationale-recherche.fr/?Projet=ANR-13-CORD-0009
[0] http://discoveryhub.co/
[0] https://www.3cixty.com/

In the framework of the Semantic Educloud project, we developed a Web ontology for the description and representation of serious games. Such ontology describes the functional and design elements of the game, the profile and virtual context of the players and furthermore the datasets from the Web of data that the game can query. The ontology has been evaluated through a prototype, which is basically a serious game quiz based on DBpedia. As future work, it is planned to implement state-of-the-art recommendation algorithms of Linked Data resources that take into account the context and players' profile. Furthermore an integration with the EDUCLOUD platform is considered: EDUCLOUD is an emerging initiative in Sophia Antipolis, for the implementation of a platform of digital educational content accessible through the cloud from a 3D portal of resources, and any interface devices (tablets, smartphones, PCs). [44]

7.2. Communities and social interactions analysis

7.2.1. Community Detection and Interest Labeling

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

7.2.1.1. Topic Modeling Based Overlapping Community Detection

Based on previous work, we conducted more experiments to evaluate the effectiveness and efficiency of the proposed tag tree based method. We used perplexity score to evaluate the performance of topic extraction. We got consistent performance when applying the model on a Flickr dataset. This work has been published in IEEE/WIC/ACM Web Intelligence 2015 [32] and Social Network Analysis and Mining Journal [13].

7.2.1.2. Temporal Analysis of User and Topic

By jointly modeling topic, expertise, time and activity, 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 based model by modeling the reverse distribution.

7.2.1.3. 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.

7.2.2. Semantic Modeling of Social, Spatiotemporal and Dedicated Networks

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

We have been working on modeling a multi-agent based recommender system. The aim of such system is to recommend a list of activities (plans) according to user preferences in order to achieve a goal. For this purpose, we propose a multi-context framework based on the well-known agent Belief-Desire-Intention (BDI) architecture [58]. First, we extend the BDI model with additional contexts in order to handle sociality. Second, we use a possibilistic approach based on the work of Da Costa Pereira & Tettamanzi [55], to reason about beliefs, desires, goals and intentions. Further, we use ontologies to represent and reason about plans and intentions. The proposed framework is detailed in a long paper that will be presented in the 8th International Conference on Agents and Artificial Intelligence in 2016 [18].

7.2.3. Collaborative Software Development Platforms

Participant: Isabelle Mirbel.

The collaborative nature of software development helped in the emergence of several online collaborative software development platforms (CSDPs). These platforms enable distributed teams of contributors to participate in the development of the various hosted projects. In such a context, the identification of relevant contributors is very important for handling efficiently the abundant requirements. However, this can be really challenging because of the fairly large number of involved contributors, especially in some distinguished projects. Moreover, the contributor profiles on a CSDP are often inadequately informative, which makes them an unqualified resource for learning about the contributors.

In this context, we proposed to identify contributors by their reputation on a CSDP. Our approach calculates reputation scores using a belief calculus, called subjective logic, according to contributors' performed roles. Knowing the reputation of anonymous contributors would enable project members to reduce the uncertainty in their future interactions with them. Moreover, we use concept lattices to classify contributors by their reputation scores, which enable us to have a comparable view on the considered contributors. Consequently, we can produce a roadmap to examine new requirements thus supporting their effective communication and prioritization.

7.2.4. Logical Foundations of Cognitive Agents

Participants: Andrea Tettamanzi, Serena Villata.

Together with Célia da Costa Pereira (I3S, UNS), we have continued an investigation about the issue of trust in multi-agent systems, and we proposed a computational model of trust based on the content of messages and on the characteristics of their sources [27].

7.2.5. Combining Argumentation Theory and Normative Reasoning with Natural Language Processing

Participants: Serena Villata, Elena Cabrio.

Together with Cristian Cardellino and Laura Alonso Alemany from the University of Cordoba (Argentina), we applied different Active Learning strategies to Information Extraction from licenses in English, with highly repetitive text, few annotated or unannotated examples available, and very fine precision needed. We showed that the most popular approach to active learning, i.e., uncertainty sampling for instance selection, does not provide a good performance in this setting. We showed that we can obtain a similar effect to that of density-based methods using uncertainty sampling, by just reversing the ranking criterion, and choosing the most certain instead of the most uncertain instances. The results of this research have been published at the CICLing [24] and JURIX [23] international conferences.

In another work, together with Alessio Palmero Aprosio (FBK Trento, Italy), we have worked on an extension of QAKiS, the system for open domain Question Answering over linked data, that allows to query DBpedia multilingual chapters. Such chapters can contain different information with respect to the english version, e.g. they provide more specificity on certain topics, or fill information gaps. In particular, we have introduced and evaluated the RADAR 2.0 framework for information reconciliation over language-specific DBpedia chapters. The framework is composed of three main modules: a module computing the confidence score of the sources depending either on the length of the related Wikipedia page or on the geographical characterization of the queried entity, a module retrieving the relations holding among the elements of the results set, and finally a module computing the reliability degree of such elements depending on the confidence assigned to the sources and the relations among them. This third module is based on bipolar argumentation theory to return the acceptability degrees. A demo of the RADAR framework is available online [0]. This contribution has been submitted to the Semantic Web Journal and is under review.

Moreover, we have proposed the BEGincs (BEG-Inconsistencies) framework, which translates a bipolar entailment graph into an argumentation graph. It then provides to the annotators sets of arguments that are supposed to be consistent. If it is not the case, the Textual Entailment system wrongly assigned some relations. Moving from single pairs to an overall graph allows for the detection of inconsistencies otherwise undiscovered. BEGincs does not identify the precise relation causing the inconsistency, but provides annotators with the consistent arguments sets, they are supported in narrowing the causes of inconsistency. The results of this research have been published at the CLIC conference [43].

7.2.6. Argumentation theory and its applications

Participants: Elena Cabrio, Serena Villata, Fabien Gandon, Andrea Tettamanzi.

[0] http://qakis.org/qakis2

Together with Celia da Costa Pereira (UNS), we have proposed a framework to measure the acceptability of an information in a multiagent system, according to *(i)* the agent's goals and the information source's goals, *(ii)* the credibility, for the agent, of the incoming information and *(iii)* the agent's beliefs (or perceptions) about the context (or situation) in which it operates. The results of this research have been published at the AAMAS international conference [27].

Moreover, together with Sahbi Benlamine, Maher Chaouachi and Claude Frasson (U. of Montreal) we have presented an empirical evaluation of the relationship between the argumentative structures of human debates and the emotions felt by the debate participants. Argumentation is often seen as a mechanism to support different forms of reasoning such that decision-making and persuasion, but all these approaches assume a purely rational behavior of the involved actors. However, humans are proved to behave differently, mixing rational and emotional attitudes to guide their actions, and it has been claimed that there exists a strong connection between the argumentation process and the emotions felt by people involved in such process. We assess this claim by means of an experiment: during several debates people's argumentation in plain English is connected and compared to the emotions automatically detected from the participants. The results of this research have been published at the IJCAI international conference [19], and submitted to the Cognitive Science journal (under review).

7.2.7. *Natural Language Argumentation on Twitter*

Participants: Tom Bosc, Elena Cabrio, Serena Villata.

A great amount of textual data is published on social media every day. For example, there are about 500 million new tweets per day on Twitter. These data reflect the opinion and thoughts of a large population and are thus potentially useful to decision-makers and marketers, among others. But processing them is challenging because of their large quantity as well as their noisiness : poor quality of writing, redundancy, presence of advertisement, etc.

The goal of this project is to build a pipeline to automatically analyze messages exchanged on Twitter and build informative and synthetic views. We study tweets under the angle of argumentation theory. First of all, the algorithm filters in argumentative tweets. Then, it describes how tweets relate to one another : tweets may support or attack other tweets, or be neutral. Finally, a visualisation of the interactions between tweets is produced. Individual parts of the pipeline are machine learning models that are trained using datasets that are crafted specially for the project. Importantly, datasets span several domains (politics, society topics, product announcements) to ensure that the approach is generic enough and will generalize to unseen topics.

7.3. Vocabularies, Semantic Web and linked data based knowledge representation

7.3.1. *SPARQL Template Transformation Language*

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

We designed and developed a generic software environment to generate Semantic Web Servers and Linked Data Navigators [0] on top of the STTL SPARQL Template Transformation Language. We designed STTL transformations from RDF to HTML that enable to set up hypertext Linked Data Navigators on local or remote (e.g. DBpedia) triple stores. This work was published at ISWC, WebIST, LNBIP and IC [26], [25], [45], [39].

We extended STTL in order to perform rule based constraint checking. Templates return boolean true (resp. false) when constraint cheking succeeds (resp. fails). We applied this extension on OWL profile conformance checking and we tested with success OWL RL, OWL EL and OWL QL profiles.

7.3.2. *SPARQL Function Language*

Participants: Olivier Corby, Catherine Faron-Zucker.

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

We started the design of a Function Language on top of SPARQL filter language. We added the `function` statement that enables users to define extension functions directly in the filter language. We added statements to the filter languages such as `let` local variables, `for` loop and `list` datatype and we integrated `select` and `construct` queries in the language. Extension functions are directly available into SPARQL queries. This solves the problem of extension function interoperability. We were able to design custom datatypes such as roman numbers, custom aggregates such as median and standard deviation, extension functions to compute the week day of a given date, approximate search functions, recursive functions with the service clause, etc. [50].

7.3.3. Graph Pattern Matching

Participants: Olivier Corby, Fuqi Song.

We proposed a heuristics-based query planning approach which allows reducing SPARQL query executing time. This approach has been developed and integrated to Corese plaform. The relevant work and results have been published at conference KES 2015 [35].

We developed a component that can improve the storage capacity of Corese software, generally speaking this approach stores large RDF literals into the file system instead of in memory. The experiments are performed based on the data set of BSBM [54] and the results suggested that with this component, it can save up to 40% RAM space without slowing down the query execution time.

We implemented and integrated similarity measurement algorithms to Corese software in order to enable approximate semantic search. The main objective is to return approximate results when there are no results in the data source corresponding to the query.

7.3.4. Dynamic Application Scheme Composition

Participant: Isabelle Mirbel.

Dynamic service composition has emerged as a promising approach to build complex runtime-adaptable applications. In this context, new approaches for bottom-up opportunistic assembly of services have emerged. However, these approaches may lead to meaningless and useless compositions. Therefore, we advocate an approach in which bottom-up discovery of services is coupled with top-down user's requirements elicitation.

In our approach, application schemes publish available behaviors from basic component assembly. Our user's requirements elicitation framework, based on previous work, offers the capability to capture high-level end-user's requirements in an iterative and incremental way and to turn them into queries to retrieve application scheme's descriptions. We adopt semantic Web languages and models as a unified framework to deal with end-user's requirements and application scheme's descriptions in order to take advantage of their reasoning and traceability capabilities. We extended previous work about requirement's modeling by providing means to represent and reason on AND and OR operators as well as contextual data. Moreover, relying on the STTL language (see Section 7.3.1 , we proposed two transformations for runtime composition: the first transformation aims at detecting the possible compositions with regards to the available applications schemes; the second one aims at building a BPMN modeling to achieve user's requirements.

7.3.5. Semantic Web Languages And Techniques for Digital Humanities

Participants: Serena Villata, Elena Cabrio, Catherine Faron-Zucker, Franck Michel.

In the framework of the Zoomathia project, we conducted three complementary works. Their results have been published at the SW4SH international workshop [22][37][38]. First, together with Cécile Callou, Chloé Martin and Johan Montagnat (UNS), we started working on the construction of a thesaurus to support multi-disciplinary studies on the transmission of zoological knowledge throughout historical periods, combining the analysis of ancient literature, iconographic and archaeozoological resources. We constructed a SKOS thesaurus based on the TAXREF taxonomical reference designed to support studies in Conservation Biology.

Second, together with Molka Tounsi (UNS), and Arnaud Zucker (UNS), we have introduced a methodology to (i) extract pertinent knowledge from medieval texts using Natural Language Processing methods, (ii) semantically enrich semi-structured zoological data and publishing it as an RDF dataset and its vocabulary, linked to other relevant Linked Data sources, and (iii) reason on this linked RDF data to help epistemologists, historians and philologists in their analysis of these ancient texts.

Third, together with Arnaud Zucker, we have proposed to adopt argumentation theory together with Semantic Web languages and techniques to provide an overall view of conflicting critiques over ancient texts, and to detect what are the different competing viewpoints and what are the strongest arguments emerging from the debate. An ontology for argumentative documents is used to annotate ancient texts, and an example of such annotation is provided about the topic of the Eternity of the species in Aristotle.

Moreover, together with Ahmed Missaoui (UNS) and Sara Tonelli (FBK Trento, Italy), we have presented the process performed to map the metadata from the Verbo-Visual-Virtual Project to the Linked Open Data cloud and the related data enrichment. Although the work was largely inspired by past efforts by other cultural heritage institutions, they face new challenges, partly related to the small size of the collection, with little-known artists and few information available from other online sources, and partly to the integration of Natural Language Processing techniques to enrich the metadata. The results of this research have been published at the AIUCD international conference.

7.3.6. Autonomous Learning of the Meaning of Objects

Participants: Valerio Basile, Elena Cabrio, Fabien Gandon.

The goal of ALOOF (CHIST-ERA) project [0] is to enable robots to tap into the ever-growing amount of knowledge available on the Web, by learning from there about the meaning of previously unseen objects, expressed in a form that makes them applicable when acting in situated environments. By searching the Web, robots will be able to learn about new objects, their specific properties, where they might be stored and so forth. To achieve this, robots need a mechanism for translating between the representations used in their real-world experience and those on the Web.

In this direction, we are building a *machine reading* pipeline to extract formally encoded knowledge from unstructured text. By combining linguistic and semantic analysis of natural language with entity linking and formal reasoning techniques, our system is capable of extracting meaningful knowledge about entities with URIs in the Linked Open Data (e.g., from DBpedia) and their relationships, encoded in standard Semantic Web fashion, i.e., RDF triples. We then employ the machine reading software to harvest the Web, targeting informative natural language resources such as educational Websites, to create a large-scale meaning bank of common sense knowledge.

7.3.7. Social Media Intelligence and Linked Knowledge

Participants: Farhad Nooralahzadeh, Elena Cabrio, 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 on the one hand, 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.

In this context, we have developed the entity discovery tools by adopting the semantic spreading activation, then we integrated it with the SMILK framework. The goal of this work was 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

[0] http://www.dis.uniroma1.it/~aloof/

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.

7.3.8. Ontology-Based Workflow Management Systems

Participants: Tuan Anh Pham, Nhan Le Thanh.

The main objective of this PhD work is to develop a Shared Workflow Management System (SWMS) using ontology engineering. Everybody can share a semi-complete workflow which is called “Workflow template”, and other people can modify and complete it to use it in their system. This customized workflow is called “Personalized workflow”. The challenges of a SWMS are to be simple, easy to use, user friendly 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, with changes compliant with the predefined rules in the workflow template? How to build an execution model to evaluate step by step a personalized workflow [34][33].

7.3.9. Semantic Mappings with a Control Flow-Based Business Workflow

Participants: Thi Hoa Hue Nguyen, Nhan Le Thanh.

The aim of this PhD work is to improve the Coloured Petri Nets (CPNs) and Ontology engineering to support the development of business process and business workflow definitions of various fields. To realize this objective, we first propose an ontological approach for representing business models in a meta-knowledge base. We introduce four basic types of manipulation operations on process models used to develop and modify business workflow patterns. Second, we propose a formal definition of semantic constraints and an O(n3)-time algorithm for detecting redundant and conflicting constraints. By relying on the CPN Ontology and sets of semantic constraints, workflow processes are semantically created. Finally, we show how to check the semantic correctness of workflow processes with the SPARQL query language [34].

7.4. Analyzing and Reasoning on Heterogeneous Semantic Graphs

7.4.1. RDF Mining

Participants: Andrea Tettamanzi, Catherine Faron-Zucker, Fabien Gandon, Tran Duc Minh, Claudia d’Amato.

We carried on our investigation in an approach to RDF mining based on grammatical evolution and possibility theory, whose aim is to mine large RDF graphs by automatically generating and testing OWL 2 axioms based on the known facts. In particular, we addressed the problem of scaling up the scoring heuristics based on falsification and possibility theory we have recently proposed [36].

7.4.2. Data and Knowledge Integration and Extraction

Participant: Andrea Tettamanzi.

Together with Somsack Inthasone of the National University of Laos, Nicolas Pasquier and Célia da Costa Pereira of I3S, we completed a survey on biodiversity and environment data mining [16].

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

7.4.3. Scalable Uncertainty Management

Participant: Andrea Tettamanzi.

Within the framework of the CNR PEPS GéoIncertitude, we proposed and studied the properties of uncertain logical gates in possibilistic network, using a problem of human geography as a motivating example and testbed [28].

7.4.4. Natural Language Question Answering

Participants: Andrea Tettamanzi, Elena Cabrio, Catherine Faron-Zucker, Amine Hallili.

We extended previous work on answering N-relation natural language questions in the commercial domain by combining an approach to learning regular expressions based on genetic programming [21].

7.4.5. Events Detection in Twitter

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

We analyze Twitter data in the objective of identifying events reported by Twitter users. Specially we have worked on two main aspects: an approach for classifying tweets as either related or not related to events and secondly we have studied an approach for disambiguating geographic entities in tweets.

We have worked on an approach for separating event-related content from the rest of Twitter messages. We have combined technics from Natural Language Processing (NLP) and Machine Learning (ML) for building a classifier model that aims at classifying tweets into two mutually exclusive classes. First of all, we apply a Named Entity Recognizer to the tweets in order to identify the occurrences of named entities and special Twitter features such as hashtags, shortened URLs or user mentions. In a second step, the named entities are replaced by their generic class in the DBpedia Ontology; we do so by using SPARQL to query the DBpedia Knowledge Base to extract the class related to each entity. Third, we use the modified content as examples to train a binary classifier. Our evaluation using different classifiers such as Naive Bayes and Long Short Term Memory have shown promising results in term of performance compared to the state of the art.

We have also worked on an approach for identifying geographic entities in Twitter. This task is challenging for two main reasons: first, a geographic term can be related to either geographic or non geographic entities (Paris can be a person or a place) and second, many geographic places might have the same name (Paris can be either the capital of France or a city in Texas). We have proposed an approach based on distant-supervision and ontology matching for identifying and disambiguate ambiguous geographic terms.

ZENITH Project-Team

7. New Results

7.1. Big Data Integration

7.1.1. CloudMdsQL, a query language for heterogeneous data stores

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

The blooming of different cloud data management infrastructures, specialized for different kinds of data and tasks, has led to a wide diversification of DBMS interfaces and the loss of a common programming paradigm. The CoherentPaaS European project addresses this problem, by providing a common programming language and holistic coherence across different cloud data stores.

In this context, we have started the design of a 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) [19], [31]. 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. Our experimental validation, with three data stores (graph, document and relational) and representative queries, shows that CloudMdsQL satisfies the five important requirements for a cloud multidatastore query language. In [32], 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.

7.1.2. Semantic Data Integration using Bio-Ontologies

Participant: Pierre Larmande.

The AgroPortal project [49] aims at developing and supporting a reference ontology repository for the agronomic domain. The ontology portal features ontology hosting, search, versioning, visualization, comment, with services for semantically annotating data with the ontologies, as well as storing and exploiting ontology alignments and data annotations. All of these within a fully semantic web compliant infrastructure. The main objective of this project is to enable straightforward use of agronomic related ontologies, avoiding data managers and researchers the burden to deal with complex knowledge engineering issues to annotate the research data. Thus, we specifically pay attention to the requirements of the agronomic community and the specificities of the crop domain. AgroPortal will offer a robust and stable platform that we anticipate will be highly valued by the community.

7.1.3. Access and Integration of Molecular Biology Data

Participants: Sarah Cohen-Boulakia, Patrick Valduriez.

The volumes of molecular biology data available on the web are constantly increasing. Accessing and integrating these data is crucial for making progress in biology. In [26], we provide all the necessary pointers to identify the reference databases capable of providing bioinformatic data for molecular biology. We also discuss the problems posed by the exploitation of these very highly heterogeneous and distributed data. Finally, in order to guide a prospective user on the choice of one of these systems, we provide an overview of the systems that provide unified access to these data.

7.2. Distributed Indexing and Searching

7.2.1. Diversified and Distributed Recommendation for Scientific Data

Participants: Esther Pacitti, Maximilien Servajean.

Recommendation is becoming a popular mechanism to help users find relevant information in large-scale data (scientific data, web). To avoid redundancy in the results, recommendation diversification has been proposed, with the objective of identifying items that are dissimilar, but nonetheless relevant to the user's interests.

We propose a new diversified search and recommendation solution suited for scientific data (i.e., plant phenotyping, botanical data) [22] We first define an original profile diversification scoring function that enables to address the problem of returning redundant items, and enhances the quality of diversification. Through experimental evaluation using two benchmarks, we showed that our scoring function gives the best compromise between diversity and relevancy. Next, to implement our new scoring function, we propose a basic Top-k threshold-based algorithm that exploits a candidate list to achieve diversification and several techniques to improve performance. First, we simplify the scoring model to reduce its computational complexity. Second, we propose two techniques to reduce the number of items in the candidate list, and thus the number of diversified scores to compute. Third, we propose different indexing scores that take into account the diversification of items and an adaptive indexing approach to reduce the number of accesses in the index dynamically based on the queries workload. The experimentation results show that our techniques yield a major reduction of response time, up to 12 times compared to a baseline greedy diversification algorithm.

We also address distributed and diversified recommendation in the context of P2P and multisite cloud [23]. We propose a new scoring function (usefulness) to cluster relevant users over a distributed overlay. Our experimental evaluation using different datasets shows major gains in recall (order of 3 times) compared with state-of-the-art solutions.

7.3. Scientific Workflows

7.3.1. Scientific Workflows: combining data analysis and simulation

Participant: Sarah Cohen-Boulakia.

While scientific workflows are increasingly popular in the bioinformatics community in some emerging application domains such as ecology, the need for data analysis is combined with the need to model complex multi-scale biological systems, possibly involving multiple simulation steps. This requires the scientific workflow to deal with retro-action to understand and predict the relationships between structure and function of these complex systems. OpenAlea (openalea.gforge.inria.fr) developped by the EPI Virtual plants is the only scientific workflow system able to uniformly address the problem, which made it successful in the scientific community.

For the first time, we proposed a conceptualisation of OpenAlea in [42]. We introduce the concept of higher-order dataflows as a means to uniformly combine classical data analysis with modeling and simulation. We provide for the first time the description of the OpenAlea system involving an original combination of features. We illustrate the demonstration on a high-throughput workflow in phenotyping, phenomics, and environmental control designed to study the interplay between plant architecture and climatic change. Ongoing work include deploying OpenAlea on a Grid technology using the SciFloware middleware.

7.3.2. Processing Scientific Workflows in Multi-site cloud

Participants: Ji Liu, Esther Pacitti, Patrick Valduriez.

As the scale of the data increases, scientific workflow management systems (SWfMSs) need to support workflow execution in High Performance Computing (HPC) environments. Because of various benefits, cloud emerges as an appropriate infrastructure for workflow execution. However, it is difficult to execute some scientific workflows in one cloud site because of geographical distribution of scientists, data and computing resources. Therefore, a scientific workflow often needs to be partitioned and executed in a multisite environment.

In [21], we define a multisite cloud architecture that is composed of traditional clouds, e.g., a pay-per-use cloud service such as Amazon EC2, private data-centers, e.g. a cloud of a scientific organization like Inria, COPPE or LNCC, and client desktop machines that have authorized access to the data-centers. We can model this architecture as a distributed system on the Internet, each site having its own computer cluster, data and programs. An important requirement is to provide distribution transparency for advanced services (i.e., workflow management, data analysis), to ease their scalability and elasticity. Current solutions for multisite clouds typically rely on application specific overlays that map the output of one task at a site to the input of another in a pipeline fashion. Instead, we define fully distributed services for data storage, intersite data movement and task scheduling.

7.3.3. Data-centric Iteration in Dynamic Workflows

Participant: Patrick Valduriez.

Dynamic workflows are scientific workflows supporting computational science simulations, typically using dynamic processes based on runtime scientific data analyses. They require the ability of adapting the workflow, at runtime, based on user input and dynamic steering. Supporting data-centric iteration is an important step towards dynamic workflows because user interaction with workflows is iterative. However, current support for iteration in scientific workflows is static and does not allow for changing data at runtime.

In [17], we propose a solution based on algebraic operators and a dynamic execution model to enable workflow adaptation based on user input and dynamic steering. We introduce the concept of iteration lineage that makes provenance data management consistent with dynamic iterative workflow changes. Lineage enables scientists to interact with workflow data and configuration at runtime through an API that triggers steering. We evaluate our approach using a novel and real large-scale workflow for uncertainty quantification on a 640-core cluster. The results show impressive execution time savings from 2.5 to 24 days, compared to non-iterative workflow execution. We verify that the maximum overhead introduced by our iterative model is less than 5% of execution time. Also, our proposed steering algorithms are very efficient and run in less than 1 millisecond, in the worst-case scenario.

7.3.4. Analyzing Related Raw Data Files through Dataflows

Participant: Patrick Valduriez.

Computer simulations may ingest and generate high numbers of raw data files. Most of these files follow a de facto standard format established by the application domain, e.g., FITS for astronomy. Although these formats are supported by a variety of programming languages, libraries and programs, analyzing thousands or millions of files requires developing specific programs. DBMS are not suited for this, because they require loading the raw data and structuring it, which gets heavy at large-scale. Systems like NoDB, RAW and FastBit, have been proposed to index and query raw data files without the overhead of using a DBMS. However, they focus on analyzing one single large file instead of several related files. In this case, when related files are produced and required for analysis, the relationship among elements within file contents must be managed manually, with specific programs to access raw data. Thus, this data management may be time-consuming and error-prone. When computer simulations are managed by a SWfMS, they can take advantage of provenance data to relate and analyze raw data files produced during workflow execution. However, SWfMS register provenance at a coarse grain, with limited analysis on elements from raw data files. When the SWfMS is dataflow-aware, it can register provenance data and the relationships among elements of raw data files altogether in a database which is useful to access the contents of a large number of files. In [24], we propose a dataflow approach for analyzing element data from several related raw data files. Our approach is complementary to the existing single raw data file analysis approaches. We validate our approach with the Montage workflow from astronomy and a workflow from Oil and Gas domain as I/O intensive case studies.

7.4. Scalable Query Processing

7.4.1. Scalable Query Processing with Big Data

Participants: Reza Akbarinia, Miguel Liroz, Patrick Valduriez.

The popular MapReduce parallel processing framework is inefficient in case of data skew, which makes the reduce side done by a few worker nodes.

In [28], [20], we propose FP-Hadoop, which makes the reduce side of MapReduce more parallel. 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.

7.5. Data Stream Mining

7.5.1. Summarizing Uncertain Data Streams

Participants: Reza Akbarinia, Florent Masseglia.

Probabilistic data management has shown growing interest to deal with uncertain data. In [29], we focus on probabilistic time series with high volumes of data, which requires efficient compression techniques. To date, most of the work on probabilistic data reduction uses synopses that minimize the error of representation wrt. the original data. However, in most cases, the compressed data will be meaningless for usual queries involving aggregation operators such as SUM or AVG. We propose *PHA* (Probabilistic Histogram Aggregation), a compression technique whose objective is to minimize the error of such queries over compressed probabilistic data. We incorporate the aggregation operator given by the end-user directly in the compression technique, and obtain much lower error in the long term. We also adopt a global error aware strategy in order to manage large sets of probabilistic time series, where the available memory is carefully balanced between the series, according to their individual variability.

7.6. Scalable Data Analysis

7.6.1. Parallel Mining of Maximally Informative k-Itemsets in Big Data

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 i) the data set is massive, and/or ii) the length K of the informative itemset to be discovered is high. In [45], 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 estimates 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 data sets. 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.

7.6.2. Frequent Itemset Mining in Massively Distributed Environments

Participants: Saber Salah, Reza Akbarinia, Florent Masseglia.

While the problem of Frequent itemset mining (FIM) has been thoroughly studied, few solutions scale. This is mainly the case when i) the amount of data tends to be very large and/or ii) the minimum support (MinSup) threshold is very low. In [46], we study the effectiveness and leverage specific data placement strategies for improving parallel FIM (PFIM) performance in MapReduce, a highly distributed computation framework. By offering a clever data placement and an optimal organization of the extraction algorithms, we show that the itemset discovery effectiveness does not only depend on the deployed algorithms. We propose ODPR (Optimal Data-Process Relationship), a solution for fast mining of frequent itemsets in MapReduce. Our method allows discovering itemsets from massive datasets, where standard solutions do not scale.

In [44], we propose a highly scalable PFIM algorithm, namely Parallel Absolute Top Down (PATD). PATD renders the mining process of very large databases (up to Terabytes) simple and compact. Its mining process is made up of only one parallel job, which dramatically reduces the mining runtime, communication cost and energy power consumption overhead, in a distributed computational platform. Based on a clever and efficient data partitioning strategy, namely Item Based Data Partitioning (IBDP), PATD mines each data partition independently, relying on an absolute minimum support (AM inSup) instead of a relative one. Through an extensive experimental evaluation using real-world data sets, we show that PATD is significantly more efficient and scalable than alternative approaches.

7.6.3. Scalable Mining of Closed Frequent Itemsets

Participants: Mehdi Zitouni, Reza Akbarinia, Florent Masseglia.

Mining big datasets poses a number of challenges which are not easily addressed by traditional mining methods, since both memory and computational requirements are hard to satisfy. One solution is to take advantage of parallel frameworks, such as MapReduce, using ordinary machines. In [48], we address the issue of mining closed frequent itemsets (CFI) from big datasets in such environments. We introduce a new parallel algorithm, called CloPN , for CFI mining. One important feature of CloPN is to use a prime number based approach to transform the data into numerical form, and then to mine closed frequent itemsets by using only multiplication and division operations. We carried out exhaustive experiments over big real world datasets to assess the performance of CloPN. The results show that our algorithm is very efficient in CFI mining from large real world datasets with up to 53 million articles.

7.6.4. Chiaroscuro

Participants: Tristan Allard, Florent Masseglia, Esther Pacitti.

The advent of on-body/at-home sensors connected to personal devices leads to the generation of fine grain highly sensitive personal data at an unprecendent rate. However, despite the promises of large scale analytics there are obvious privacy concerns that prevent individuals to share their personnal data. In [30], we propose Chiaroscuro, a complete solution for clustering personal data with strong privacy guarantees. The execution sequence produced by Chiaroscuro is massively distributed on personal devices, coping with arbitrary connections and disconnections. Chiaroscuro builds on our novel data structure, called Diptych, which allows the participating devices to collaborate privately by combining encryption with differential privacy. Our solution yields a high clustering quality while minimizing the impact of the differentially private perturbation. Our study show that Chiaroscuro is both correct and secure.

7.6.5. Large-scale Recognition of Visual and Audio Entities

Participants: Valentin Leveau, Alexis Joly, Patrick Valduriez.

We improved our work on the retrieval of visual identities by introducing a supervised classification layer on top of the large-scale instance-based matching layer. We introduce a new match kernel based on the inverse rank of the Shared Nearest Neighbors (SNN) combined with local geometric constraints [40]. To avoid overfiting and reduce processing costs, the dimensionality of the resulting over-complete representation is further reduced by hierarchically pooling the raw consistent matches according to their spatial position in the training images. The final image representation is obtained by concatenating the resulting feature vectors at several resolutions. Learning from these representations using a logistic regression classifier is shown to provide excellent fine-grained classification performance. In [38], we transpose our new SNN match kernel to the case of audio contents (applied to bird sounds recognition). Thus, the spatial pooling of geometrically consistent visual matches is replaced by a temporal pooling of temporally consistent audio matches. The resulting classification system obtained the second best results at the LifeCLEF bird identification challenge 2015 [36], the largest challenge of this kind ever organized (1000 bird species, 33K audio recordings).

7.6.6. Crowd-sourced Biodiversity Data Production through Pl@ntNet

Participants: Alexis Joly, Julien Champ, Jean-Christophe Lombardo, Antoine Affouard.

Initiated in the context of a citizen sciences project with botanists of the AMAP laboratory and the Tela Botanica social network, Pl@ntNet [18] is an innovative collaborative plateform focused on image-based plant identification as a mean to enlist new contributors and boost the production of biodiversity data and knowledge. Since 2010, several hundreds of thousands of geo-tagged and dated plant photographs were collected and revised by tens of thousands of novice, amateur and expert botanists. A content-based identification tool, available as both web and mobile applications, is synchronized with the growing data and allows any user to query or enrich the system with new observations. As a concrete new result, the cumulative number of downloads of the iPhone or Android app did reach 1M in October 2015. One of the main novelty in 2015 was the introduction of deep learning technologies in order to improve classification performance as well as the quality and speed of the content-based image retrieval.

A comparative study that we conducted in the context of the LifeCLEF [0] plant identification challenge did actually confirm that deep convolutional neural networks definitely outperforms the best fine-grained classification models on the aggregation of hand-crafted visual features [33]. Thus, we integrated this technology in the Pl@ntNet platform and exploited it in two ways: (i) for extracting more relevant (local and global) visual features to be indexed and searched within our efficient content-based indexing and retrieval framework (SnoopIm software) (ii) for reranking the species returned by the content-based search engine so as to increase the average reciprocal rank of the correct species while keeping a good level of interpretability of the returned results.

7.6.7. Crowd-sourced Biodiversity Data Production through LifeCLE

Participants: Alexis Joly, Julien Champ, Jean-Christophe Lombardo, Antoine Affouard.

We continued sharing the data produced by the Pl@ntNet platform with the international research community through the animation of the LifeCLEF research platform and the set-up of three new challenges, one related to plant images, one to bird sounds and one to fish videos. More than 200 research groups registered to at least one of the challenges and about 20 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 2015 overview paper [37] and more detailed analyses are provided in technical reports for the plant task [35] and the bird task [36]. We also report on an experimental study aimed at evaluating how state-of-art computer vision systems perform in identifying plants compared to human expertise [15],. A subset of the evaluation dataset used within LifeCLEF 2014 plant identification challenge was shared with volunteers of diverse expertise, ranging from leading experts of the targeted flora to inexperienced test subjects. In total, 16 human runs were collected and evaluated comparatively to the 27 machine-based runs of LifeCLEF challenge. The main outcome of the experiment was that machines are still far from outperforming the best expert botanists but they are clearly competing with some experienced botanists specialists of other floras.

[0] www.lifeclef.org

ALICE Project-Team

7. New Results

7.1. Dihedral Angle-Based Maps of Tetrahedral Meshes

Participants: Nicolas Ray, Bruno Lévy.

This work is a collaboration with Gilles-Philippe Paillé (visiting), Pierre Poulin (U. de Montréal) and Alla Sheffer (UBC).

Given a 2D triangulation, it is well known that it is reasonably easy to reconstruct the shape of all the triangles from the sole data of the angles at the triangle corners, provided that they satisfy some constraints. In this project, we studied how this idea can be generalized in the volumetric setting. In other words, we proposed a geometric representation of a tetrahedral mesh that is solely based on dihedral angles, and what are the constraints that these dihedral angles need to satisfy to make that possible. We first show that the shape of a tetrahedral mesh is completely defined by its dihedral angles. This proof leads to a set of angular constraints that must be satisfied for an immersion to exist in $\mathbb{R}^3$. This formulation lets us easily specify conditions to avoid inverted tetrahedra and multiply-covered vertices, thus leading to locally injective maps. We then present a constrained optimization method that modifies input angles when they do not satisfy constraints. Additionally, we develop a fast spectral reconstruction method to robustly recover positions from dihedral angles. We demonstrate the applicability of our representation with examples of volume parameterization, shape interpolation, mesh optimization, connectivity shapes, and mesh compression. This work has been published in Transactions on Graphics [17].

7.2. 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 a two steps procedure that defines the desired orientation of the hexahedra with a frame field, then integrates this frame field to generate a deformed 3D grid inside the object. Hexahedra are generated where the grid is non degenerated and not too distorted, and tetrahedra will fill the remaining volume.

7.2.1. Frame Field Generation

A frame field must define the orientation of a cube (the less deformed hexahedron) everywhere inside the object. This object is very difficult to manipulate because it has to be invariant by rotation of 90 degrees around each of its facet normal vector. In [26] we have designed a fast algorithm that is able to define a smooth frame field constrained to be aligned with the object boundary. We represent frames by spherical harmonics (as introduced in [33]) and greatly improve the state of the art thanks to an expression of the boundary constraints that keep the objective function of the optimisation problem very close to quadratic.

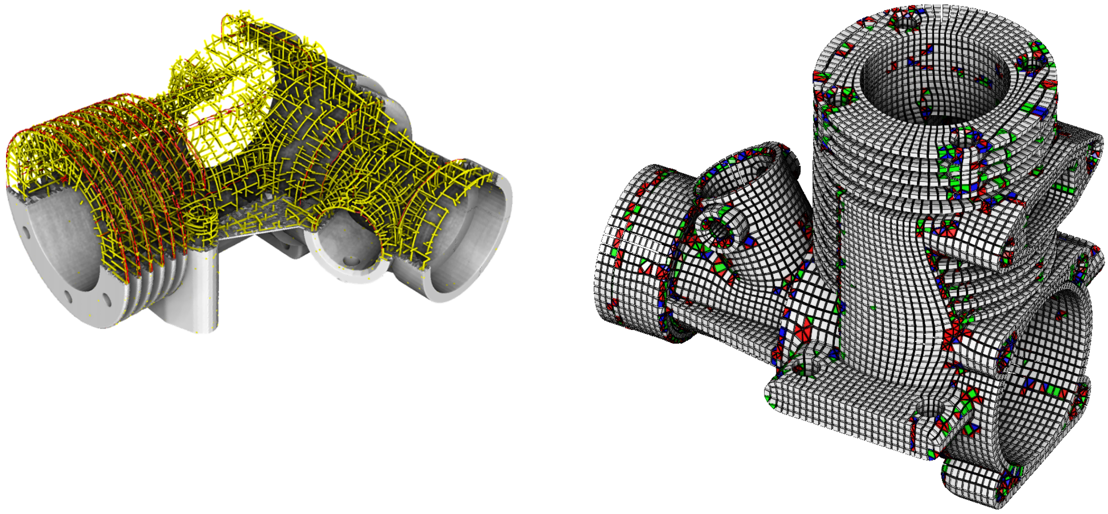


Figure 1. Our Hexahedral dominant meshes generation pipeline have two steps: the generation of a smooth frame field (yellow curves on the left image), and the construction of a mesh with hexahedron that aligns with the frame field.

7.2.2. Generation of Hexahedral-dominant Meshes

The generation of the hex dominant meshes is performed in two steps: place a deformed 3D grid inside the object such that it is aligned with the frame field, then use it to produce hexahedra and fill the rest of the volume with tetrahedra. We developed two solutions for the first step: a 3D extension of PGP [40] and an adaptation of Cubecover [39]. Both solutions have pros and cons so we plan to make them cooperate in a near future. The conversion of this result in an hexahedral-dominant mesh is a very complex problem for which we have a fair solution: we extract a point set from the deformed 3D grid, generate a tetrahedral mesh of the object that is constrained to include the point set in its vertices. From this tetrahedral mesh, we merge sets of tetrahedra into hexahedra with an extension of [37]. We are now working on an alternative solution that will generate hexahedra directly from the deformed 3D grid, and extract the boundary of the rest of the volume as a 2D mesh. From this mesh, we will try to produce more hexahedra by adapting existing combinatorial methods [27].

7.2.3. 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, but the best function basis for each case. For hexahedral dominant meshes, we want to determine if the benefit of using hexahedra deserves having specific function bases devoted to properly link tetrahedral and hexahedral elements. We are developing a new function basis, tailored to non-conformal mixed hexahedra-tetrahedra meshes. Using a combination of tri-linear and quadratic hexahedra, it is possible to construct a space of continuous functions even on a non-conformal mesh. We are now proceeding to analyse the properties of this function space, both theoretically and experimentally. This topic is addressed in the (ongoing) Ph.D. thesis of Maxence Reberol.

7.3. Semi-discrete Optimal Transport in 3D

Participant: Bruno Lévy.

This work introduces a practical algorithm to compute the optimal transport map between a piecewise linear density and a sum of Dirac masses in 3D. In this semi-discrete setting, Aurenhammer *et al.* showed that the optimal transport map is determined by the weights of a power diagram [28]. The optimal weights are computed by minimizing a convex objective function with a quasi-Newton method. To evaluate the value and gradient of this objective function, we propose an efficient and robust algorithm that computes at each iteration the intersection between a power diagram and the tetrahedral mesh that defines the measure. Like in the multilevel proposed by Mérigot, we use a hierarchical algorithm, that uses nested point sets to discretize the source measure.
We think this work may lead to interesting discretizations of the physics, that include the conservation laws (conservation of energy, conservation of momentum ...) deep in their definition, as explained by Jean-David Benamou and Yann Brenier in their fluid dynamics formulation of optimal transport [30].

This work was published in the journal *Mathematical Modeling and Analysis* [10].

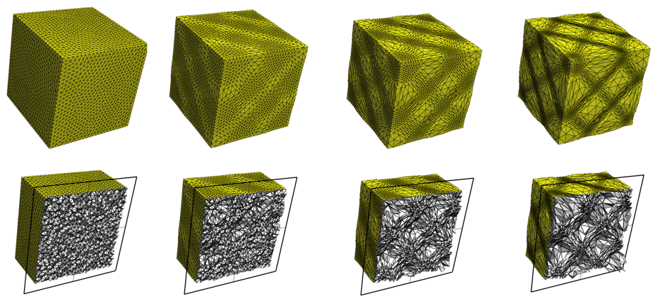


Figure 2. Semi-discrete optimal transport from a constant density to a varying one (product of sines).

7.4. By-example Synthesis of Structurally Sound Shapes

Participants: Jonas Martínez Bayona, Jérémie Dumas, Sylvain Lefebvre.

This is a collaboration with Li-Yi Wei (HKU) on a first project, An Lu (Inria/TU. Munich), Jun Wu (TU. Munich) and Christian Dick (TU. Munich) on a second project.

This work is at the heart of the ERC ShapeForge and considers the by-example synthesis of shapes under structural constraints. We considered two views of the problem that lead to different methodologies.

In a first approach, our goal is to cover a surface with a pattern – an operation akin to texturing in Computer Graphics. The pattern is however used to define the final shape, by determining which parts of the surface are solid or empty. The method operates on a thin voxel shell and does not require any parametrization of the input surface. The pattern is synthesized using a novel formulation for by-example pattern synthesis along surfaces. It is analyzed for structural weaknesses and this information is fed back to the pattern synthesizer, so that seamless reinforcements are added to the structure. We collaborated with researchers from T.U. Munich to analyze the structural behaviour of our structures, and developed a fast evaluation scheme that can be used within our optimization loop to guarantee structural soundness of the resulting design. The work was published in ACM Transactions on Graphics in 2015 [9] (proceedings of SIGGRAPH 2015).

In a second approach we considered the synthesis of shapes that are as rigid as possible under specific boundary conditions and using a prescribed amount of material, while resembling a given input example pattern, as illustrated in Figure 3 . Our method is inspired by the field of topology optimization, where rigid shapes are optimized but without any appearance constraints. Our algorithm generates shapes that resemble the input exemplar while being within a user specified percentage of the most rigid shape obtained without the appearance objective. The work was published in ACM Transactions on Graphics in 2015 [12] (proceedings of SIGGRAPH Asia 2015).

Figure 3. Left. A chair automatically synthesized from a load scenario and an example pattern. The rigidity of the chair is within controlled bounds of a shape optimized without appearance objective. Right. A table design automatically synthesized.

7.5. Modeling for Fabrication

We pursued our research regarding automatic modeling techniques for fabrication, where an algorithm takes into account fabrication constraints to simplify the modeling process. This year we have worked on three projects in this area: the modeling of mechanisms from incomplete 2D definitions, the modeling of self-supporting tight enclosures to assist the fabrication process, and the interactive sculpting of support-free objects.

7.5.1. 3D Fabrication of 2D Mechanisms

Participants: Jean Hergel, Sylvain Lefebvre.

This project considered the automatic modeling of 3D mechanisms from an under-specified 2D model of the mechanism. Our approach casts the synthesis problem as an edge orientation problem in a graph, where graph nodes represent parts of the mechanisms and edges capture their interactions as analyzed by the 2D simulation of the mechanism. The edge orientation determines which parts include which others. Once all inclusions have been determined, we formulate a CSP to solve for the layering problem: each part is assigned a depth 'layer' in 3D. We finally compute the final geometry through CSG (boolean combinations of shapes). This work has been published in Computer Graphics Forum (proceedings of Eurographics 2015) [8]. It received an honorable mention from the best paper committee.

7.5.2. Self-supporting Tight Enclosures

Participants: Samuel Hornus, Sylvain Lefebvre, Frédéric Claux, Jérémie Dumas.

The aim of this project was to develop a technique to automatically generated a tight enclosure in the free space around an object. The challenge was to ensure that the enclosure stays close to the object and be as thin as possible while still being printable without collapsing. Such an enclosure finds at least two important applications : 1. as a protective skin to avoid artifacts when 3D-printing a multi-material object. 2. for generating as-large-as-possible cavities inside the printed object in order to minimized material usage and print time. The work is available as an Inria technical report [22].

7.5.3. *Interactive Sculpting of Support-free Objects*

Participants: Tim-Christopher Reiner, Sylvain Lefebvre.

Tim Reiner, former PhD student at the Karlsruhe Institute of Technology, joined the team on a Post-Doc position to explore new ideas in the context of modeling, rendering, and fabrication. Starting in March 2015, he developed a voxel-based environment for interactive modeling. In a research project together with Sylvain Lefebvre, our team has derived novel techniques for sculpting support-free 3D shapes. These shapes have the property that they do not require support structures during fabrication on fused deposition modeling or resin-based printers. This work is currently under review.

7.6. Intersection Detection via Gauss Maps; a Review and New Techniques

Participant: Samuel Hornus.

We have revisited the problem of deciding whether two convex objects intersect or not. A systematic view of the problem for polyhedra led us to a unified view of several techniques developed in the computer graphics community and to a new and very fast technique specialized to pairs of tetrahedra. A novel view of the problem as a minimization problem over the 2-sphere led us to the description of new interesting links between the set of planes separating two objects and the silhouette edges of their Minkowski difference. From there, we devised a new algorithm for separating two arbitrary convex objects that is a little bit faster and much more robust than the state of the art technique of Gilbert, Johnson and Keerthi [31]. The work has been summarized in [21].

7.7. Fractal Geometry

Participant: Dmitry Sokolov.

This is a collaboration with Christian Gentil (LE2I), Gilles Gouatay (LSIS), Anton Mishkinis (LE2I).

Additive manufacturing enables for the first time the physical realization of objects having complex geometries. Good approximations of fractals, in particular, can now be manufactured in a variety of materials, including metals. The application domains of fabricated fractal geometries are vast, from the be design of “fractal” micro-strip antennas, to the creation of meta-materials.

The main challenge with traditional fractals is the control of the resulting geometry. For example, it is quite challenging to get the desired shape using the system of fractal homeomorphisms proposed by Barnsley [29]. We elaborate here a new type of modeling system, using the facilities of existing CAGD software, while extending their capabilities and their application areas. This new type of modeling system will offer designers (engineers in industry) and creators (visual artists, stylists, designers, architects, etc.) new opportunities to design and produce a quick mock-up, a prototype or a single object. Our approach is to expand the possibilities of a standard CAD system by including fractal shapes while preserving ease of use for end users.

This year we published two papers on the subject [20], [16].

AVIZ Project-Team

7. New Results

7.1. Design Considerations for Composite Physical Visualizations

Participants: Mathieu Le Goc [correspondant], Pierre Dragicevic, Samuel Huron, Jean-Daniel Fekete.

Physical visualization has existed for thousands of years, yet the Information Visualization community is just starting to study it. Many current physical visualizations are monolithic, static, and not interactive. Some of them are made of multiple individual objects that can be rearranged in order to represent a variety of informative configurations. We call them composite physical visualizations. A major benefit of such visualizations is that they support modularity and updatability, but their design space is not well understood.

We show [28] that composite physical visualizations can be classified according to two orthogonal dimensions: i) their level of actuation and ii) their manipulability. Among existing systems, some have a high manipulability but no support for actuation, while others are fully actuated but not manipulable. Only a few systems are combining both qualities and none supports both full manipulability and full actuation. We discuss the tradeoffs between these two dimensions, and identify the opportunities and challenges for future research and design.

7.2. Design Considerations for Enhancing Word-Scale Visualizations with Interaction

Participants: Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, Petra Isenberg.

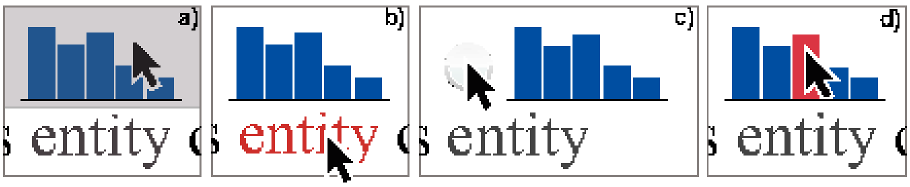


Figure 8. Illustration of where interaction can take place in the context of word-scale visualizations.

This paper presents a design space for interaction with word-scale visualizations. Most sparklines and word-scale visualizations are static and do not support any interaction. However, when word-scale visualizations are used in digital environments, interaction can enhance their use by allowing various data manipulation and management operations. Our design space covers where (Figure 8), when, and how interaction can be triggered for word-scale visualizations embedded in a text document. It also includes how and when to transition from a view where the text with word-scale visualizations is the focus (document-centric view) to a view in which the visualizations becomes the reading focus (visualization-centric view).

7.3. Drawing Characteristics for Reproducing Traditional Hand-Made Stippling

Participants: Domingo Martín, Vicente Del Sol, Celia Romo, Tobias Isenberg [correspondant].

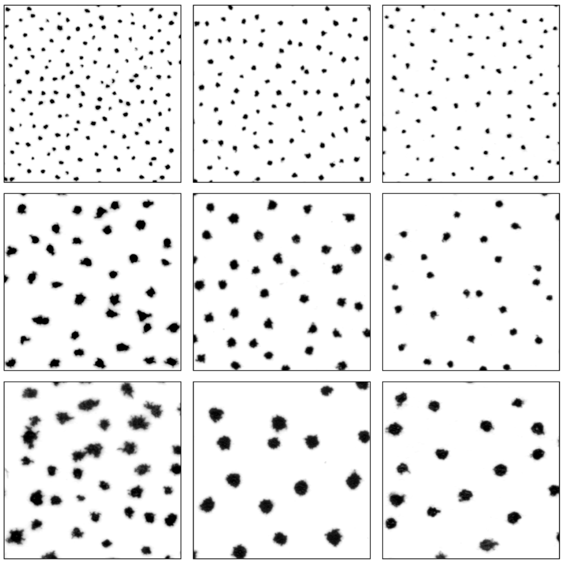


Figure 9. Samples from the stipple dot experiments.

We contribute an in-depth analysis of the characteristics of traditional stippling and relate these to common practices in NPAR stippling techniques as well as to the abilities and limitations of existing printing and display technology. We focus specifically on the properties of stipple dots and consider the dimensions and attributes of pens and paper types used in artistic practice (see Figure 9). With our analysis we work toward an understanding of the requirements for digital stippling, with the ultimate goal to provide tools to artists and illustrators that can replicate the stippling process faithfully in the digital domain. From the results of our study we provide a dataset for use in new example-based stippling techniques, derive a taxonomy of characteristics and conditions for the reproduction of stippling, and define future directions of work.

More on the project Web page: http://tobias.isenberg.cc/VideosAndDemos/Martin2015DCR.

7.4. Evaluation of an IEC Framework for Guided Visual Search

Participants: Nadia Boukhelifa [correspondant], Anastasia Bezerianos, Waldo Cancino, Evelyne Lutton.

We evaluated and analysed a framework for Evolutionary Visual Exploration (EVE) [13] (Figure 10) 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. This method smoothly combines automatically calculated metrics and user input in order to propose pertinent views to the user. We revisited this framework and a prototype application that was developed as a demonstrator, and summarized our previous study with domain experts and its main findings. We then reported on results from a new user study with a clear predefined task that examined how users leveragde the system and how the system evolved to match their needs.

While previously we showed that using EVE, domain experts were able to formulate interesting hypotheses and reach new insights when exploring freely, the new findings indicated that users, guided by the interactive evolutionary algorithm, were able to converge quickly to an interesting view of their data when a clear task was specified. We provided a detailed analysis of how users interact with an evolutionary algorithm and how the system responded to their exploration strategies and evaluation patterns. This line of work aims at building a bridge between the domains of visual analytics and interactive evolution. The benefits are

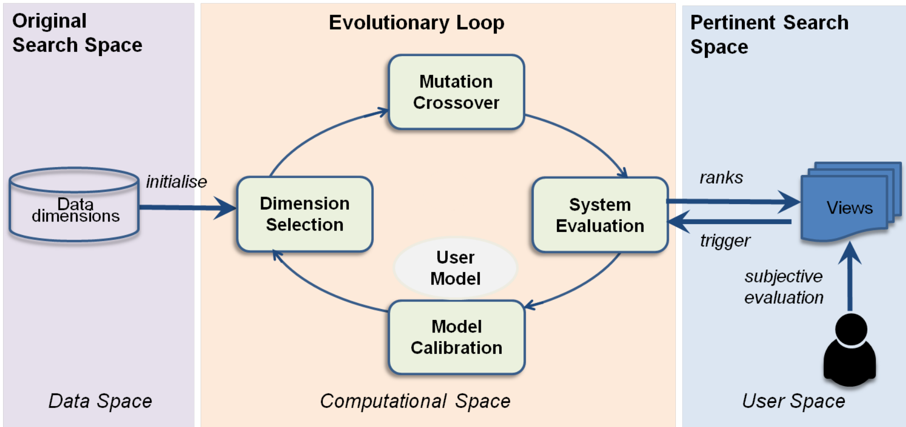


Figure 10. Illustration of the evolutionary visual exploration framework.

numerous, in particular for evaluating Interactive Evolutionary Computation (IEC) techniques based on user study methodologies.

Next, we summarized and reflected upon our experience in evaluating our guided exploratory visualization system [34]. This system guided users in their exploration of multidimensional datasets to pertinent views of their data, where the notion of pertinence is defined by automatic indicators, such as the amount of visual patterns in the view, and subjective user feedback obtained during their interaction with the tool. To evaluate this type of system, we argued for deploying a collection of validation methods that are: user-centered, observing the utility and effectiveness of the system for the end-user; and algorithm-centered, analysing the computational behaviour of the system. We reported on observations and lessons learnt from working with expert users both for the design and the evaluation of our system.

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

7.5. Exploring the Effect of Word-Scale Visualizations on Reading Behavior

Participants: Pascal Goffin, Wesley Willett, Anastasia Bezerianos, Petra Isenberg.

We studied how the integration of small visualizations (word-scale visualizations) into a sentence affects reading speed and memorization during a brief reading task. In particular, we were interested in how different placement types—with their inherent text appearance and layout changes—affect readers. We designed a quantitative study in which we gave sentences with or without small visualizations for participants to read (study conditions are shows in Figure 11). Then, we invited them to answer questions on the sentences. We found that the information encoded in the visualizations is more prominent and easily remembered than information in the written text, but that different placement options had little to no effect on reading performance, even if participants had different preferences.

7.6. Exploration of the Brain's White Matter Structure through Visual Abstraction and Multi-Scale Local Fiber Tract Contraction

Participants: Maarten H. Everts, Eric Begue, Henk Bekker, Jos B. T. M. Roerdink, Tobias Isenberg [correspondant].

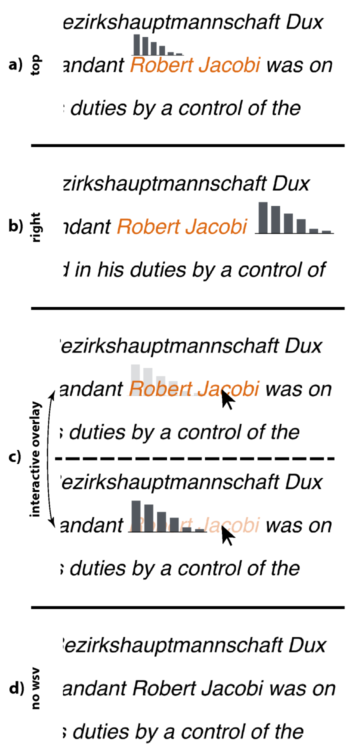


Figure 11. Illustration of the study conditions.

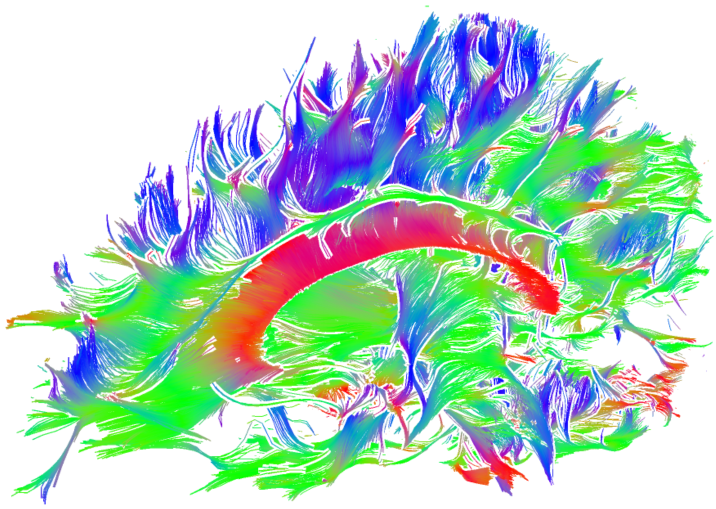


Figure 12. Contraction of the white matter fiber tracts in the brain.

We developed a visualization technique for brain fiber tracts from DTI data that provides insight into the structure of white matter through visual abstraction. We achieve this abstraction by analyzing the local similarity of tract segment directions at different scales using a stepwise increase of the search range. Next, locally similar tract segments are moved toward each other in an iterative process, resulting in a local contraction of tracts perpendicular to the local tract direction at a given scale. This not only leads to the abstraction of the global structure of the white matter as represented by the tracts, but also creates volumetric voids (see Figure 12). This increase of empty space decreases the mutual occlusion of tracts and, consequently, results in a better understanding of the brain's three-dimensional fiber tract structure. Our implementation supports an interactive and continuous transition between the original and the abstracted representations via various scale levels of similarity. We also support the selection of groups of tracts, which are highlighted and rendered with the abstracted visualization as context.

More on the project Web page: http://tobias.isenberg.cc/VideosAndDemos/Everts2015EBW.

7.7. Interactive Illustrative Line Styles and Line Style Transfer Functions for Flow Visualization

Participants: Maarten H. Everts, Henk Bekker, Jos B. T. M. Roerdink, Tobias Isenberg [correspondant].

We present a flexible illustrative line style model for the visualization of streamline data. Our model partitions view-oriented line strips into parallel bands whose basic visual properties can be controlled independently. We thus extend previous line stylization techniques specifically for visualization purposes by allowing the parametrization of these bands based on the local line data attributes. Moreover, our approach supports emphasis and abstraction by introducing line style transfer functions that map local line attribute values to complete line styles. With a flexible GPU implementation of this line style model we enable the interactive exploration of visual representations of streamlines. We demonstrate the effectiveness of our model by applying it to 3D flow field datasets (see Figure 13).

More on the project Web page: http://tobias.isenberg.cc/VideosAndDemos/Everts2015IIL.

Figure 13. Illustrative line styles applied to streamline visualization of a complex 3D flow.

7.8. Research Agenda for Data Physicalization

Participants: Yvonne Jansen, Pierre Dragicevic [correspondant], Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, Kasper Hornbæk.

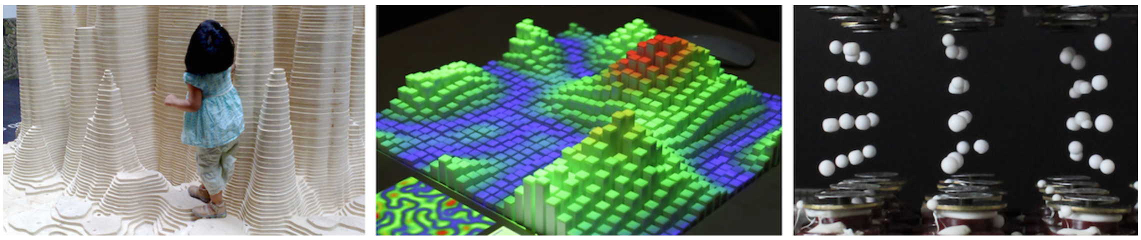


Figure 14. Three examples of data physicalizations.

Physical representations of data have existed for thousands of years. Yet it is now that advances in digital fabrication, actuated tangible interfaces, and shape-changing displays are spurring an emerging area of research that we call Data Physicalization. It aims to help people explore, understand, and communicate data using computer-supported physical data representations. We call these representations physicalizations, analogously to visualizations – their purely visual counterpart. We joined our efforts with research teams from Europe and published a research agenda where we go beyond the focused research questions addressed so far by delineating the research area, synthesizing its open challenges and laying out opportunities for future work. Examples can be seen in Figure 14 .

More on the Data Physicalization Wiki: dataphys.org/.

7.9. Storytelling and Engagement

Participants: Jeremy Boy, Jean-Daniel Fekete, Françoise Detienne.

We conducted three web-based field experiments, in which we evaluated the impact of using initial narrative visualization techniques and storytelling on user-engagement with exploratory information visualizations. We conducted these experiments on a popular news and opinion outle, and on a popular visualization gallery website. While data journalism exposes visualizations to a large public, we do not know how effectively this public makes sense of interactive graphics, and in particular if people explore them to gain additional insight to that provided by the journalists. In contrast to our hypotheses, our results indicated that augmenting exploratory visualizations with introductory 'stories' does not seem to increase user-engagement in exploration.

Many online data graphics use narrative design elements to explain a given dataset in a straightforward and compelling way. According to New York Times graphic editors Mike Bostock and Shan Carter, these explanatory graphics are preferable for data-journalism, as they have the advantage of exposing up-front what the main insights from the data are, without making people '‘have to work for them.” However, most only provide limited interactivity, which reduces the potential for personal extraction of insight. In essence and by definition, Information visualization (Infovis) is interactive and exploratory. Thus, finding ways to make exploratory graphics more accessible and engaging to people is important, because if open/public/civic data is to truly empower people, then these people should be able to use appropriate tools to gain their own insights and knowledge—not only that provided by journalists in articles written or designed from a specific perspective. We explored the potential of narrative visualization techniques and storytelling to trigger this desired user-engagement. By engagement, we specifically mean a user's investment in the exploration of a visualization.

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 indepth observational studies and controlled experiments which contribute to theories and frameworks that unify our findings and help us generate new, advanced interaction techniques. Although we continue to explore the theory of Instrumental Interaction in the context of multi-surface environments [23], we are also extending it into a wider framework we call *information substrates*. This has resulted in several prototypes, such as Webstrates [18]. We also continue to study elementary interaction tasks in large-scale environments, such as pointing [11] and object manipulation [15].

Information substrates – "Instrumental interaction" argues that, since our interaction with the physical world is often mediated by tools, or instruments, we should do the same in the digital world. Our work on multisurface environments has demonstrated the value of this model, for example, to support distributed interfaces in which the user controls the content of a wall-sized display using handheld devices [23]. Instrumental interaction does not, however, describe the "objects of interest" that instruments interact with, nor does it explain how an object becomes an instrument, nor how users appropriate them in unexpected ways (the principle of "co-adaptation").

"Information substrates" embrace a wider scope than instrumental interaction: both objects and instruments are "substrates" that hold information and behavior, and can be combined in arbitrary ways. What makes an object an instrument is defined not by what it is but by how the user uses it. We started to explore this concept wiyh Webstrates [23], a web-based environment in which content and tools are embedded in the same information substrate—in this case the Document Object Model (DOM) (Figure 7).

Our work on information substrates has influenced other projects in the group. For example, our work on tools to help programmers parallelize and optimize their code [22] uses coordinated views of the code: a traditional text view and a graphical polyhedric visualization (Figure 2). These two substrates afford different types of manipulation by the user, but share the same underlying information, i.e. the algorithm being designed. The SketchSliders technique [20], described in the following section, provides users with an easily customizable approach to control complex visual displays. SketchSliders act as a substrate for creating slider instruments, which are independent from but tightly coupled with the visual display they control. By letting users define their own sliders, we solve the long-standing problem of combining power and simplicity. Finally, the ColorLab prototypes [17], described in the following section, provide artists and graphic designers with substrates that offer novel ways to interact with and display color relationships.

Interaction in the large – ExSitu and its predecessor InSitu have a long history of studying the most fundamental action in visual environments: pointing. We recently published an extensive 64-page journal article [11] on our studies of pointing on large, wall-sized displays. In such environments, users must be able to point from a distance, typically up to a few meters from the screen, with great accuracy. Existing techniques are ill-suited for this task, due to the combination of the high index of difficulty and the constraint that users must be able to move around in the room while pointing.

We have designed and tested a number of techniques, including dual-mode techniques that combine coarse pointing with direct techniques, such as ray-casting or using the orientation of the head, and fine pointing with relative techniques, such as using a hand-held touchpad 3 . Rather than proposing the "ultimate" pointing technique for such environments, we provide a set of criteria, a set of techniques derived from those criteria, and a calibration technique for optimizing the transfer functions used by relative pointing techniques under extreme conditions.

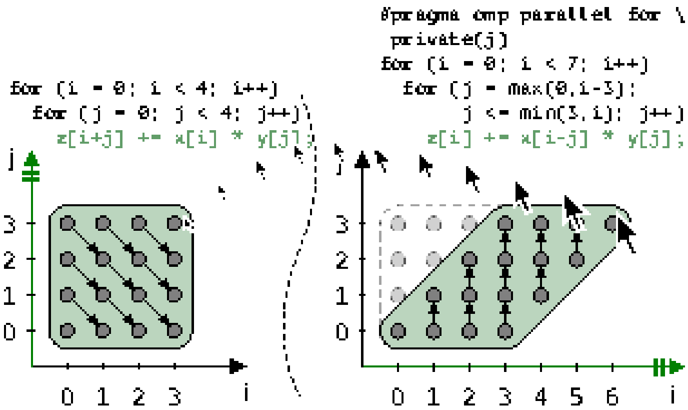


Figure 2. Performing a skew transformation to parallelize polynomial multiplication. The code is automatically transformed from its original form (left) to the skewed one (right).

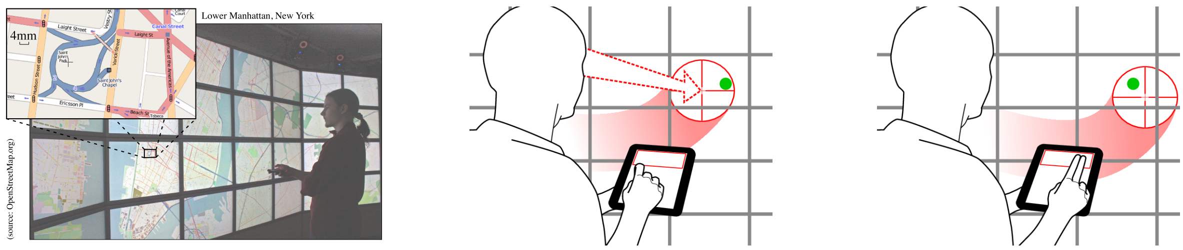


Figure 3. The challenge of pointing on a wall from a distance on a ultra-high resolution wall-sized display (left). Two of the pointing techniques that we evaluated: coarse pointing using the orientation of the head (center) vs. a two-finger swipe (right). In both cases, a one-finger swipe controls precise pointing.

In collaboration with the Inria REVES group in Sophia Antipolis, we proposed a framework for analyzing 3D object manipulation in immersive environments [15]. We decomposed 3D object manipulation into the component movements, taking into account both physical constraints and mechanics. We then fabricated five physical devices that simulate these movements in a measurable way under experimental conditions. We implemented the devices in an immersive environment and conducted an experiment to evaluate direct finger-based against ray-based object manipulation. We identified the compromises required when designing devices that (i) are reproducible in both real and virtual settings, and (ii) can be used in experiments to measure user performance.

7.2. Creativity

Participants: Sarah Fdili Alaoui, Michel Beaudouin-Lafon, Ghita Jalal, Wendy Mackay, Joseph Malloch, Nolwenn Maudet, 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. This year, we conducted studies and created tools for a variety of such users. Based on contextual interviews with artists, designers and scientists, we created the *Color Portraits* design space [17] to characterize color manipulation activities, which influenced the design of a set of color manipulation tools (*Color Lab*). We designed *BricoSketch* [21] to enable professional illustrators to work at different levels of detail on paper. We studied how makers *remix* each others' designs by analyzing metadata from over 175,000 digital designs from Thingiverse [19]. We created *SketchSliders* [20] to help scientists explore their data by sketching and manipulating free-form interactive controllers. Finally, we studied the meaning and use of the term *evaluation* within the NIME (New Interfaces for Musical Expression) community [14].

Our studies of these "extreme users" allows us to obtain empirical grounding for the theoretical concepts of instrumental interaction, information substrates and co-adaptive systems. We expect to transfer what we learn to the design of creative tools, first for expert users, then for non-specialists and non-professional users.

Color Portraits – We conducted contextual interviews with 16 participants, who provided detailed examples of how they used color to create 69 different artistic or technical artifacts [17]. Based on results from these interviews, we created the Color Portraits design space to help identify color manipulation requirements that are poorly addressed by today's color manipulation tools. We then developed a set of novel color-manipulation tools that test the generative power of the design space. We presented these to users as probes. Our observations of how users interacted with the color probes provide implications for the design of more advanced tools.

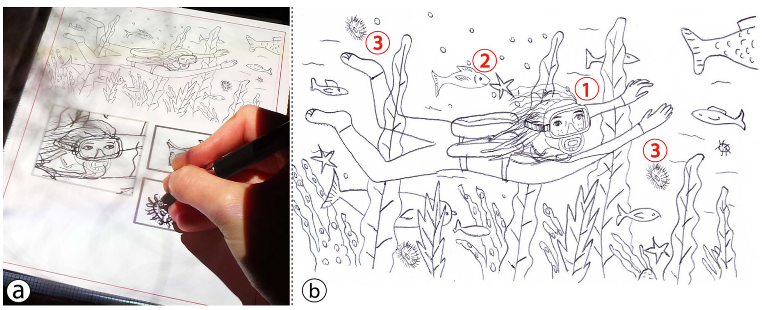


Figure 4. An artist drawing on paper with BricoSketch. (a) The artist has created three views on paper to draw parts of the illustration with higher detail: (1) the head of the diver, (2) a fish, and (3) an urchin. (b) The final composition after blending the partial views together.

BricoSketch – We conducted interviews with four professional illustrators and investigated how they use technology and paper in their creative process [21]. We also studied the evolution of the work of one of these illustrators for a period of two years. In interaction with this artist, we designed BricoSketch. BricoSketch enables illustrators to interactively create partial views of their drawings. Such views can be transposed and rescaled. Artists can then use them to create variations of their illustrations or add details with higher drawing precision. Our implementation is based on interactive paper technology that allows for above-the-surface interaction and supports traditional drawing tools such as common pens and pencils.

Remixing Designs – We investigated [19] how makers remix digital designs for physical objects on “Thingiverse”, a well-established online 3D-printing maker community. We collected metadata from over 175,000 digital designs and analyzed the *remixing graph* – links between sources and remixes that primarily exhibit a tree-like inheritance structure. We also used this data to identify particularly influential and surprising “Things”, which we further examined via qualitative case studies. We concluded with specific suggestions for online design repositories and design software so as to provide better support for remixing, and thus build stronger online maker communities.

SketchSliders – We developed SketchSliders [20], range sliders that users can freely sketch on a mobile device to parametrize and customize their data exploration on a wall display. With a small combination of sketches and gestures, users can create complex interactive controllers, such as slider branches and data transformation sliders 5 . In addition to their natural custom shape, the sketched sliders can also be enhanced by interaction aids such as slider cursors, markers and distribution visualizations. We evaluated the sketching interface with six visualization experts and found that SketchSliders accommodate a wide range of exploration strategies, as well as help users focus and customize their visual explorations.

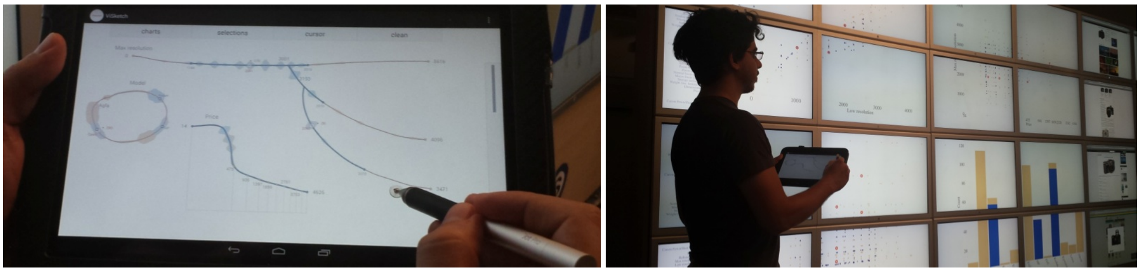


Figure 5. SketchSliders let users directly sketch viusalization controllers to explore multidimensional datasets.

Evaluation for NIME – We explored the use of evaluation techniques and terminology within the past three years of the New Interfaces for Music Expression (NIME) conference [14]. We categorized each paper that mentioned evaluation according to five criteria: a) targets and stakeholders considered, b) goals set, c) criteria used, d) methods used, and e) duration of evaluation. Results suggest that the NIME community does not share a common culture with respect to evaluation, with little consistency regarding use of the term. This paper raises the issue of evaluation within NIME community, with the goal of using it more consistently and effectively in the future.

7.3. 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 started to investigate how to support telepresence among large, heterogeneous interactive spaces [24], [25]. In particular, we studied how accurately a user can interpret deictic gestures in a video feed of a remote user [12]. These deictic gestures are important for conveying non-verbal cues for communication between remote users. We also created *Webstrates* [18], an environment for exploring shareable dynamic media and the concept of *information substrate*.

Telepresence among large, heterogeneous interactive spaces – Large interactive spaces are powerful tools that can help scientific, industrial and business users to collaborate on large and complex data sets. In order to reach their full potential, these spaces must not only support local collaboration, but also collaboration with remote users, who may have significantly different display and interaction capabilities, such as a wall-display connected to an immersive CAVE.

We explain why supporting telepresence across large interactive spaces is critical for remote collaboration [24]. We have also started to explore how such asymmetric interaction capabilities provide interesting opportunities for new collaboration strategies in large interactive spaces [25].

Accuracy of deictic gestures for telepresence – In the context of telepresence on large wall-sized displays, we investigated how accurately a user can interpret the video feed of a remote user showing a shared object on the display, by looking at it or by looking and pointing at it (Figure 6) [12]. We also analyzed how sensitive distance and angle errors are to the relative position between the remote viewer and the video feed. We showed that users can accurately determine the target, that eye gaze alone is more accurate than when combined with the hand, and that the relative position between the viewer and the video feed has little effect on accuracy. These findings can inform the design of future telepresence systems for wall-sized displays.

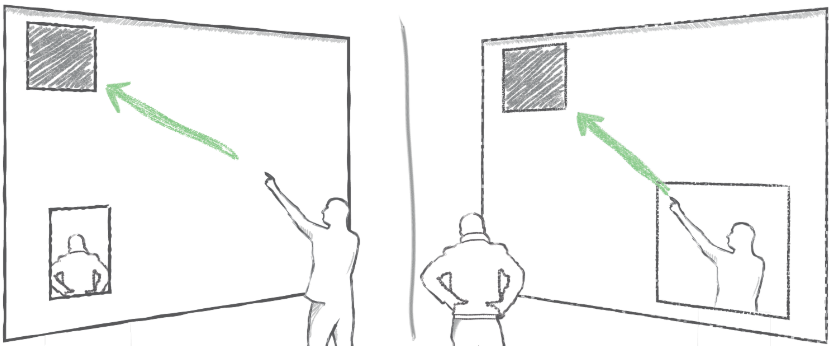


Figure 6. Users working on shared objects using two remote wall-sized displays : a user (left) shows a shared object by pointing at it and the remote user (right) can see which object is being shown through the video feed.

Webstrates – In collaboration with Université of Aarhus (Denmark) and Institut Mines Telecom, we created Webstrates [18], a system inspired by Alan Kay's early vision of interactive dynamic media. Webstrates is based on web technology: web pages served by the Webstrates server can be shared in real time among multiple users, on any web-enabled device. By using transclusion, a webstrate page can include other Webstrates. Webstrates can also include code, making them dynamic and interactive. A Webstrate that can act on another, transcluded Webstrate, is similar to an editor on a classical desktop environment. However the distinction between content and tools, documents and applications is blurred, e.g. content can be used as a tool, and tools can be shared like regular content. We implemented two case studies to illustrate Webstrates (Figure 7). We authored the article collaboratively, using functionally and visually different editors that we could personalize and extend at run-time. We also used Webstrates to orchestrate a presentation, using multiple devices to control the presentation, to let the audience participate and the session chair organize the session. We demonstrated the simplicity and generative power of Webstrates with three additional prototypes and evaluated them from a systems perspective. Webstrates runs in our WildOS middleware on the WILD and WILDER rooms, and is used for some of our projects on telepresence.

Figure 7. Sample uses of Webstrates: (a) Collaborative document authoring with different editors personalized at run-time; (b) Multiple devices used to sketch a figure (tablet 1), see it in a print preview (tablet 2), and adjust it in a graphics editor (laptop). (c) Distributed talk controlled remotely by a speaker with a separate interface for audience participation.

GRAPHDECO Project-Team

7. New Results

7.1. Computer-Assisted Design with Heterogeneous Representations

7.1.1. BendFields: Regularized Curvature Fields from Rough Concept Sketches

Participants: Adrien Bousseau, Emmanuel Iarussi.

Designers frequently draw curvature lines to convey bending of smooth surfaces in concept sketches. We present a method to extrapolate curvature lines in a rough concept sketch, recovering the intended 3D curvature field and surface normal at each pixel of the sketch (Fig. 4). This 3D information allows us to enrich the sketch with 3D-looking shading and texturing. We first introduce the concept of *regularized curvature lines* that model the lines designers draw over curved surfaces, encompassing curvature lines and their extension as geodesics over flat or umbilical regions. We build on this concept to define the orthogonal cross field that assigns two regularized curvature lines to each point of a 3D surface. Our algorithm first estimates the projection of this cross field in the drawing, which is non-orthogonal due to foreshortening. We formulate this estimation as a scattered interpolation of the strokes drawn in the sketch, which makes our method robust to sketchy lines that are typical for design sketches. Our interpolation relies on a novel smoothness energy that we derive from our definition of regularized curvature lines. Optimizing this energy subject to the stroke constraints produces a dense non-orthogonal 2D cross field, which we then lift to 3D by imposing orthogonality. Thus, one central concept of our approach is the generalization of existing cross field algorithms to the non-orthogonal case. We demonstrate our algorithm on a variety of concept sketches with various levels of sketchiness. We also compare our approach with existing work that takes clean vector drawings as input.

This work is a collaboration with David Bommes from Titane project team at Inria Sophia-Antipolis, now at RWTH Aachen University. It has been published at ACM Transactions on Graphics (TOG) [7].

7.1.2. Line Drawing Interpretation in a Multi-View Context

Participant: Adrien Bousseau.

Many design tasks involve the creation of new objects in the context of an existing scene. Existing work in computer vision only provides partial support for such tasks. On the one hand, multi-view stereo algorithms allow the reconstruction of real-world scenes, while on the other hand algorithms for line-drawing interpretation do not take context into account. Our work combines the strength of these two domains to interpret line drawings of imaginary objects drawn over photographs of an existing scene (Fig. 5). The main challenge we face is to identify the existing 3D structure that correlates with the line drawing while also allowing the creation of new structure that is not present in the real world. We propose a labeling algorithm to tackle this problem, where some of the labels capture dominant orientations of the real scene while a free label allows the discovery of new orientations in the imaginary scene. We illustrate our algorithm by interpreting line drawings for urban planing, home remodeling, furniture design and cultural heritage.

This work is a collaboration with Jean-Dominique Favreau and Florent Lafarge from Titane project team, Inria Sophia-Antipolis. It has been published at the Conference on Computer Vision and Pattern Recognition (CVPR) [14].

7.1.3. WrapIt: Computer-Assisted Crafting of Wire Wrapped Jewelry

Participants: Adrien Bousseau, Emmanuel Iarussi.

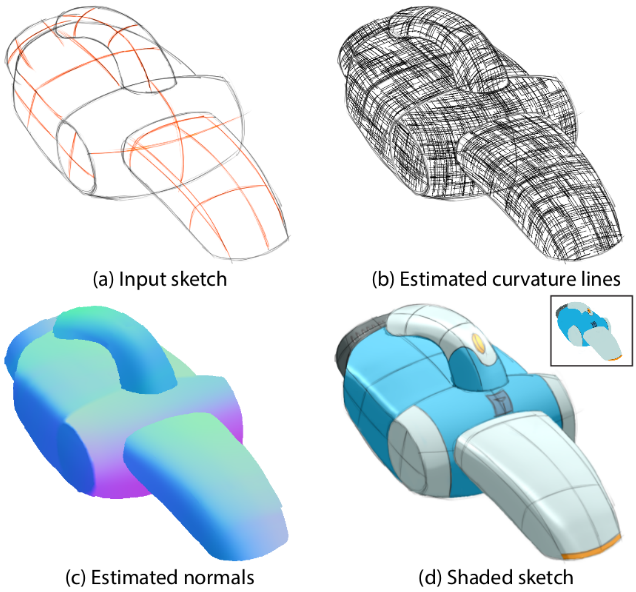


Figure 4. Our method [7] takes as input a rough design sketch with annotated curvature lines (a). We propose a novel smoothness energy to propagate the curvature information to all pixels (b), which allows us to recover surface normals (c) and compute shading (d).

Figure 5. Our method [14] takes as input several photographs of a scene, along with a line drawing of a new object (left). We exploit the dominant orientations of the existing scene to reconstruct the line drawing in 3D (right).

Wire wrapping is a traditional form of handmade jewelry that involves bending metal wire to create intricate shapes. The technique appeals to novices and casual crafters because of its low cost, accessibility and unique aesthetic. We present a computational design tool that addresses the two main challenges of creating 2D wire-wrapped jewelry: decomposing an input drawing into a set of wires, and bending the wires to give them shape (Fig. 6). Our main contribution is an automatic wire decomposition algorithm that segments a drawing into a small number of wires based on aesthetic and fabrication principles. We formulate the task as a constrained graph labeling problem and present a stochastic optimization approach that produces good results for a variety of inputs. Given a decomposition, our system generates a 3D-printed custom support structure, or *jig*, that helps users bend the wire into the appropriate shape. We validated our wire decomposition algorithm against existing wire-wrapped designs, and used our end-to-end system to create new jewelry from clipart drawings. We also evaluated our approach with novice users, who were able to create various pieces of jewelry in less than half an hour.

This work is a collaboration with Wilmot Li from Adobe, San Francisco. The project was initiated by a 3-months visit of Emmanuel Iarussi at Adobe. It has been published at ACM Transactions on Graphics (Proc. SIGGRAPH Asia) [8].

Figure 6. Our system [8] helps novices convert a line drawing (left) into a real piece of jewelry (right).

7.1.4. 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 objects by *novices*. Achieving this goal first requires a deeper understanding of how novices currently generate, explore, and communicate design ideas with traditional tools, i.e., sketches on paper and hands-on prototyping materials. We describe a study framed around two all-day design charrettes where participants perform a complete design process including ideation sketching, concept development and presentation, fabrication planning documentation and collaborative fabrication of hand-crafted prototypes. 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. We observed that while participants had no formal training in design, they made use of advanced visualization techniques to convey 3D concepts. Participants also extensively used physical materials (paper, foam, cardboard) both to support concept exploration and to communicate their ideas to collaborators. We deduce from these observations recommendations for the conception of design tools adapted to the needs and skills of novices.

This work is a collaboration with Wendy McKay, Theophanis Tsandilas and Lora Oehlberg from the InSitu project team - Inria Saclay, in the context of the ANR DRAO project. It is conditionally accepted to ACM CHI 2016.

7.1.5. Vectorizing Rough Line Drawings

Participant: Adrien Bousseau.

Our goal in this project is to convert rough, freehand bitmap sketches to clean vector drawings, keeping three main objectives in mind: (i) the vectorial curves should approximate well the input drawing, (ii) the drawing should be composed of a small number of curves with few control points to preserve the compactness and editability of vector graphics, and (iii) the algorithm should support user guidance to disambiguate the multiple interpretations inherent to artistic inputs. Unfortunately, existing vectorization algorithms only partly satisfy these requirements. In particular, while most methods employ curve fitting to satisfy the first objective of data fidelity, this fitting is performed locally and is often sub-optimal with respect to our second objective of *low complexity*. To achieve our objectives, we propose to cast line drawing vectorization as a global optimization that balances data fidelity with model complexity. We express data fidelity as the goodness of fit of Bézier curve segments, and we express model complexity as the number and degree of curve segments that compose the output drawing. Our algorithm produces clean, compact and editable vector drawings from bitmap sketches.

This ongoing work is a collaboration with Jean-Dominique Favreau and Florent Lafarge from Titane project team, Inria Sophia-Antipolis.

7.1.6. Exploring Design Spaces with Sketch-Based Rendering

Participant: Adrien Bousseau.

Designers often start product design by drawing many quick and imperfect sketches. These sketches typically capture shape variations of a concept from different viewpoints. We introduce *sketch-based rendering* as a way to help designers explore the design space induced by such sketches. Our interactive tool allows designers to interpolate between the sketches, providing a continuous, 3D-like visualization of the concept and its variations without requiring explicit 3D information.

We propose an iterative algorithm to match and warp between sketches using little user interaction. We designed this algorithm to address the specific challenges inherent to concept sketches, in particular the fact that they are dominated by contours rather than color or texture, and that these contours should not distort during interpolation. We also describe how to approximate the relative camera positions of different sketches from the magnitude of their 2D motion fields. This approach allows plausible 3D-like camera movements despite the presence of sketch distortions and variations that prevent standard camera calibration. Our tool, thus, fills a gap in the initial stage of the product design pipeline by allowing designers and their patrons to make better informed choices before proceeding to more expensive 3D modeling and prototyping.

This ongoing work is a collaboration with Ishan Darolia and Vinay Namboodiri from IIT Kampur and Rahul Arora and Karan Singh from University of Toronto.

7.1.7. Sketch-Based Inverse Procedural Modeling

Participant: Adrien Bousseau.

Designing and modeling 3D objects is a crucial skill in various areas of entertainment, science, and engineering. However, this task is notoriously hard and unintuitive, especially for novices. Prior work has addressed the modeling problem from many different directions. Sketch-based modeling exploits human intuition and experience in drawing objects. Nevertheless, the quality of the final 3D model depends on the sketching skills of the user, the amount of details added to the drawing, and ability to resolve inherent ambiguities of the sketching process. Another popular direction is procedural modeling, which has been successfully used to create detailed and complex cities, realistic and growing vegetation, and other man-made objects. But procedural modeling is difficult to control and thus hard to use as an exploratory design tool making it accessible only to experts. Our goal in this project is to leverage both the intuitiveness, freedom and flexibility of sketching and the precision, exactness, and detail amplification of procedural modeling. Users of our system begin to sketch a 3D model using a mouse or a digital pen on a tablet. After only a few strokes, our algorithm finds a compact 3D procedural representation that matches the sketch while augmented it with geometric details.

This ongoing work is a collaboration with Gen Nishida, Bedrich Benes, Ignacio Garcia Dorado and Daniel Aliaga from Purdue University.

7.1.8. A data-based approach to retrieve the viewpoint of a design sketch

Participants: Johanna Delanoy, Adrien Bousseau.

Designing objects requires frequent transitions from a 2D representation, the sketch, to a 3D one. Because 3D modeling is time consuming, it is made only during late phases of the design process. Our long term goal is to allow designers to automatically generate 3D models from their sketches. In this work, we address the preliminary step of recovering the viewpoint under which the object is drawn. We adopt a data-driven approach where we build correspondences between the sketch and 3D objects of the same class from a database. In particular, we relate the curvature lines and contours of the 3D objects to similar lines commonly drawn by designers using histograms of orientation. The 3D objects from the database are then used to vote for the viewpoints and the more probable ones are chosen. Our results on design sketches suggest that using both contours and curvature lines give higher precision than using either one. In particular, curvature information improves viewpoint retrieval when details of the objects are different from the sketch.

The work has been published in the journal Revue Française d'Informatique Graphique and presented at the 28th Journées de l'Association Française d'Informatique Graphique [5].

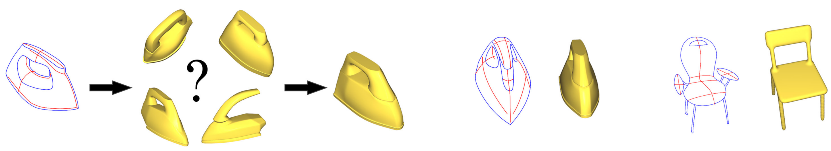


Figure 7. Our method [5] allows to retrieve the viewpoint of a design sketch, using a collection of 3D objects.

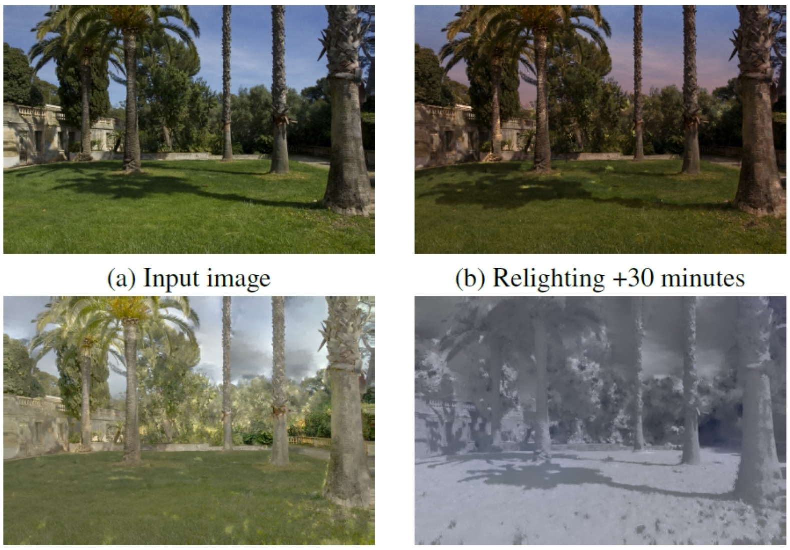
7.2. Graphics with Uncertainty and Heterogeneous Content

7.2.1. Multi-View Intrinsic Images for Outdoors Scenes with an Application to Relighting

Participants: Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Stefan Popov, Adrien Bousseau, George Drettakis.

We introduce a method to compute intrinsic images for a multi-view set of outdoor photos with cast shadows, taken under the same lighting (Fig. 8). We use an automatic 3D reconstruction from these photos and the sun direction as input and decompose each image into reflectance and shading layers, despite the inaccuracies and missing data of the 3D model. Our approach is based on two key ideas. First, we progressively improve the accuracy of the parameters of our image formation model by performing iterative estimation and combining 3D lighting simulation with 2D image optimization methods. Second we use the image formation model to express reflectance as a function of discrete visibility values for shadow and light, which allows us to introduce a robust visibility classifier for pairs of points in a scene. This classifier is used for shadow labelling, allowing us to compute high quality reflectance and shading layers. Our multi-view intrinsic decomposition is of sufficient quality to allow relighting of the input images. We create shadow-caster geometry which preserves shadow silhouettes and using the intrinsic layers, we can perform multi-view relighting with moving cast shadows. We present results on several multi-view datasets, and show how it is now possible to perform image-based rendering with changing illumination conditions.

This work was published in ACM Transactions on Graphics [2].



(a) Input image (b) Relighting +30 minutes (c) Reflectance (d) Shading

Figure 8. Our algorithm [2] decomposes outdoor images (a) into reflectance (c) and shading (d). This decomposition enables relighting of photographs by rendering new shadows in the shading layer (b).

This work is part of an industrial partnership with Autodesk and has been published in ACM Transactions on Graphics [2].

7.2.2. A Bayesian Approach for Selective Image-Based Rendering using Superpixels

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

Many recent Image-Based Rendering (IBR) algorithms have been proposed each having different strengths and weaknesses, depending on 3D reconstruction quality and scene content. Each algorithm operates with a set of hypotheses about the scene and the novel views, resulting in different quality/speed trade-offs in different image regions. We developed a principled approach to select the algorithm with the best quality/speed trade-off in each region. To do this, we propose a Bayesian approach, modeling the rendering quality, the rendering process and the validity of the assumptions of each algorithm. We then choose the algorithm to use with Maximum a Posteriori estimation. We demonstrate the utility of our approach on recent IBR algorithms which use oversegmentation and are based on planar reprojection and shape-preserving warps respectively. Our algorithm selects the best rendering algorithm for each superpixel in a preprocessing step; at runtime our selective IBR uses this choice to achieve significant speedup at equivalent or better quality compared to previous algorithms. The work has been published in the International Conference on 3D Vision (3DV) - 2015 [15].

7.2.3. Uncertainty Modeling for Principled Interactive Image-Based Rendering

Participants: Rodrigo Ortiz Cayon, George Drettakis.

Despite recent advances in IBR methods, they are limited in regions of the scene which are badly or completely unreconstructed. Such regions have varying degrees of uncertainty, which previous solutions treat with heuristic methods. Currently we attempt to develop a comprehensive model of uncertainty for interactive IBR. Regions with high uncertainty would feed an iterative multi-view depth synthesis algorithm. For the rendering we will formalize an unified IBR algorithm, which provides a good quality/speed tradeoff by combining the

Figure 9. In top-left, we use planes fronto-parallel to the input view which fail for trees and slanted planes. Using local plane estimation top-right the result is improved, especially for slanted planes (blue box). Using the shape preserving warp bottom-left of the warping method we previously developed, better results are achieved for the tree (red box), but the quality of the slanted planes is worse. Our algorithm [15] bottom-right makes the correct choice locally, giving the best solution in each case.

advantages of forward warping and depth-based backprojection algorithms and includes plausible stereoscopic rendering for unreconstructed volumetric regions.

7.2.4. Multi-view Inpainting

Participants: Theo Thonat, George Drettakis.

We are developing a new approach for removing objects in multi-view image datasets. For a given target image from which we remove objects, we use Image-Based Rendering for reprojecting the other images into the target and for regions not visible in any other image we use inpainting techniques. The difficulties reside in formalizing the unified algorithm and enforcing multi-view consistency. This is an ongoing project in collaboration with Adobe Research (E. Shechtman and S. Paris).

7.2.5. 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. Then we define an algorithm for maximal-entropy sampling of the parameters of the noise functions so as to meet prescribed statistics to reproduce. 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. We are currently investigating a learning method so as to inject into the model the appropriate prescribed statistics deduced from an input exemplar image.

This ongoing work is a collaboration with Ian Jermyn from Durham University and will be submitted for publication in 2016.

7.2.6. Unifying Color and Texture Transfer for Predictive Appearance Manipulation

Participants: Fumio Okura, Kenneth Vanhoey, Adrien Bousseau, George Drettakis.

Recent color transfer methods use local information to learn the transformation from a Source to an Exemplar image, and then transfer this appearance change to a Target image (figure 10 (a) to (d)). These solutions achieve successful results for general mood changes, e.g., changing the appearance of an image from “sunny” to “overcast”. However, they fail to create new image content, such as leaves on a bare tree (figure 10 (d)). Texture transfer, on the other hand, can synthesize such content but tends to destroy image structure (figure 10 (e)). We propose the first algorithm that unifies color and texture transfer, outperforming both by automatically leveraging their respective strengths (figure 10 (f)). A key novelty in our approach resides in teasing apart appearance changes that can be modeled simply as changes in color versus those that require new image content to be generated. Our method starts with an analysis phase which evaluates the success of color transfer on the Source/Exemplar scene. To do so, color transfer parameters are learned on this pair, and applied on the Source. The color transferred Source image is then evaluated against the Exemplar which serves as a ground truth, using texture distance metrics (textons in our case). This provides information on the localization of success and failure of color transfer on this scene. This analysis then drives the synthesis: a selective, iterative texture transfer algorithm that simultaneously predicts the success of color transfer on the Target and synthesizes new content using texture transfer where needed. Synthesis exploits a dense pixel matching between the Source/Exemplar scene, on which information is learned, and the Target/Output scene, on which we want to synthesize. The algorithm iterates between synthesizing the new scene by locally using either color or texture transfer, and improving the dense matching on the scene being synthesized. As a result, it leverages the best of both techniques on a variety of scenes by transferring large temporal changes between photographs, such as change of season and flooding. We demonstrate this with seasonal changes on vegetation (e.g., trees) and snow, and on examples involving flooding.

This work is a collaboration with Alexei Efros from UC Berkeley in the context of the associate team CRISP. It has been published in Computer Graphics Forum [9] and was accepted and presented at the Eurographics symposium on Rendering.

Figure 10. Illustration of our appearance prediction method [9]. The future appearance of a target image (c) is predicted (f) based on the knowledge learned from a quasi-aligned source-exemplar pair ((a) and (b)) which characterizes an analogous transformation. The key insight is to selectively operate color transfer ((d): only operate rigid local color histogram transformations, i.e., change the background's overall mood) or texture transfer ((e): copying pixels or patches from the exemplar, i.e. synthesize the three's leaves) where suitable, so as to obtain an improved result (f).

7.2.7. Simplification of Triangle Meshes with Digitized Radiance

Participant: Kenneth Vanhoey.

Very accurate view-dependent surface color of virtual objects can be represented by outgoing radiance of the surface. This data forms a surface light field, which is inherently 4-dimensional, as the color is varying both spatially and directionally. Acquiring this data and reconstructing a surface light field of a real-world object can result in very large datasets, which are very realistic, but tedious to store and render. In this project, we consider the processing of outgoing radiance stored as a vertex attribute of triangle meshes, and especially propose a principled simplification technique. We show that when reducing the global memory footprint of such acquired objects, smartly reducing the spatial resolution, as opposed to the directional resolution, is an effective strategy for overall appearance preservation. To define such simplification, we define a new metric to guide an iterative edge collapse algorithm. Its purpose is to measure the visual damage introduced when operating a local simplification. Therefore, we first derive mathematical tools to calculate with radiance functions on the surface: interpolation, gradient computation and distance measurements. Then we derive a metric using these tools. We particularly ensure that the mathematical interpolation used in the metric is coherent with the non-linear interpolation we use for rendering, which makes the math coherent with the rendered object. As a result we show that both synthetic and acquired objects benefit from our radiance-aware simplification process: at equal memory footprint, visual quality is improved compared to state of the art alternatives.

This work is a collaboration with the ICube laboratory, Strasbourg, France. It was published in the Computer Graphics Journal [11] and was accepted and presented at Computer Graphics International 2015 in Strasbourg, France.

7.2.8. Video based rendering for vehicles

Participants: Abdelaziz Djelouah, Georgios Koulieris, George Drettakis.

The main objective of image based rendering methods is to provide high quality free-view point navigation in 3D scenes using only a limited set of pictures. Despite the good visual quality achieved by most recent methods, the results still look unrealistic because of the static nature of the rendered scenes. This project is in the general context of enriching image based rendering experience by adding dynamic elements and we are particularly interested by adding vehicles.

Vehicles represent an important proportion of the dynamic elements in any urban scene and adding an object with such a complex appearance model has many challenges. First, contrary to classic IBR the number of viewpoints is limited because all input videos must be recorded at the same time. Also, because of this limited number of viewpoints, using classic multi-view reconstruction methods does not produce good results. Instead we use 3D stock models as proxy for the cars. The first step is the registration of the 3D model with the input videos. Then, using the 3D model, the input videos are processed to extract the different visual layers (base color, reflections, transparency, etc.). Finally, the objective is to find the appropriate way to combine the 3D model and the extracted layers to provide the most realistic image from any viewpoint.

This ongoing work is a collaboration with Gabriel Brostow from University College London in the context of the CR-PLAY EU project and with Alexei Efros from UC Berkeley.

7.2.9. Finger-Based Manipulation in Immersive Spaces and the Real World

Immersive environments that approximate natural interaction with physical 3D objects are designed to increase the user's sense of presence and improve performance by allowing users to transfer existing skills and expertise from real to virtual environments. However, limitations of current Virtual Reality technologies, e.g., low-fidelity real-time physics simulations and tracking problems, make it difficult to ascertain the full potential of finger-based 3D manipulation techniques. This project decomposes 3D object manipulation into the component movements, taking into account both physical constraints and mechanics. We fabricate five physical devices that simulate these movements in a measurable way under experimental conditions. We then implement the devices in an immersive environment and conduct an experiment to evaluate direct finger-based against ray-based object manipulation. The key contribution of this work is the careful design and creation of physical and virtual devices to study physics-based 3D object manipulation in a rigorous manner in both real and virtual setups.

This work was presented at IEEE Symposium on 3D User Interfaces [12], and is in collaboration with the EXSITU Inria group in Paris (T. Tsandilas, W. Mackay, L. Oehlberg).

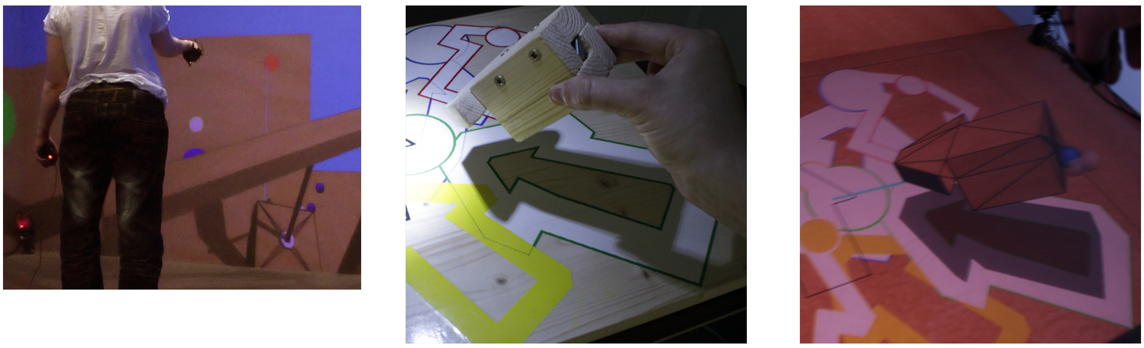


Figure 11. A user in our immersive environment (left) for finger-based manipulation [12]. Completing a 6 DoF manipulation task in real (center) and virtual (right) settings.

7.2.10. Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation

Participants: Georgios Koulieris, George Drettakis.

Comfortable, high-quality 3D stereo viewing is becoming a requirement for interactive applications today. The main challenge of this project is 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 train a classifier to learn these correlations using an eye-tracker which provides the ground-truth object being looked at. The classifer is used at runtime to predict object category – and thus gaze – during game play, based on the current state of game variables. We use this prediction to propose a dynamic disparity manipulation method, which provides rich and comfortable depth. We evaluate 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 is a collaboration with the Technical University of Crete (K. Mania) and Cottbuss University (D. Cunningham), and will be presented at IEEE VR 2016.

7.2.11. Compiling High Performance Recursive Filters

Infinite impulse response (IIR) or recursive filters, are essential for image processing because they turn expensive large-footprint convolutions into operations that have a constant cost per pixel regardless of kernel size. However, their recursive nature constrains the order in which pixels can be computed, severely limiting both parallelism within a filter and memory locality across multiple filters. Prior research has developed algorithms that can compute IIR filters with image tiles. Using a divide-and-recombine strategy inspired by parallel prefix sum, they expose greater parallelism and exploit producer-consumer locality in pipelines of IIR filters over multi-dimensional images. While the principles are simple, it is hard, given a recursive filter, to derive a corresponding tile-parallel algorithm, and even harder to implement and debug it. We show that parallel and locality-aware implementations of IIR filter pipelines can be obtained through program transformations , which we mechanize through a domain-specific compiler. We show that the composition of a small set of transformations suffices to cover the space of possible strategies. We also demonstrate that the tiled implementations can be automatically scheduled in hardware-specific manners using a small set of generic heuristics. The programmer specifies the basic recursive filters, and the choice of transformation requires only a few lines of code. Our compiler then generates high-performance implementations that are an order of magnitude faster than standard GPU implementations, and outperform hand tuned tiled implementations of specialized algorithms which require orders of magnitude more programming effort – a few lines of code instead of a few thousand lines per pipeline. This work was presented the High Performance Computing conference and is a collaboration with F. Durand, J. Ragan-Kelley and G. Chaurasia of MIT and S. Paris of Adobe [13].

7.2.12. Probabilistic Connections for Bidirectional Path Tracing

Participants: Sefan Popov, George Drettakis.

Bidirectional path tracing (BDPT) with Multiple Importance Sampling is one of the most versatile unbiased rendering algorithms today. BDPT repeatedly generates sub-paths from the eye and the lights, which are connected for each pixel and then discarded. Unfortunately, many such bidirectional connections turn out to have low contribution to the solution. The key observation in this projects is that we can importance sample connections to an eye sub-path by considering multiple light sub-paths at once and creating connections probabilistically. We do this by storing light paths, and estimating probability mass functions of the discrete set of possible connections to all light paths. This has two key advantages: we efficiently create connections with low variance by Monte Carlo sampling, and we reuse light paths across different eye paths. We also introduce a caching scheme by deriving an approximation to sub-path contribution which avoids high-dimensional path distance computations. Our approach builds on caching methods developed in the different context of VPLs. Our Probabilistic Connections for Bidirectional Path Tracing approach raises a major challenge, since reuse results in high variance due to correlation between paths. We analyze the problem of path correlation and derive a conservative upper bound of the variance, with computationally tractable sample weights. We present results of our method which shows significant improvement over previous unbiased global illumination methods, and evaluate our algorithmic choices.

This work was in collaboration with R. Ramamoorthi (UCSD) and F. Durand (MIT) and appeared in the Eurographics Symposium on Rendering [10].

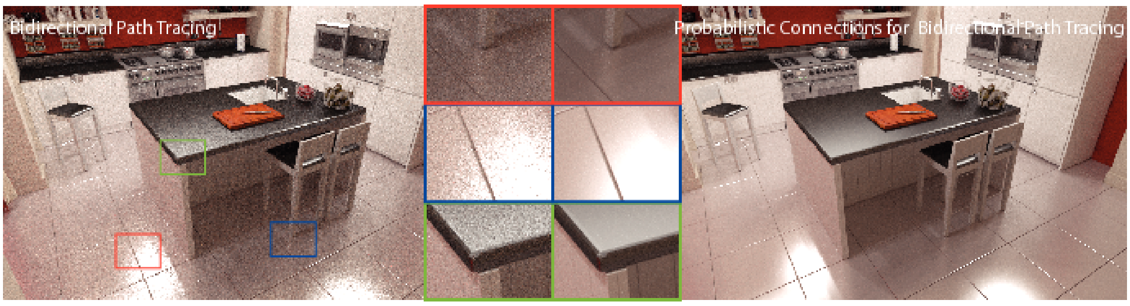


Figure 12. Our Probabilistic Connections for Bidirectional Path Tracing [10] approach importance samples connections to an eye sub-path, and greatly reduces variance, by considering and reusing multiple light sub-paths at once. Our approach (right) achieves much higher quality than bidirectional path-tracing on the left for the same computation time.

HYBRID Project-Team

7. New Results

7.1. 3D User Interfaces

7.1.1. Novel 3D Interactive Techniques

THING: Introducing a Tablet-based Interaction Technique for Controlling 3D Hand Models Merwan Achibet, Anatole Lécuyer and Maud Marchal

The hands of virtual characters are highly complex 3D models that can be tedious and time-consuming to animate with current methods. We introduced the *THING* [17], a novel tablet-based approach that leverages multi-touch interaction for a quick and precise control of a 3D hand's pose 2 . The flexion/extension and abduction/adduction of the virtual fingers can be controlled for each finger individually or for several fingers in parallel through sliding motions on the surface of the tablet. We designed two variants of THING: **(1)***MobileTHING*, which maps the spatial location and orientation of the tablet to that of the virtual hand, and **(2)***DesktopTHING*, which combines multi-touch controls of fingers with traditional mouse controls for the global position and orientation of the hand model. We compared the usability of THING against mouse-only controls and a data glove in two controlled experiments. Results show that DesktopTHING was significantly preferred by users while providing performance similar to data gloves. Together, these results could pave the way to the introduction of novel hybrid user interfaces based on tablets and computer mice in future animation pipelines. This work was done in collaboration with Géry Casiez (Inria team MJOLNIR).

Figure 2. THING enables the control of 3D hand models (in blue) by sliding fingers along sliders arranged in a morphologically-consistent pattern on the tablet's screen. This creates a strong correspondence between user's input and pose of the controlled hand. Here, the user closes the virtual hand and then points the index finger.

Plasticity for 3D User Interfaces: New Models for Devices and Interaction Techniques Jérémy Lachoche and Bruno Arnaldi

We have introduced new models for device and interaction techniques to overcome plasticity limitations in Virtual Reality (VR) and Augmented Reality (AR) [26]. We aimed to provide developers with solutions to use and create interaction techniques that fit to the 3D application tasks and to the input and output devices available. The device model describes input and output devices and includes capabilities, limitations and representations in the real world. We also propose a new way to develop interaction techniques with an approach based on PAC and ARCH models [43]. These techniques are implemented independently from the specific devices used thanks to the proposed device model. Moreover, our approach aims to facilitate the portability of interaction techniques over different target OS and 3D frameworks. This work was done in collaboration with Thierry Duval (Lab-STICC), Éric Maisel (ENIB) and Jérome Royan (IRT B-Com).

Dealing with Frame Cancellation for Stereoscopic Displays in 3D User Interfaces Jérémy Lacoche, Morgan Le Chénéchal, Valérie Gouranton and Bruno Arnaldi

We explored new methods to reduce ocular discomfort when interacting with stereoscopic content, focusing on frame cancellation [27]. Frame cancellation appears when a virtual object in negative parallax (front of the screen) is clipped by the screen edges; stereopsis cue lets observers perceive the object popping-out from the screen while occlusion cue provides observers with an opposite signal. Such a situation is not possible in the real world. This explains some visual discomfort for observers and leads to a poor depth perception of the virtual scene. This issue is directly linked to the physical limitations of the display size that may not cover the entire field of view of the observer. To deal with these physical constraints we introduce two new methods in the context of interactive applications. The first method consists in two new rendering effects based on progressive transparency that aim to preserve the popping-out effect of the stereo. The second method focuses on adapting the interaction of the user, not allowing him to place virtual objects in an area subject to frame cancellation. This work was done in collaboration with Sébastien Chalmé (IRT B-Com), Thierry Duval (Lab-STICC) and Éric Maisel (ENIB).

7.1.2. Understanding Human Perception in VR

Distance Estimation in Large Immersive Projection Systems, Revisited Ferran Argelaguet and Anatole Lécuyer

When walking within an immersive projection environment, accommodation distance, parallax and angular resolution vary according to the distance between the user and the projection walls which can influence spatial perception. As CAVE-like virtual environments get bigger, accurate spatial perception within the projection setup becomes increasingly important for application domains that require the user to be able to naturally explore a virtual environment by moving through the physical interaction space. In this work we performed two experiments which analyze how distance estimation is biased when accommodation distance, parallax and angular resolution vary [23]. The experiments were conducted in a large immersive projection setup with up to ten meter interaction range. The results showed that both accommodation distance and parallax have a strong asymmetric effect on distance judgments. We found an increased distance underestimation for positive parallax conditions as the accommodation-convergence difference increased. In contrast, we found less distance overestimation for negative and zero parallax conditions. Our findings also showed that angular resolution has a negligible effect on distance estimation. This work was done in collaboration with Anne-Hélène Olivier (MIMETIC) and Gerd Bruder (University of Hamburg).

Virtual Proxemics: Locomotion in the Presence of Obstacles in Large Immersive Projection Environments Ferran Argelaguet, Anatole Lécuyer

In the real world we navigate with ease by walking in the presence of obstacles, we develop avoidance strategies and behaviors which govern the way we locomote in the proximity of physical objects and other persons during everyday tasks. With the advances of virtual reality technology, it becomes important to gain an understanding of how these behaviors are affected in a virtual reality application. In this work, we analyzed the walking and collision avoidance behavior when avoiding real and virtual static obstacles [19]. In order to generalize our study, we considered both anthropomorphic and inanimate objects, each having his virtual and real counterpart. The results showed that users exhibit different locomotion behaviors in the presence of real and virtual obstacles, and in the presence of anthropomorphic and inanimate objects. Precisely, the results showed a decrease of walking speed as well as an increase of the clearance distance (i. e., the minimal distance between the walker and the obstacle) when facing virtual obstacles compared to real ones. Moreover, our results suggest that users act differently due to their perception of the obstacle: users keep more distance when the obstacle is anthropomorphic compared to an inanimate object and when the orientation of anthropomorphic obstacle is from the profile compared to a front position. We discussed implications on future large shared immersive projection spaces. This work was done in collaboration with Anne-Héléne Olivier (MIMETIC), Julien Pettré (MIMETIC) and Gerd Bruder (University of Hamburg).

7.1.3. Sports and Virtual Reality

A Methodology for Introducing Competitive Anxiety and Pressure in VR Sports Training Ferran Argelaguet and Anatole Lécuyer

Athletes' performance is influenced by internal and external factors, including their psychological state and environmental factors, especially during competition. As a consequence, current training programs include stress management. In this work, we explored whether highly immersive systems can be used for such training programs [11]. First, we proposed methodological guidelines to design sport training scenarios both on considering the elements that a training routine must have, and how external factors might influence the participant. The proposed guidelines are based on flow and social-evaluative threat theories. Second, to illustrate and validate our methodology, we designed an experiment reproducing a 10m Olympic pistol shooting competition 3 . We analyzed whether changes in the environment are able to induce changes in user performance, physiological responses and the subjective perception of the task. The simulation included stressors in order to raise a social-evaluative threat, such as aggressive public behavior or unforced errors, increasing the pressure while performing the task. The results showed significant differences in the user behavior and in their subjective impressions, trends in the physiological data were also observed. Taken together our results suggest that highly immersive systems could be further used for training systems in sports. This work was done in collaboration with Frank Multon (MIMETIC).

Figure 3. The proposed methodology was illustrated and evaluated in a virtual Olympic shooting experiment. The experiment was conducted in a wide immersive projection system being able to enclose a ten meter wide shooting range with six virtual opponents and one participant.

7.1.4. Experiencing the Past in Virtual Reality

An Immersive Virtual Sailing on the 18 th -Century Ship Le Boullongne Jean-Baptiste Barreau, Florian Nouviale and Valérie Gouranton

This work is the result of the collaboration between historians and computer scientists whose goal was the digital reconstitution of "Le Boullongne", an 18th-century merchant ship of "La Compagnie des Indes orientale" [12]. This ship has now disappeared and its reconstitution aims at understanding on-board living conditions. Three distinct research laboratories have participated in this project so far. The first, a department of naval history, worked on historical documents, especially the logbooks describing all traveling events of the ship. The second, a research laboratory in archeology, archaeoscience and history, proposed a 3D model of the ship based on the original naval architectural plans. The third, a computer science research laboratory, implemented a simulation of the ship sailing in virtual reality. This work focuses on the reconstitution of the ship in virtual reality, aiming at restoring a realistic interactive naval simulation: the 3D model of the ship has been integrated in an ocean simulation, with a physical rendering of the buoyancy. The simulation allows a user to walk around on the ship, at a scale of 1:1, and even steer it through a natural interaction. Several characteristics of the simulation reinforce the sensation of being on-board: (1) A sonic environment mixing spatialized sounds (gulls flying, a whale swimming, wood cracking, cannons firing) and global soundscape (ocean and wind). (2) The meteorology of the simulation is dynamically modifiable; the user can increase the swell height and speed. The global illumination and wind sound vary in accordance with these parameters. The buoyancy simulation entails realistic movements of the ship. (3) Several interactions are proposed allowing the user to steer the ship with his/her hand, walk around on the ship, fire the cannons, and modify the weather. (4) Three animated sailors accompany the user in his/her sailing experience. They are wearing realistic period costumes. The immersive simulation has allowed historians to embark on "Le Boullongne" and to better understand how life was organized on-board. It has also been presented at several public exhibitions, in CAVE-like structures and HMD. This work was done in collaboration with Ronan Gaugne (Univ. Rennes 1), Yann Bernard (CReAAH) and Sylviane Llinares (CERHIO, UBS Lorient).

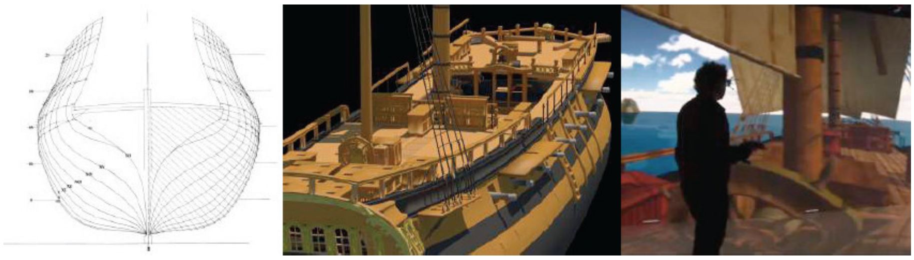


Figure 4. Digital reconstitution of "Le Boullongne". From architectural plans to virtual reality implementation.

Touching and interacting with inaccessible cultural heritage Valérie Gouranton and Bruno Arnaldi

Sense of touch provides a particular access to our environment, enabling a tangible relation with it. In the particular use case of cultural heritage, touching the past, apart from being a universal dream, can provide essential information to analyze, understand, or restore artifacts. However, archaeological objects cannot always offer a tangible access, either because they have been destroyed or too damaged, or because they are part of a larger assembly. In other cases, it is the context of use that has become inaccessible, as it is related to an extinct activity. In [15] we proposed a workflow based on a combination of computed tomography, 3D images, and 3D printing to provide concrete access to cultural heritage, and we illustrate this workflow in different contexts of inaccessibility. These technologies are already used in cultural heritage, but seldom combined, and mostly for exceptional artifacts. We proposed to combine these technologies in case studies corresponding to relevant archaeological situations.

This work was done in collaboration with Théophane Nicolas (INRAP), Ronan Gaugne (Univ. Rennes 1), Cédric Tavernier (Image ET) and Quentin Petit (CNRS).

3D reconstruction of the loyola sugar plantation and virtual reality applications Jean-Baptiste Barreau, Valérie Gouranton

Discovered in 1988, the Loyola sugar plantation, owned by the Jesuits in French Guiana, is a major plantation of colonial history and slavery. Ongoing archaeological excavations have uncovered the Jesuit's house and the outbuildings usually associated with a plantation such as a chapel and its cemetery, a blacksmith shop, a pottery, the remains of the entire sugar production (a windmill, a boiler and a dryer), coffee and indigo warehouses etc. Based on our findings and our network with 3D graphic designers and researchers in virtual reality, a 3D restitution integrated within a virtual reality platform was initiated to develop a better understanding of the plantation and its surrounding landscape. A specific work on the interactive changes of sunlight and animal sounds aimed to reconstruct a coherent evolution during one day of the site's environment [21].

This work was done in collaboration with Quentin Petit (CNRS),Yann Bernard (CReAAH), Reginald Auger (Laval University, Canada), Yannick Le Roux (Laval University, French Guiana) Ronan Gaugne (IMMERSIA), and Cédric Tavernier (Image ET).

7.2. Physically-Based Simulation and Multisensory Feedback

7.2.1. Interactive Physically-Based Simulation

Aggregate constraints for virtual manipulation with soft fingers, Maud Marchal, Anthony Talvas

Figure 5. Interaction with deformable fingers generates many interconnected contact points which are expensive to solve with friction. Our approach aggregates contact constraints per phalanx with torsional friction. The subsequent increase in performance allows real time dexterous manipulation of virtual objects using soft fingers.

Interactive dexterous manipulation of virtual objects remains a complex challenge that requires both appropriate hand models and accurate physically-based simulation of interactions. In [16], we proposed an approach based on novel aggregate constraints for simulating dexterous grasping using soft fingers. Our approach aims at improving the computation of contact mechanics when many contact points are involved, by aggregating the multiple contact constraints into a minimal set of constraints. We also introduced a method for non-uniform pressure distribution over the contact surface, to adapt the response when touching sharp edges. We used the Coulomb-Contensou friction model to efficiently simulate tangential and torsional friction. We showed through different use cases that our aggregate constraint formulation is well-suited for simulating interactively dexterous manipulation of virtual objects through soft fingers, and efficiently reduces the computation time of constraint solving. This work was done in collaboration with Christian Duriez (Inria team DEFROST) and Miguel Otaduy (Univ. Rey Juan Carlos, Madrid, Spain).

7.2.2. Multimodal Feedback

Elastic-Arm: Human-scale passive feedback for augmenting interaction and perception in virtual environments Merwan Achibet, Adrien Girard, Maud Marchal, Anatole Lécuyer

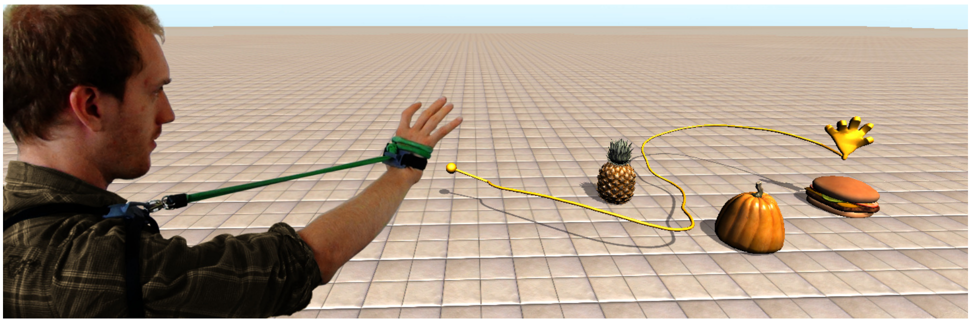


Figure 6. The Elastic-Arm is a body-mounted armature that provides egocentric passive haptic feedback. It presents an alternative to more complex active haptic devices that are generally less adapted to large immersive environments. In this example, a user performs a selection task by stretching his virtual arm using a combination of the Bubble and Go-Go techniques reimplemented with our system.

Haptic feedback is known to improve 3D interaction in virtual environments but current haptic interfaces remain complex and tailored to desktop interaction. In [18], we introduced the *ElasticArm*, a novel approach 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 her shoulder. As a result, a progressive resistance force is perceived when extending the arm. This haptic feedback can be incorporated with various 3D interaction techniques and we illustrate the possibilities offered by our system through several use cases based on well-known examples such as the *Bubble* technique, *Redirected Touching*, and pseudo-haptics. These illustrative use cases provide users with haptic feedback during selection and navigation tasks but they also enhance their perception of the virtual environment. Taken together, these examples suggest that the Elastic-Arm can be transposed in numerous applications and with various 3D interaction metaphors in which a mobile haptic feedback can be beneficial. It could also pave the way for the design of new interaction techniques based on human-scale egocentric haptic feedback.

Visual vibrations to simulate taps on different materials Maud Marchal, Anatole Lécuyer

In [40], we presented a haptic visualization technique for conveying material type through visual feedback, expressed as visible decaying sinusoidal vibration resulting from tapping an object. The technique employs cartoon-inspired visual effects and modulates the scale of the vibration to comply with visual perception. The results of a user study showed that participants could successfully perceive three types of material (rubber, wood, and aluminum) using our novel visual effect. This work was done in collaboration with Taku Hachisu and Hiroyuki Kajimoto (Univ. of Electro Communication, Tokyo, Japan).

7.2.3. GPU-based Collision Detection in Virtual Environments

GPU Ray-Traced Collision Detection: Fine Pipeline Reorganization François Lehericey, Valérie Gouranton, Bruno Arnaldi

Ray-tracing algorithms can be used to render a virtual scene and to detect collisions between objects. Numerous ray-tracing algorithms have been proposed which use data structures optimized for specific cases (rigid objects, deformable objects, etc.). Some solutions try to optimize performance by combining several algorithms to use the most efficient algorithm for each ray. In [31], we presented a ray-traced collision detection pipeline that improves the performance on a graphic processing unit (GPU) when several ray-tracing algorithms are used.

When combining several ray-tracing algorithms on a GPU, a well-known drawback is thread divergence among work-groups that can cause loss of performance by causing idle threads. We avoid branch divergence by dividing the ray tracing into three steps with appended buffers in between. We also show that prediction can be used to avoid unnecessary synchronizations between the CPU and GPU. Applied to a narrow-phase collision detection algorithm, results show an improvement of performance up to 2.7 times.

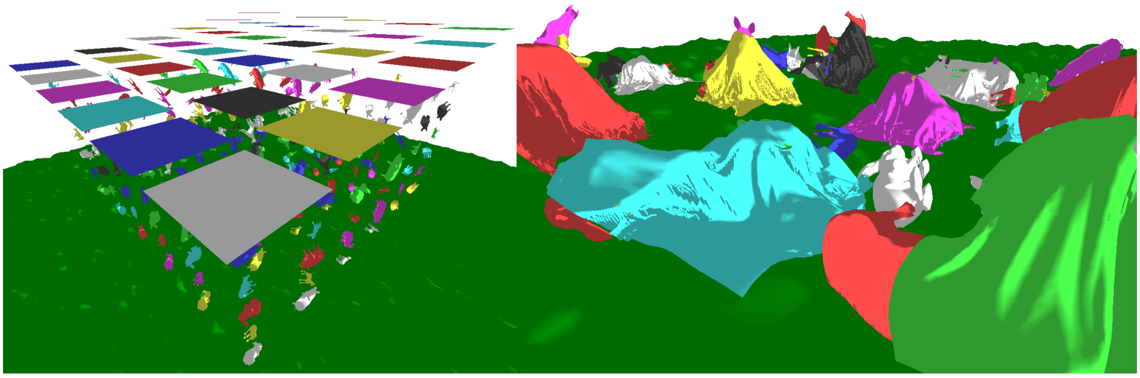


Figure 7. 216 concave objects fall on an irregular ground and 36 deformable sheets fall over them [31].

GPU Ray-Traced Collision Detection for Cloth Simulation François Lehericey, Valérie Gouranton, Bruno Arnaldi

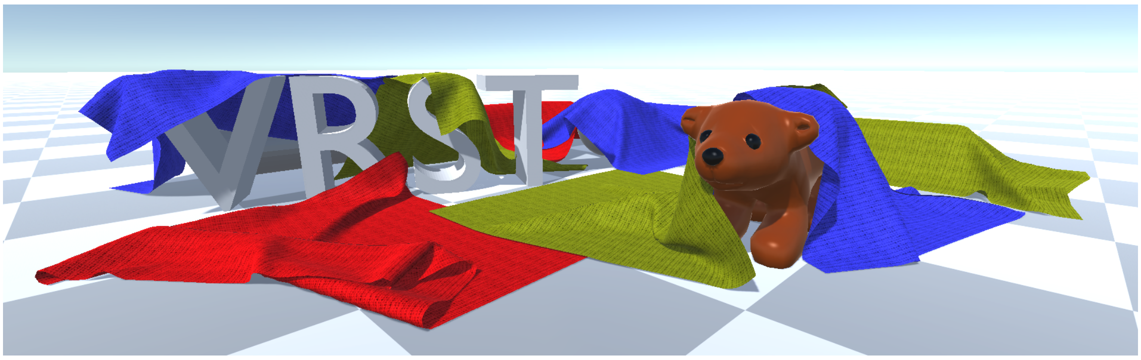


Figure 8. Our method can perform collision detection between clothes and handle self collision detection [30].

In [30], we proposed a method to perform collision detection with cloths with ray-tracing at an interactive frame-rate. Our method is able to perform collision detection between cloths and volumetric objects (rigid or deformable) as well as collision detection between cloths (including auto-collision). Our method casts rays between objects to perform collision detection, and an inversion-handling algorithm is introduced to correct errors introduced by discrete simulations. GPU computing is used to improve the performances by parallelizing the ray-tracing. Our implementation handles scenes containing deformable objects at an interactive frame-rate, with collision detection lasting a few milliseconds.

7.2.4. Medical Applications

Real-time tracking of deformable targets in 3D ultrasound images Maud Marchal

In [35], [36], we presented a novel approach for tracking a deformable anatomical target within 3D ultrasound volumes. Our method is able to estimate deformations caused by the physiological motions of the patient. The displacements of moving structures are estimated from an intensity-based approach combined with a physically-based model and has therefore the advantage to be less sensitive to the image noise. Furthermore, our method does not use any fiducial marker and has real-time capabilities. The accuracy of our method is evaluated on real data acquired from an organic phantom. The validation is performed on different types of motions comprising rigid and non-rigid motions. Thus, our approach opens novel possibilities for computer-assisted interventions where deformable organs are involved.

Our approach was also evaluated on the MICCAI CLUST'15 challenge 3D database. We achieved a mean tracking error of 1.78 mm with an average computation time of 350 ms per frame, ranking our method first during the on-site challenge [34]. This work was done in collaboration with Lucas Royer, Anthony Le Bras and Guillaume Dardenne (IRT bcom), and Alexandre Krupa (Inria team LAGADIC).

Statistical study of parameters for deep brain stimulation automatic pre-operative planning of electrodes trajectories Maud Marchal

Automatic methods for pre-operative trajectory planning of electrodes in Deep Brain Stimulation are usually based on the search for a path that resolves a set of surgical constraints to propose an optimal trajectory. In [13], we studied the use of parameters based on real trajectories of surgeons. For that purpose we firstly retrieve the actual weighting factors used by neurosurgeons thanks to a retrospective study, secondly we compare the results from two different hospitals to evaluate their similarity, and thirdly we compare these trends to the weighting factors usually empirically set in most current approaches. We proposed two approaches, one based on a stochastic sampling and the other on an exhaustive search. In each case, we get a sample of combinations of weighting factors along with a measure of their quality, i.e. the similarity between the optimal trajectory they lead to and the trajectory manually planned by the surgeon as a reference. Then visual and statistical analysis are performed on the number of occurrences and on the rank means. We performed our study on 56 retrospective cases from two different hospitals.We could observe a trend of the occurrence of each weight on the number of occurrences. We also proved that each weight had a significant influence on the ranking. Additionally, we observed no influence of the medical center parameters, suggesting that the trends were comparable in both hospitals. Finally, the obtained trends were confronted to the usual weights chosen by the community, showing some common points but also some discrepancies. These results tend to show a predominance of the choice of a trajectory close to a standard direction. Secondly, the avoidance of the vessels or sulci seems to be sought in the surroundings of the standard position. The avoidance of the ventricles seem to be less predominant, but this could be due to the already reasonable distance between the standard direction and the ventricles. The similarity of results between two medical centers tend to show that it is not an exceptional practice. This work was done in collaboration with Caroline Essert and Antonio Capobianco (Univ. Strasbourg), Claire Haegelen and Pierre Jannin (LTSI, Rennes), Sara Fernandez-Vidal, Carine Karachi and Eric Bardinet (Institut du Cerveau et de la Moëlle Epinière, Paris).

7.3. Collaborative Virtual Environments

Asymetric Remote Collaboration in Mixed Reality: Awareness and Navigation Morgan Le Chénéchal, Valérie Gouranton and Bruno Arnaldi

We first focused on the lack of mutual awareness that may appear in many situations and we evaluated different ways to present the distant user and his actions in the Virtual Environment (VE) in order to understand his perception and cognitive process. We focused on a common case consisting in estimating accurately the time at which a distant user analyzed the meaning of a remotely pointed object. Amongst others, our experimental results presented at CTS [28], show that expertise of the users influences on how they estimate the distant activity and the type of applied strategies.

Then, in a similar asymmetric setup, we proposed a demo at IEEE VR to deal with real estate business. In this context, it is quite difficult for estate agents to make customers understand the potential and the volumes of free spaces. The demo aimed to solve these issues based on a laying out scenario in which a seller and a customer collaborate. As the roles of both users are different, we proposed an asymmetric collaboration where the two users do not use the same interaction setup and do not benefit from the same interaction capabilities.

Last, we focused on a remote collaborative maintenance scenario in which a remote expert helps an operator in performing a physical task 9 . Our system is based on a VR setup for the remote expert in order to virtually co-locates him in the real workspace, and an AR interface for the display of the helping gestures to the agent. In a preliminary user study, we evaluated the performance of our system in a navigation task, and we presented results at ICAT-EGVE [29].

This work was done in collaboration with Thierry Duval (Lab-STICC) and Jérome Royan (IRT B-Com).

Figure 9. Remote collaborative maintenance using mixed reality.

High-Level Components for Developing Collaborative and Interactive Virtual Environments Rozenn Bouville, Valérie Gouranton, Thomas Boggini, Florian Nouviale and Bruno Arnaldi

We proposed a framework called #FIVE (Framework for Interactive Virtual Environments) for the development of interactive and collaborative virtual environments [22]. It has been developed for an easier and a faster design and development of virtual reality applications. It was designed with a constant focus on re-usability with as few hypotheses as possible on the final application in which it could be used. Whatever the chosen implementation for the Virtual Environment (VE), #FIVE : (1) provides a toolkit that eases the declaration of possible actions and behaviours of objects in the VE, (2) provides a toolkit that facilitates the setting and the management of collaborative interactions in a VE, (3) is compliant with distribution of the VE on different setups and (4) proposes guidelines to efficiently create a collabo-rative and interactive VE. It is composed of several modules, among them, two core modules : the relation engine and the collaborative interaction engine. On the one hand, the relation engine manages the relations between the objects of the environment. On the other hand, the collaborative interaction engine manages how users can collaboratively control objects. The modules that compose the #FIVE framework can be used either independently or simultaneously , depending on the requirements of the application. They can also communicate and work with other modules thanks to an API. For instance, a scenario engine can be plugged to any or both of the #FIVE modules if the application is scenario-based. #FIVE has already been used in VR applications by several members of our team (see section 6.5). The feedbacks are rather positive and we intend to further develop #FIVE with additional functionalities, notably by extending it to the control of avatars whether they are controlled by a user or by the system.

High-Level Components for Developing Collaborative Scenarios Guillaume Claude, Valérie Gouranton and Bruno Arnaldi

We were interested in the description of activities of actors in Collaborative Virtual Environments for Training to team working on procedures. We have proposed #SEVEN, a model for the description of procedures as

Collaborative Virtuel Environments Scenarios (see also section 6.6). In [25] we have demonstrated the abilities of this model to be adapted to a wide range of use cases. We showed that it can adapt its abstraction level to the required guidance level and describe more or less complex unfolding of events. In [24] we have provided a novel approach to the distribution of the actions between the actors of the simulation by using an action filtering model in conjunction with a reactive team model. The action filtering model uses data about the actors such as their abilities or their rights. Our reactive team model can be used to define relationships between the team members and the effects of inner rules of the team upon the involvement of the actors in the procedure. To our knowledge, our solution is the closest to the existing models proposed by the social science domain known as role theory. Our work has been applied to several domains, including the training of scrub nurses to neurosurgery procedures 10 .

Figure 10. The #FIVE and #SEVEN models used in the S3PM project to provide an interactive environment and define collaborative scenarios and handle dynamic team structures in a surgical context.

7.4. Brain-Computer Interfaces

7.4.1. Novel Usages of BCI

Mind-Window: Real-Time Brain Activity Visualization Using Tablet-Based Augmented Reality and EEG for Single or Multiple Users, Anatole Lécuyer, Jonathan Mercier, Maud Marchal

Figure 11. Our novel "Mind-Window" approach enables one or multiple users to visualize the brain activity of a person in real-time by using tablets and augmented reality. It proposes to see through the tablet a virtual brain model "as if the skull is transparent". The display of the virtual brain is updated in real-time according to the real brain activity of the person which is measured thanks to an EEG headset.

We introduced a novel approach, called the “Mind-Window”, for real-time visualization of brain activity [33]. The Mind-Window enables one or multiple users to visualize the brain activity of another person as if his/her skull was transparent. Our approach relies on the use of multiple tablet PCs that the observers can move around the head of the observed person wearing an electroencephalography cap (EEG). A 3D virtual brain model is superimposed to the head of the observed person using augmented reality by tracking a 3D marker placed on top the head. The EEG cap records the electrical fields emitted by the brain, and they are processed in real-time to update the display of the virtual brain model. Several visualization techniques are proposed such as an interactive cutting plane which can be manipulated with touch-based inputs on the tablet. The Mind-Window approach could be used for medical applications, e.g. by providing a simple way for physicians to diagnose and observe brain activity of patients. Teachers could also use our system to teach brain anatomy/activity and EEG features, e.g., electrodes localization, electrical patterns, etc. Finally, video conferences or video games could be “brain-augmented”, making use of the Mind-Window for entertainment purposes.

B-C-Invisibility Power: Optical Camouflage Based on Mental Activity in Augmented Reality, Anatole Lécuyer, Jonathan Mercier, Maud Marchal

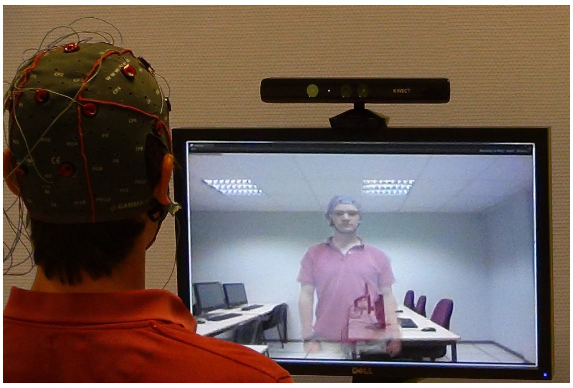


Figure 12. The "B-C-Invisibility power” enables users to become virtually invisible by performing mental tasks. Brain signals are extracted using EEG electrodes and analyzed within the BCI.

In the context of the ANR project HOMO-TEXTILUS which focuses on the design of novel "smart clothes", we introduced a kind of "invisibility cloak": an interactive approach for using Brain-Computer Interfaces for controlling optical camouflage called "B-C-Invisibility power”. We proposed to combine augmented reality and BCI technologies to design a system which somehow provides the "power of becoming invisible” [32]. Our optical camouflage is obtained on a PC monitor combined with an optical tracking system. A cut out image of the user is computed from a live video stream and superimposed to the prerecorded background image using a transparency effect. The transparency level is controlled by the output of a BCI, making the user able to control her invisibility directly with mental activity. The mental task required to increase/decrease the invisibility is related to a concentration/relaxation state. Results from a preliminary study based on a simple video-game inspired by the Harry Potter universe could notably show that, compared to a standard control made with a keyboard, controlling the optical camouflage directly with the BCI could enhance the user experience and the feeling of “having a super-power”.

7.4.2. BCI Methodology and Techniques

A methodological framework for applications combining BCI and videogames, Anatole Lécuyer

We have proposed a user-centered methodological framework [41] to guide design and evaluation of applications based on Brain-Computer Interface (BCI). Our framework is based on the contributions of ergonomics to ensure that these applications are well suited for end-users. It provides methods, criteria and metrics to perform the phases of the human-centered design process aiming to understand the context of use, specify the user needs and evaluate the solutions in order to define design choices. Several ergonomic methods (e.g., interviews, longitudinal studies, user based testing), objective metrics (e.g., task success, number of errors) and subjective metrics (e.g., mark assigned to an item) are suggested to define and measure the usefulness, usability, acceptability, hedonic qualities, appealingness, emotions related to user experience, immersion and presence to be respected. The benefits and contributions of our user centred framework for the ergonomic design of videogames based on BCI were also discussed.

This work was done in collaboration with Fabien Lotte (Inria team POTIOC).

Feasibility and specificity of simultaneous EEG and fMRI, Marsel Mano, Lorraine Perronnet, Jussi Lindgren, Anatole Lécuyer

In the field of fMRI, Arterial Spin Labeling (ASL) imaging relies on control and label radio-frequency pulses. This generates alternate gradient patterns as well as higher specific absorption rate (SAR). To date, only a few studies have addressed the issue of connecting EEG signal to ASL perfusion. Furthermore, previous studies have shown reduced blood-oxygen-level dependent (BOLD) signal-to-noise ratio (SNR) in the presence of EEG. ASL being a low SNR technique, the aim of this study was to assess ASL-EEG at 3T in terms of safety as well as EEG and magnetic resonance signal quality. Our experimental results show that ASL-EEG can be safely performed [20] [38]. Standard ASL acquisitions generated more than 2.5-fold SAR increase compared to a standard BOLD echo planar imaging sequence. This corresponded to up to 4°C temperature increase on the bundle, yet not exceeding 36°C. Gradient artifact correction of the EEG signal by average artifact subtraction was generally good for BOLD-EEG and ASL-EEG. However, residual gradient artifacts affecting 1% of the pulsed ASL-EEG data have to be considered. Further research is needed to understand the artifact variability and to develop an appropriate correction strategy. No residual artifacts were observed for alternating control and label pulses ASL-EEG. Neither a change of the number of reference volumes for artifact subtraction nor an independent component analysis could help tackle this gradient artifact correction issue. Regarding magnetic resonance imaging, a 20% loss in SNR was observed when compared to acquisitions performed without EEG. Taken together our results suggest that EEG and ASL MRI can be simultanously combined for the purpose of real-time experiments which could for instance be envisionned in our HEMISFER project.

This work was done in collaboration with VISAGES team.

ILDA Team

7. New Results

7.1. An Evaluation of Interactive Map Comparison Techniques

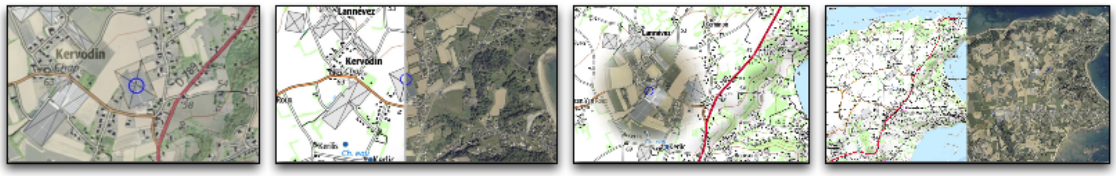


Figure 4. Empirical evaluation of multiplexing strategies using juxtaposition or overlaying for spatially-registered map comparison tasks [4]. Research conducted in the context of ANR project MapMuxing with IGN (Institut National de l'Information Géographique et Forestière).

Geovisualization applications typically organize data into layers. These layers hold different types of geographical features, describe different characteristics of the same features, or represent those features at different points in time. Layers can be composited in various ways, most often employing a juxtaposition or superimposition strategy, to produce maps that users can explore interactively. From an HCI perspective, one of the main challenges is to design interactive compositions that optimize the legibility of the resulting map and that ease layer comparison. We characterized five representative techniques, and empirically evaluated them using a set of real-world maps in which we purposefully introduced six types of differences amenable to inter-layer visual comparison. We discussed the merits of these techniques in terms of visual interference, user attention and scanning strategy. Those results can help inform the design of map-based visualizations for supporting geo-analysis tasks in many application areas.

This work was published at ACM CHI 2015 [4], and received an honorable mention (top 5% of all submissions).

7.2. Reciprocal Drag and Drop

Drag-and-drop has become ubiquitous, both on desktop computers and touch-sensitive surfaces. It is used to move and edit the geometry of elements in graphics editors, to adjust parameters using controllers such as sliders, or to manage views (e.g., moving and resizing windows, panning maps). Reverting changes made via a drag-and-drop usually entails performing the reciprocal drag-and-drop action. This can be costly as users have to remember the previous position of the object and put it back precisely. We introduced the DnD^{-1} model that handles all past locations of graphical objects. We redesigned the Dwell-and-Spring widget to interact with this history. Applications can implement DnD^{-1} to enable users to perform reciprocal drag-and-drop to any past location for both individual objects and groups of objects. We performed two user studies, whose results show that users understand DnD^{-1}, and that Dwell-and-Spring enables them to interact with this model effectively.

This work was published in ACM ToCHI [1].

7.3. SketchSliders: Sketching Widgets for Visual Exploration on Wall Displays

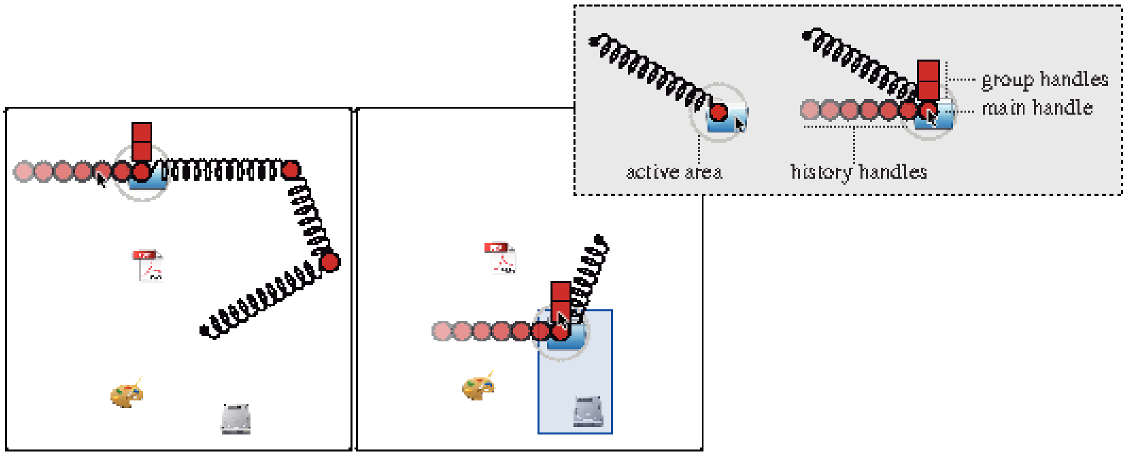


Figure 5. Navigating a graphical object's direct manipulation history as captured by the DnD^{-1} model, using the Dwell-and-Spring widget.

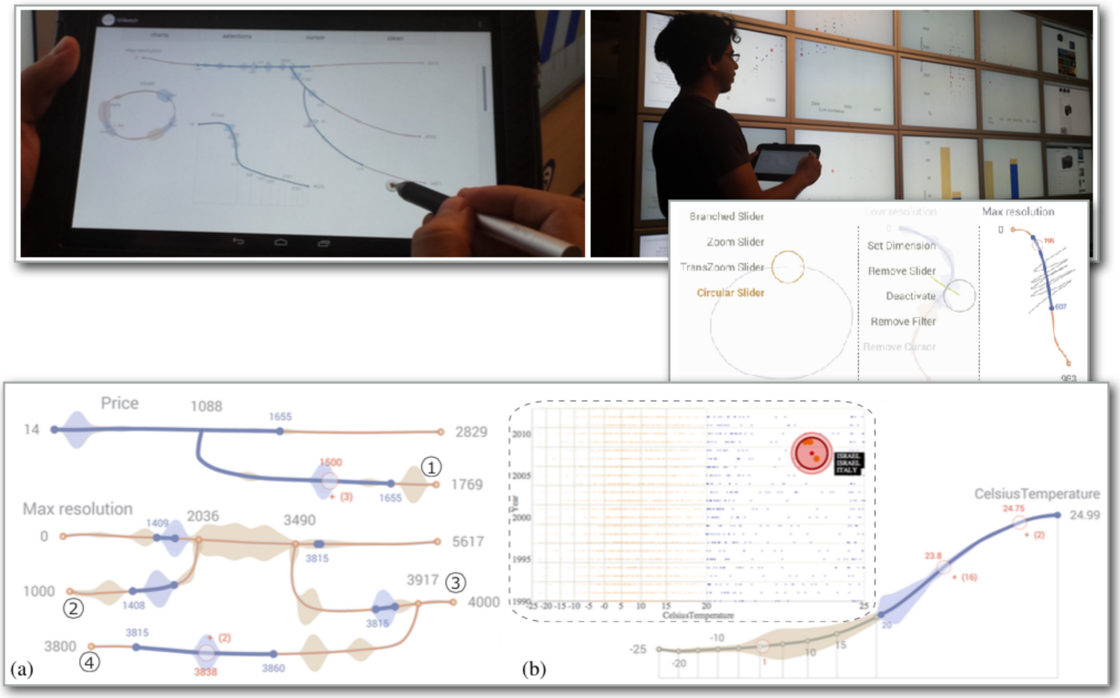


Figure 6. (top) The user sketching their sliders on the fly (left), to interact with their data on the wall display (right). Menus and simple gestures (middle) are enough to create complex sliders (bottom) that can help explore data at different granularities.

Given our interest in how to effectively interact with wall displays, we have started investigating ways to empower end users, by allowing them to easily create themselves their interfaces. We introduced a sketching interface that runs on mobile devices, and allows users to explore multi-dimensional datasets on wall displays by sketching on the fly the interactive controllers they require. We demonstrated this concept with *SketchSliders*, range sliders that users can freely sketch on the mobile surface to customize their exploration. A small combination of sketches and gestures allows the creation of complex interactive sliders, such as circular sliders for periodic data, slider branches for detailed interaction, and fisheye transformation sliders. We augmented sliders with a suite of tools, such as markers, slider cursors, and approximate views of data distributions. These designs were inspired by a design study with three visualization experts, and validated through a user study with six experts using our system.

This work was published at ACM CHI 2015 [9], and received an honorable mention (top 5% of all submissions).

7.4. Ultra-high-resolution Wall-sized Displays

We have worked on the following other projects, also related to the interactive visualization of large datasets on ultra-high-resolution wall displays:

- Mid-air Pointing on Ultra-Walls [5]. The size and resolution of ultra-high resolution wall-sized displays (“ultra-walls”) make traditional pointing techniques inadequate for precision pointing. We studied mid-air pointing techniques that can be combined with other, domain-specific interactions. We explored the limits of existing single-mode remote pointing techniques and demonstrated theoretically that they do not support high-precision pointing on ultra-walls. We then explored solutions to improve mid-air pointing efficiency: a tunable acceleration function and a framework for dual-precision (DP) techniques, both with precise tuning guidelines.
- WallTweet: A Knowledge Ecosystem for Supporting Situation Awareness [20]. Tweets are an important source of information during large-scale events, like tornados or terrorist attacks. Yet, tweets are hard to visualize and put in a geographical context: large quantities of tweets get sent in a short period, that vary greatly in content and relevance with respect to the crisis at hand. WallTweet is a tweet visualization designed for wall displays and aimed at improving the situation awareness of users monitoring a crisis event utilizing tweets.
- The monitoring of road trafic data on wall-sized displays [15]. Road traffic is a complex system that can be very unstable. A little perturbation can lead to a traffic-crippling congestion. To avoid such situations, researchers attempt to model traffic in order to prevent congestions and optimize traffic flow. Traffic is also continually monitored by operators in traffic control rooms. We designed an interactive system to monitor traffic on a wall display, that is coupled to traffic modeling algorithms. The system enables users to interactively adjust traffic parameter settings and visualize the impact of these adjustments at both a local and global scale.

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. Developable Surfaces

Participants: Antoine Begault, Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer, Camille Schreck.

Figure 5. Example of developable model generation from sketches in [10] and for virtual paper interaction [14].

Developable surfaces are surfaces which can be unflattened on a plane without being stretched nor squeezed. In other words, they can be made from 2D pattern without change of lengths. They are usually hard to model efficiently as the length condition is non linear. We developed this year two different applications for developable surfaces, once applied for leather product designer, and the other one to virtual paper deformation.

We developed a method to generate 3D models for garments and leather products from designer sketches. Given two or three orthogonal sketched views depicting the silhouette, the seams, and the folds, we automatically compute a 3D developable surface and the corresponding 2D patterns which fits the silhouette and exhibits the designed folds. Our method can handle complex cases where the 2D silhouette actually correspond to a non planar and discontinuous curve on the 3D surface. We also proposed a new efficient approach to improves the developability of the resulting surface while preserving the pre-designed folds. This work has been published in ACM Transactions on Graphics [10], and we presented it in SIGGRAPH Asia in November.

Within the PhD work of Camille Shreck, we developed the first interactive 3D virtual model of crumpled paper. Deforming virtual paper is especially challenging to model efficiently has crumpling can be seen as singularities on the surface, leading therefore to non smooth surfaces which do not fit well to standard physically based deformation model. We proposed in this work a new geometrical representation of surface especially adapted to model non smooth developable surfaces as a set of planes, cylinders, and generalized cones meeting at the discontinuities of the surface. Our model can dynamically adapts to the surface deformation and to new crumples, while been associated to an optimal mesh triangulation containing very few triangles. Our interactive deformation model interleaves a standard Finite Element Model on the coarse triangular mesh to guide the general deformation, with a geometrical steps adapting our surface structure to optimally sample the degrees of freedom of the crumpled paper. This work as been accepted for publication in ACM Transaction on Graphics [14], has been presented at the conference WomEncourage [29], and as a communication in AFIG [33].

6.1.2. Procedural models for shape assemblies

Participants: Marie-Paule Cani, Damien Rohmer, Ulysse Vimont.

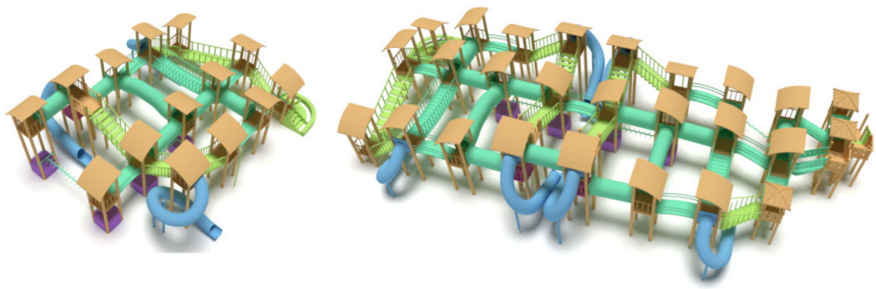


Figure 6. Example of shape assemblies before and after subpart-remplacement in [11].

A popular mode of shape synthesis involves mixing and matching parts from different objects to form a coherent whole. In collaboration with the University College London, Universiteit Utrecht, and KAUST, we proposed a method to automatically detect replaceable subparts within a complex assembly. In this work, we model the geometrical assembly as a graph where each node represent a single component, and the edges represents inter-part connectivity. Our method analyses this graph to detect similar inter-part connectivity enabling to exchange or mix sub-structures to synthesise new geometrical models. This work has been published in Eurographics [11].

6.1.3. Toward Functional CAD assemblies

Participants: Pablo Covès, Harold Vilmart, Robin Roussel, Damien Rohmer, Marie-Paule Cani, Jean-Claude Léon.

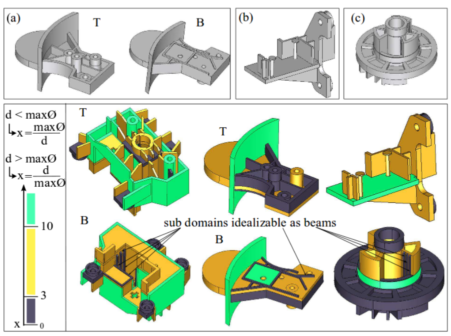


Figure 7. Example of shape idealization in [5].

We chose to focus on man-made objects to tackle the topic of shape assemblies. This is two-folds since CAD models of virtual industrial prototypes provide an excellent, real-size test-bed for our methods. Moreover, this is perfectly fitting the demand from industrial partners such as EDF and Airbus Group. On a complementary basis, we have initiated a partnership with UCL (University College London) to address function-preserving assembly deformation.

Assemblies representing products are most often reduced to a collection of independent CAD models representing each component. To our knowledge, there has been no approach proposed to generate CAD assembly models from 3D scans. An approach is initiated with a partnership with LIRIS (R. Chaine and J. Digne) and EDF in the framework of a Rhône-Alpes region project (Potasse) starting with the PhD of P. Coves.

Following the work of [43],[5], [38], partnership with Inria GRAPHIK team (F. Ulliana) has been set up and a deductive logic framework has been coupled to the SALOME plateform with the insertion of an ontology describing a subset of a product structure. This partnership is developed with the internship of H. Vilmart to evolve toward an intrinsic, knowledge-based representation of a product structure that takes into account the isometries of components using our prior work about symmetry analyses [42]. The description of components through this product structure aims at supporting the generation of CAD assembly models from 3D scans to be able to derive functionally meaningful constraints of relative positions of components extracted from scans.

In the scope of the ERC Expressive, a partnership has been set up with N. Mitra (UCL) with the starting PhD of R. Roussel addressing function-preserving assembly deformation.

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 collision 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. *Real-time physically-based models*

Participants: Marie-Paule Cani, Francois Faure, Pierre-Luc Manteaux, Richard Malgat, Matthieu Nesme.

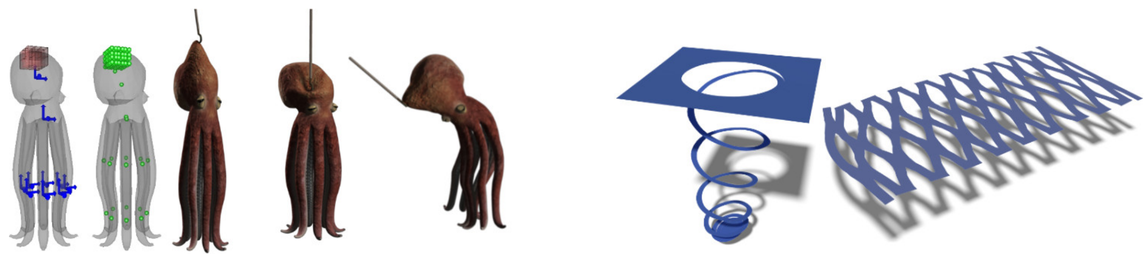


Figure 8. Left: Mixing a coarse frame-based simulation to a local FEM patch from [24]. Right: Frame based simulation for surface cutting in [25].

We keep on improving fundamental tools in physical simulation, such as new insight on constrained dynamics [15] at Siggraph. This allows more stable simulations of thin inextensible objects. A new extension of our volumetric contact approach (Siggraph 2010 and 2012) has been proposed [17] to apply rotational reaction to contact according to the shape of the contact area.
We have proposed an original approach to multi-resolution simulation, in which arbitrary deformation fields at different scales can be combined in a physically sound way[24]. This contrasts with the refinement of a given technique, such as hierarchical splines or adaptive meshes.
Following the success of frame-based elastic models (Siggraph 2011), a real-time animation framework provided in SOFA and currently used in many of our applications with external partners, we proposed an extension to the cutting of surface objects [25], in collaboration with Berkeley, where Pierre-Luc Manteaux spent 4 months at the end of 2014.

6.2.2. Simulating paper material with sound

Participants: Marie-Paule Cani, Pierre-Luc Manteaux, Damien Rohmer, Camille Schreck.

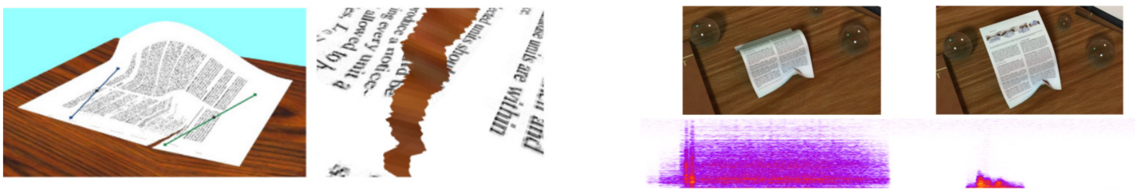


Figure 9. Left: Example of our paper tearing model in [23]. Right: Our sound synthesis for paper crumpling in [34].

Extending our results on animating paper crumpling, we proposed to synthesise the sound associated to paper material. We proposed a real time model dedicated to paper tearing. In this work, we model the specific case when two hands are tearing a flat sheet of paper on a table, in this case we synthesise procedurally the geometrical deformation of the sheet using conical surface, the tearing using a procedural noise map, and the tearing sound as a modified white noise depending on the speed of action. This work has been published in Motion in Games conference [23].

We are also developing a sound synthesis method for paper crumpling. The geometrical surface deformation is analysed to drive a procedurally synthesized friction sound and a data driven crumpling sound. We are currently developing this work and did a first communication to AFIG conference [34].

6.2.3. Animating anatomy

Participants: Armelle Bauer, Ali Hamadi Dicko, Francois Faure, Olivier Palombi, Damien Rohmer.

A real-time spine simulation model leveraging the multi-model capabilities of SOFA was presented in an international conference on biomechanics [7]. We also used a biomechanical model to regularize real-time motion capture and display, and performed live demos at the Emerging Technologies show of Siggraph Asia, Kobe, Japan [41].

We are developing an ontology-based virtual human embryo development model. In one side, a dedicated ontology stores the anatomical knowledge about organs' geometry, relations, and development rules. On the other side, we synthesize an animated visual 3D model using the informations of the ontology. This work can be seen as a first step toward interactive development anatomy teaching, or simulation, based on an ontology storing existing medical knowledge. This work has been published in the Journal of Biomedical Semantics [12].

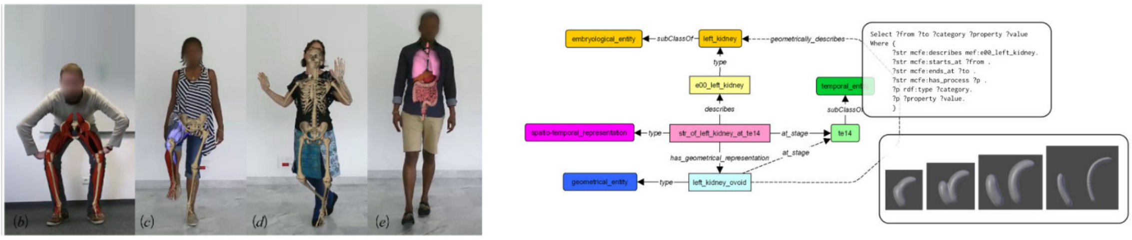


Figure 10. Left: Example of the Living Book of Anatomy in [41]. Right: Example of kidney development ontology and its 3D modeling in [12].

6.3. Knowledge-based Models for Narrative Design

- **Scientist in charge**: Rémi Ronfard.
- **Other permanent researchers**: Marie-Paule Cani, 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 direction tools

Participants: Adela Barbulescu, Rémi Ronfard.

Figure 11. Live demo at the EXPERIMENTA exhibition in Grenoble.

During the third year of Adela Barbulescu's PhD thesis, we proposed a solution for converting a neutral speech animation of a virtual actor (talking head) to an expressive animation. Using a database of expressive audiovisual speech recordings, we learned generative models of audiovisual prosody for 16 dramatic attitudes (seductive, hesitant, jealous, scandalized, etc.) and proposed methods for transferring them to novel examples. Our results demonstrate that the parameters which describe an expressive performance present person-specific signatures and can be generated using spatio-temporal trajectories; parameters such as voice spectrum can be obtained at frame-level, while voice pitch, eyebrow raising or head movement depend both on the frame and the temporal position at phrase-level. This work was presented at the first joint conference on facial animation and audio-visual speech processing [16] and in a live demo at the EXPERIMENTA exhibition in Grenoble, and was seen by 1200 visitors.

6.3.2. Virtual cinematogaphy

Participants: Quentin Galvane, Rémi Ronfard.

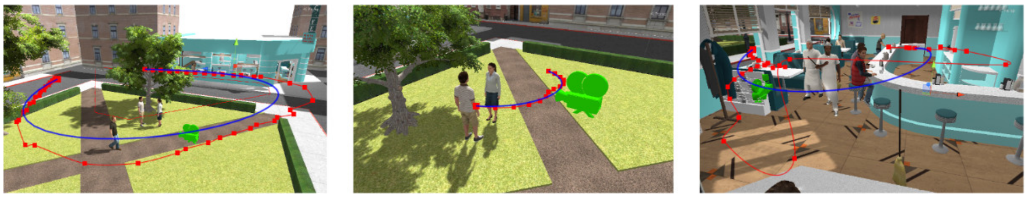


Figure 12. Automatic camera trajectory in [18].

During the third year of Quentin Galvane's Phd thesis, we proposed a solution for planning complex camera trajectories in crowded animation scenes [2] Galvane [18]. This work was done in a collaboration with Marc Christie in Rennes.

We also published new results from Vineet Gandhi's PhD thesis (defended in 2014) on the generation of cinematographic rushes from single-view recordings of theatre performances [32]. In that paper, we demonstrate how to use our algorithms to generate a large range of dynamic shot compositions from a single static view, a process which we call "vertical editing". Our patent application on this topic was reviewed positively and is being extended.

Those techniques were used to automatically generated cinematographically pleasant rushes from a monitor camera during rehearsals at Theatre des Celestins, as part of ANR project "Spectacle-en-Lignes". Results of the projects are described in two papers [28], Steiner [30] and we presented them to a professional audience during the Avignon theatre festival. This work was done in a collaboration with the Institut de Recherche et d'Innovation (IRI) at Centre Pompidou and the SYLEX team at LIRIS.

6.3.3. Film editing & narrative design

Participants: Quentin Galvane, Rémi Ronfard.

We proposed a new computational model for film editing at the AAAI artificial intelligence conference, which is based on semi-Makov chains [20]. Our model significantly extends previous work by explicitly taking into account the crucial aspect of timing (pacing) in film editing. Our proposal is illustrated with a reconstruction of a famous scene of the movie "Back to the future" in 3D animation, and a comparison of our automatic film editing algorithms with the director's version. Results are further discussed in two companion papers [31], Galvane [19]. This work was done in a collaboration with Marc Christie in Rennes. Future work is being planned to extend this important work to the case of live-action video (as described in the previous section) and to generalize for the case non-linear film editing including temporal ellipses and flashbacks.

Figure 13. Example of automatic movie sequence editing in [20].

6.4. Creating and Interacting with Virtual Prototypes

- **Scientist in charge**: Jean-Claude Léon.
- **Other permanent researchers**: Marie-Paule Cani, 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 shape hierarchies

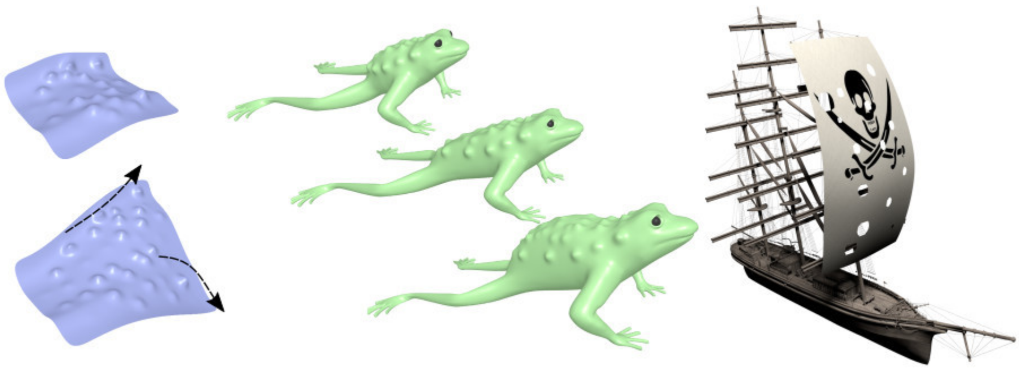


Figure 14. Our method in [13] allows surface stretching while the geometrical details are continuously duplicated.

Sculpting paradigm has been successfully applied to deform simple smooth surfaces. More complex objects representing virtual characters or real-life objects are however modeled as hierarchy of shapes with elements, sub-elements and details. Applying sculpting deformation to such objects is challenging as every parts of the hierarchy should stay coherent through the deformation.

When an object can be represented as a smooth underlying surface and a set of singular details, we proposed a real-time deformation approach enabling to freely stretch or squeeze the 3D object while continuously maintaining the details' appearance. Instead of stretching or squeezing the details the same way than the smooth underlying surface, we duplicate or merge them smoothly while ensuring that the distribution of details has the same caracteristic than the original one. We published this work in Shape Modeling International [13].

In the case of more general object hierarchies, we are developing a new methodology to apply generic deformation into complex assemblies while preserving their properties in extending the shape grammar approach into our new *deformation grammar*. We presented our preliminary results as a communication in the GTMG conference [35].

6.4.2. Sketching and sculpting Virtual Worlds

Participants: Marie-Paule Cani, Guillaume Cordonnier, Ulysse Vimont.

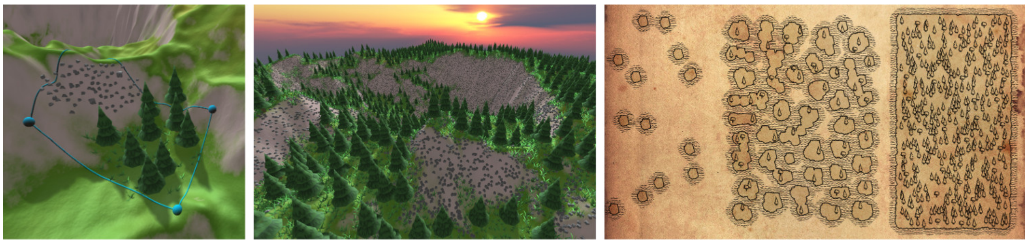


Figure 15. Example of maps generated using our World Brush method [8].

Modeling virtual worlds is particular challenging: the fractal-like distribution of details in terrain shapes makes them easy to identify, but very difficult to design using standard modeling software, even for expert users. Moreover, virtual worlds involve distributions of different categories of contents over terrains, such as vegetation, houses, roads or rivers. Efficiently modeling these sets of elements, which are statistically correlated, is indeed a challenge.

This year, our contributions to tackle these issues were two-folds:

Firstly, we investigated the use of a plate tectonics metaphor to generate plausible terrains from a simple vector map representing the location of the main rivers and mountain picks. The method uses a Voronoi tesselation of pick locations to automatically generate tectonic plates which themselves drive terrain folds. Hydraulic erosion is then used to further sculpt the terrain and add details, while the specified rivers are considered to maintain consistency with the input map. This work was published in [27]. A more accurate modeling of large scale fluvial erosion and plates tectonics phenomena was investigated in Guillaume Cordonnier's master thesis and is the object of his PhD, which started in October 2015.

Secondly, we proposed a paint-based interface to tackle the problem of easily populating a terrain with distributions of objects (trees, rocks, grass, houses, etc) or of graph-like structures such as rivers and roads. The key point of our solution is to learn statistics about distributions of elements and their correlation with other distributions, with graph structures, or with terrain slope, and store the resulting histograms as "colors" in a palette interface. After creating a few local distribution manually, the user selects them with a pipette tool, and is able to reuse them with a brush. We also provided a gradient tool to interpolate between two such "colors" an move tool enabling, for instance to move groups of trees and rocks over a terrain while maintaining the adequate correlation with local slope, and a deformation interface based on seam carving enabling to seamlessly stretch or compress a region of virtual world. This work, a collaboration between Arnaud Emilien when defended his phD in December 2014, Ulysse Vimont, Marie-Paule Cani, and Bedrich Benes from Purdue University, was published at Siggraph 2015 [8].

6.4.3. Sketching and sculpting Motion

Participants: Marie-Paule Cani, Martin Guay, Kevin Jordao, Rémi Ronfard.

Sketching and sculpting methods were restricted so far to the design of static shapes. One of our research goals has been to extend these interaction metaphors to motion design. This year, this included three specific contributions.

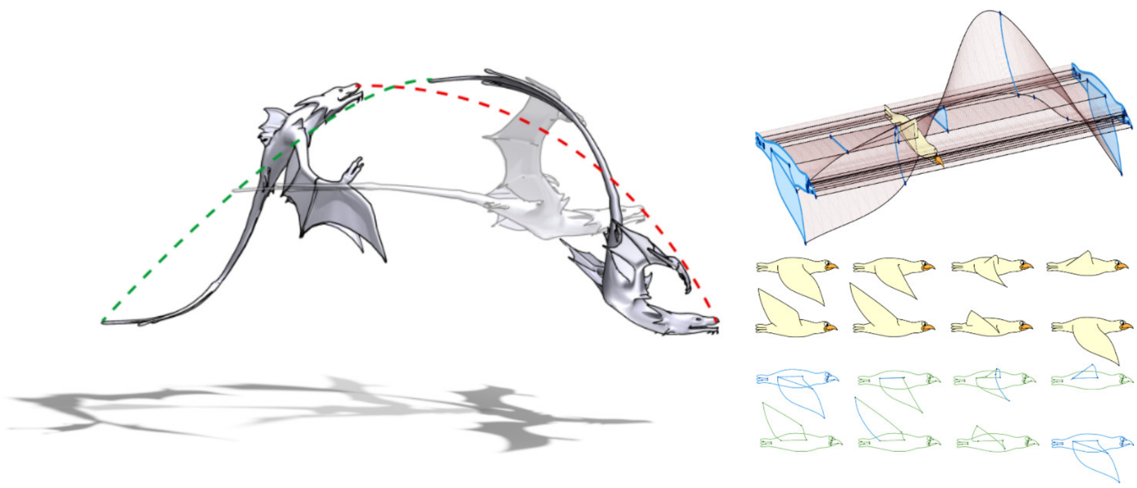


Figure 16. Left: Result of our Sculpting Motion method [9]. Right: Result of our Vector Graphics method enabling to design vectorial animation [6].

Firstly, to handle sketch-based representation of motion in the 2D case, we extended the static vector graphics complex data structure, which we had introduced at Siggraph last year, to vector graphics animations with time-varying topology [6]. This second paper was presented at Siggraph again this year. The proposed data structure is augmented with a rich set of editing operations, which can be used to quickly interpolate 2D drawings with different topologies. This work was done within a collaboration with Boris Dalstein and Michiel van de Panne from UBC, Canada.

Secondly, following a first method enabling to sculpt crowd animations (Jordao, Eurographics 2014), we developed a painting interface enabling to specify both density and main directions of motion in an animated crowd. The resulting system is still based on crowd-patches, i.e. the crowd motion is an assembly of local trajectories defined in interconnected patches. Our new painting system, called Crowd-Art, uses discrete changes in loop trajectories to evolve the number of in/out constraints in each patch until the requested density and directions are best matched. See [22]. This concluded Kevin Jordao's PhD thesis, co-advided by Julien Pettre from the MimeTIC team and in collaboration with Marc Christie, defended in December 2015.

Lastly, we developed the first expressive interface to interactively sketch and progressively sculpt and refine character motion. Our solution is based on a space-time sketching metaphor: The user sketches a single space-time stroke, which is used to initialize a series of dynamic lines of action, serving as intermediates to animate the character's model. Motion and shape deformation can be immediately replayed from this single stroke, since it sets at the same time shape, trajectory and speed (defined from the drawing speed). Thanks to visual feedback, the user can easily refine the resulting motion by editing specific lines of actions at fixed times, or by composing several motions together. This work, published at Siggraph, is one of the first methods enabling arbitrary motion to be defined from scratch by a beginner [9]. Together to another work enabling to add dynamics to character motion [21], this concluded Martin Guay's PhD thesis, defended in June 2015.

MANAO Project-Team

7. New Results

7.1. Analysis and Simulation

7.1.1. Parametrization of BRDFs

Opaque materials are represented in computer graphics by Bi-directional Reflectance Distribution Functions (BRDF), which are 4D functions of light and view direction. Dealing with such a high dimensionality is problematic for the modeling and rendering of material appearance. The choice of a BRDF parametrization greatly simplifies this task by identifying the axis where most variations occur in common opaque materials. The 4D parametrization of Rusinkiewicz [86] is classically used in graphics, in particular because of its direct connection to micro-facet theory. Alternative parametrization by Neumann et al. [71] and Stark et al. [91] have been proposed, but are restricted to 2D parametrizations, and hence a restricted class of materials.

We have extended the work of Neumann et al. [71] and Stark et al. [91] to a pair of 4D BRDF parameterizations with explicit changes of variables. Revealing some of their mathematical properties and relationships to Rusinkiewicz' parametrization allows us to better understand their benefits and drawbacks for representing measured BRDFs. Our preliminary study suggests that the alternative parametrization inspired by Stark et al. [91] is superior, and should thus be considered in future work involving BRDFs.

7.1.2. New BRDF Model and Diffraction Effects identification

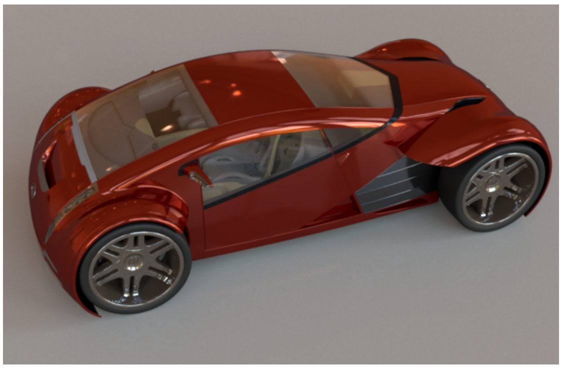
Finding the appropriate BRDF model, with meaningful physical parameters, that can represent accurately measured data remains a challenging task. In [20], we show that two different physical phenomena are present in measured reflectance: reflection and diffraction. Taking both into account, we present a reflectance model [24] that is compact and a very good approximation (cf. Figure 8) of measured reflectance. Designers can act on model parameters, related to surface properties, to create new materials.

7.1.3. Statistical analysis of BRDFs

On the one hand, a BRDF is a complex 4D function, which should ensure reciprocity and energy conservation laws. On the other hand, when computing radiance reaching the eye from a surface point, the view direction is held fixed. In this respect, we are only interested in a 2D BRDF slice that acts as a filter on the local environment lighting. In [21], our goal is to understand the statistical properties of such a filter as a function of viewing elevation. To this end, we have conducted a study of measured BRDFs where we have computed statistical moments for each viewing angle. We show that some moments are correlated together across dimensions and orders, while some others are close to zero and may safely be discarded. Our study opens the way to novel applications such as moment-based manipulation of measured BRDFs, material estimation and image-based material editing. It also puts empirical and physically-based material models in a new perspective, by revealing their effect as view-dependent filters.

7.1.4. Importance Sampling of Real Light Sources

Realistic images can be rendered by simulating light transport with Monte Carlo methods. The possibility to use realistic light sources for synthesizing images greatly contributes to their physical realism. Among existing models, the ones based on light fields are attractive due to their ability to capture faithfully the far-field and near-field effects as well as their possibility of being acquired directly. Since acquired light sources have arbitrary frequencies and possibly high dimensions (4D), using such light sources for realistic rendering leads to performance problems. We have investigated [12] how to balance the accuracy of the representation and the efficiency of the simulation (cf. Figure 9). The work relies on generating high quality samples from the input light sources for unbiased Monte Carlo estimation [67]. This is a foundation work that has leaded to new sampling techniques for physically-based rendering with light field light sources. The results show that physically accurate rendering with realistic light sources can be achieved in real time.



(a) Car model (Path tracing with adaptive sampling, 128 to 16384 samples per pixel), with smoother `red-metallic-paint` body, `chrome` wheels, `white-marble` floor. Inside the car: `pickled-oak-260`, `specular-white-phenolic`

(b) Kitchen model (Energy Redistribution Path Tracing, 5h), with `colonial-maple-223` cupboards, `chrome` tap, sink and handles, `nickel` kitchen wall, `alumina-oxide` oven door, `white-marble` and `black-obsidian` tiles, `brass` bowls, `aluminium` glasses.

Figure 8. Example scenes using our new BRDF model [24] to represent different measured materials.

Figure 9. Our new light importance sampling technique estimates direct lighting interactively (7-9 fps) with only 200 samples per pixel that are distributed among the different images of the light field luminaire. The car headlights are represented by the same light field composed of 11 $\times$ 9 images (256 $\times$ 256 pixels).

7.1.5. Exact Relations between Wave and Ray Aberrations

The aberrations of an optical system can be described in terms of the wave aberrations, defined as the departure from the ideal spherical wavefront; or the ray aberrations, which are in turn the deviations from the paraxial ray intersections measured in the image plane. The classical connection between the two descriptions is an approximation, the error of which has, so far, not been quantified analytically.

We derive [13] exact analytical equations for computing the wavefront surface, the aberrated ray directions, and the transverse ray aberrations in terms of the wave aberrations (a.k.a., Optical Path Difference) and the reference sphere. We introduce precise conditions for a function to be an OPD function, show that every such function has an associated wavefront, and study the error arising from the classical approximation. We establish strict conditions for the error to be small. We illustrate our results with numerical simulations. Our results show that large numerical apertures and OPD functions with strong gradients yield larger approximation errors.

7.2. From Acquisition to Display

7.2.1. Lytro Microscope

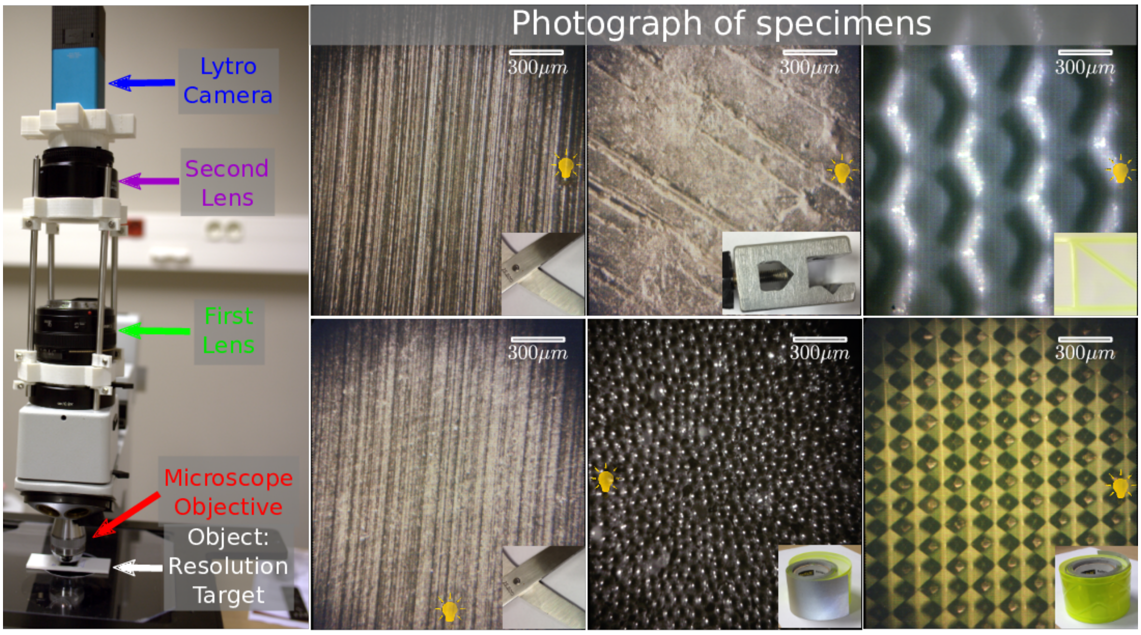
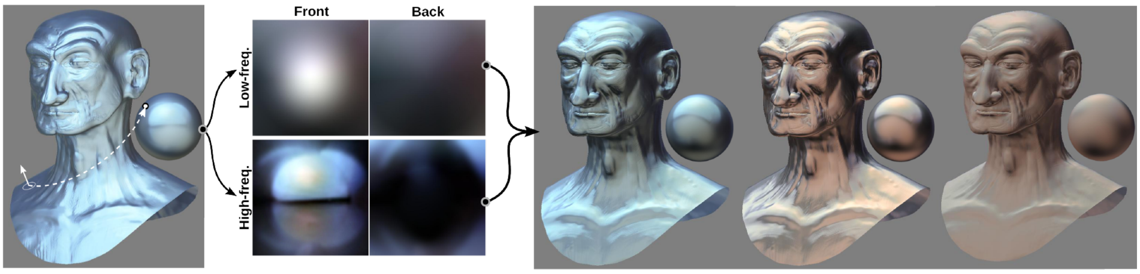


Figure 10. Light field microscopy with a consumer light field camera. Light fields can be beneficial for microscopic applications because they provide 3D information on a sample. Access to the technology, has, however, been limited by the need for custom-building the device. Our work enables an easy entry-level experimentation with the technology. (left) Light field microscope with a consumer light field camera. (right) Example specimens photographed with our system.

We explore [22] the use of inexpensive consumer light-field camera technology for the purpose of light-field microscopy. Our experiments are based on the Lytro (first generation) camera. Unfortunately, the optical systems of the Lytro and those of microscopes are not compatible, leading to a loss of light-field information due to angular and spatial vignetting when directly recording microscopic pictures. We therefore consider an adaptation of the Lytro optical system. We demonstrate that using the Lytro directly as an ocular replacement, leads to unacceptable spatial vignetting. However, we also found a setting that allows the use of the Lytro camera in a virtual imaging mode which prevents the information loss to a large extent. We analyze the new virtual imaging mode and use it in two different setups for implementing light-field microscopy using a Lytro camera. As a practical result, we show that the camera can be used for low magnification work, as e.g. common in quality control, surface characterization, etc. (cf. Figure 10) We achieve a maximum spatial resolution of about 6.25 micrometers, albeit at a limited SNR for the side views.

7.3. Editing and Modeling

7.3.1. MatCap Decomposition for Dynamic Appearance Manipulation



(a) Input MatCap (b) Our decomposition (c) Rotated lighting (d) Color change (e) Rougher look

Figure 11. Our method decomposes a MatCap into a representation that permits dynamic appearance manipulation via image filters and transforms. (a) An input MatCap applied to a sculpted head model (with a lookup based on screen-space normals). (b) The low- & high-frequency (akin to diffuse & specular) components of our representation stored in dual paraboloid maps. (c) A rotation of our representation orients lighting toward the top-left direction. (d) Color changes applied to each component. (e) A rougher-looking material obtained by bluring, warping and decreasing the intensity of the high-frequency component.

In sculpting software, MatCaps are often used by artists as a simple and efficient way to design appearance. Similar to LitSpheres, they convey material appearance into a single image of a sphere, which can be easily transferred to an individual 3D object. Their main purpose is to capture plausible material appearance without having to specify lighting and material separately. However, this also restricts their usability, since material or lighting cannot later be modified independently. Manipulations as simple as rotating lighting with respect to the view are not possible. In [23], we show how to decompose a MatCap into a new representation that permits dynamic appearance manipulation. We consider that the material of the depicted sphere acts as a filter in the image, and we introduce an algorithm that estimates a few relevant filter parameters interactively. We show that these parameters are sufficient to convert the input MatCap into our new representation, which enables real-time appearance manipulations through simple image re-filtering operations. This includes lighting rotations, the painting of additional reflections, material variations, selective color changes and silhouette effects that mimic Fresnel or asperity scattering (cf. Figure 11).

7.3.2. Dynamic On-Mesh Procedural Generation

In collaboration with Technicolor, we developed a method to generate procedural models with global structures, such as growth plants, on existing surfaces at interactive time [18]. Our approach extends shape grammars to enable context-sensitive procedural generation on the GPU. To this end, we unified the representation of external contexts as texture maps, which can be spatially varying parameters controlling the grammar expansion through very fast texture fetches (e.g., a density map). External contexts also include the shape of the

Figure 12. Given a base mesh and a procedural grammar of ivy growth, our GPU-based marching rule generated the ivy geometry on-the-fly in parallel with interactive performance. In this example, the grammar expansion is guided through a user-friendly painting interface.

underlying surface itself that we represent as a texture atlas of geometry images. Extrusion along the surface is then performed by a marching rule working in texture space using indirection pointers. We also introduce a lightweight deformation mechanism of the generated geometry maintaining a C1 continuity between the terminal primitives while taking into account the shape and trajectory variations. Our method is entirely implemented on the GPU and it allows to dynamically generate highly detailed models on surfaces at interactive time (cf. Figure 12). Finally, by combining marching rules and generic contexts, users can easily guide the growing process by directly painting on the surface with a live feedback of the generated model. This provides friendly editing in production environments.

7.3.3. Boolean on general 3D meshes

Computing Boolean operations (Booleans) of 3D polyhedra/meshes is a basic and essential task in many domains, such as computational geometry, computer-aided design, and constructive solid geometry. Booleans are challenging to compute when dealing with meshes, because of topological changes, geometric degeneracies, etc. Most prior art techniques either suffer from robustness issues, deal with a restricted class of input/output meshes, or provide only approximate results.

We overcome these limitations and introduced an exact and robust approach performing on general surface meshes (closed and orientable) [11]. Our method is based on a few geometric and topological predicates that allow to handle all input/output cases considered as degenerate in existing solutions, such as voids, non-manifold, disconnected, and unbounded meshes, and to robustly deal with special input configurations. Our experimentation showed that our more general approach is also more robust and more efficient than Maya's implementation ($\times 3$), CGAL's robust Nef polyhedra ($\times 5$), and recent plane-based approaches.

During this work, we also developed a complete benchmark intended to validate Boolean algorithms under relevant and challenging scenarios, and we successfully ascertain both our algorithm and implementation with it.

7.3.4. Extending MLS surfaces

Moving least squares (MLS) surface approximation is a popular tool for the processing and reconstruction of non-structured and noisy point clouds. We introduce [14] a new variant improving the approximation quality when the underlying surface is assumed to be locally developable, which is often the case in point clouds coming from the acquisition of manufactured objects. Our approach follows Levin's classical MLS procedure: the point cloud is locally approximated by a bivariate quadratic polynomial height-field defined in a local tangent frame. The a priori developability knowledge is introduced by constraining the fitted poly-nomials to have a zero-Gaussian curvature leading to the actual fit of so-called parabolic cylinders. When the local developability assumption cannot be made unambiguously, our fitted parabolic cylinders seamlessly degenerate to linear approximations. We show that our novel MLS kernel reconstructs more locally-developable surfaces than previous MLS methods while being faithful to the data.

MAVERICK Project-Team

6. New Results

6.1. Single Scattering in participating media with refractive boundaries

Participant: Nicolas Holzschuch [contact].

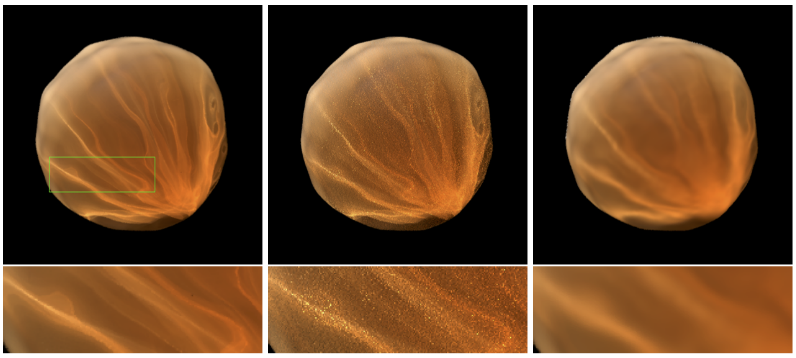


Figure 2. Single scattering: comparison between our algorithm and existing methods (equal computation time) on a translucent sphere illuminated by a point light source from behind.

Volume caustics are high-frequency effects appearing in participating media with low opacity, when refractive interfaces are focusing the light rays (see Figure 2). Refractions make them hard to compute, since screen locality does not correlate with spatial locality in the medium. We have developed a new method for accurate computation of single scattering effects in a participating media enclosed by refractive interfaces. Our algorithm is based on the observation that although radiance along each camera ray is irregular, contributions from individual triangles are smooth. Our method gives more accurate results than existing methods, faster. It uses minimal information and requires no precomputation or additional data structures. This paper was published in the *Computer Graphics Forum* journal [3] and presented at the *Eurographics Symposium on Rendering*.

6.2. Diffraction effects in reflectance properties

Participant: Nicolas Holzschuch [contact].

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, it has a metallic or plastic appearance. The reflectance model (BRDF) is essential for photorealistic pictures. Measured reflectance provide high realism, at the expense of memory cost. Parametric models are compact, but it is difficult to find the right parameters from measured reflectance.

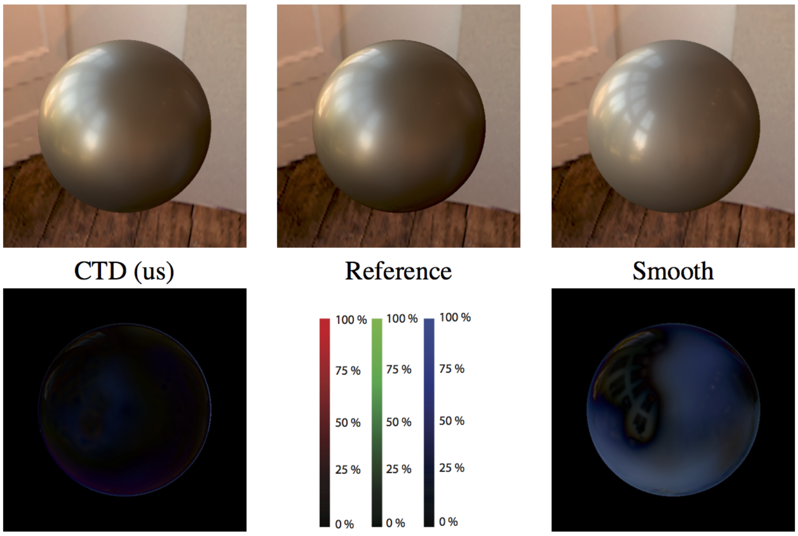


Figure 3. Comparison between our model combining reflection and diffraction (left), measured data (center) and state-of-the-art model (right).

Many parametric models are based on a physical representation of the surface micro-geometry and how it interacts with incoming light. The Cook-Torrance model assumes that light follows the principles of optical geometry: it is reflected by the surface micro-geometry but also potentially occluded. The diffraction model assumes that the micro-geometry diffracts the incoming light. This reflectance model has an intrinsic wavelength dependency. Previous experiments have shown that fitting measured materials to parametric models is hard. Heuristic models based on either Cook-Torrance or diffraction are complex, with many parameters. Our research has shown that both effects (optical geometry and diffraction) are present in most measured materials [6]. Based on this knowledge, we have proposed a new reflectance model, that accurately represents measured reflectance [10]. This model combines optical geometry for the specular peak and diffraction effects for wide-angle scattering.

6.3. Efficient and Accurate Spherical Kernel Integrals using Isotropic Decomposition

Participant: Cyril Soler [contact].

Spherical filtering is fundamental to many problems in image synthesis, such as computing the reflected light over a surface or anti-aliasing mirror reflections over a pixel. This operation is challenging since the profile of spherical filters (e.g., the view-evaluated BRDF or the geometry-warped pixel footprint, above) typically exhibits both spatial-and rotational-variation at each pixel, precluding precomputed solutions. We accelerate complex spherical filtering tasks using isotropic spherical decomposition (ISD), decomposing spherical filters into a linear combination of simpler isotropic kernels. Our general ISD is flexible to the choice of the isotropic kernels, and we demonstrate practical realizations of ISD on several problems in rendering: shading and prefiltering with spatially-varying BRDFs, anti-aliasing environment mapped mirror reflections, and filtering of noisy reflectance data. Compared to previous basis-space rendering solutions, our shading solution generates ground truth-quality results at interactive rates, avoiding costly reconstruction and large approximation errors. This paper was published in *ACM Transactions on Graphics* [4] and presented at Siggraph Asia 2015.

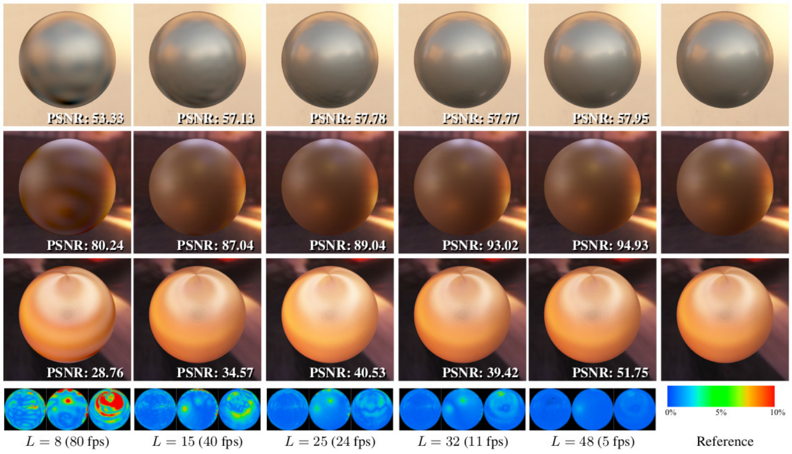


Figure 4. Convergence of spherical ISD shading for increasing L. Top to bottom: isotropic alum-bronze (with `pisa` *illumination), isotropic gold-paint and anisotropic yellow-satin (both using* `grace cathedral` *illumination). Reference images were ray-traced using 300K samples per pixel. False color error images in the bottom row visually illustrate the convergence of our RZH approximation to the reference rendering. Note that the dark region in the center of the spheres on the last row of renderings is indeed part of the underlying reflectance input data.*

6.4. Color transfer guided by summary statistics

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

Color transfer Colorization

Figure 5. Our framework allows for automatic local color transfer (left) and colorization (right) based on textural properties.

We have targeted 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 paper, we propose a unified framework that uses the textural content of the images to guide the color transfer and colorization. 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. This work was presented at the AFIG conference [7]. An extended version is available as a research report [9].

6.5. Programmable 2D Arrangements for Element Texture Design

Participants: Hugo Loi, Thomas Hurtut, Romain Vergne, Joëlle Thollot [contact].

We introduce a programmable method for designing stationary 2D arrangements for element textures, namely textures made of small geometric elements. These textures are ubiquitous in numerous applications of computer-aided illustration. Previous methods, whether they be example-based or layout-based, lack control and can produce a limited range of possible arrangements. Our approach targets technical artists who will design an arrangement by writing a script. These scripts are using three types of operators: partitioning operators for defining the broad-scale organization of the arrangement, mapping operators for controlling the local organization of elements, and merging operators for mixing different arrangements. These operators are designed so as to guarantee a stationary result meaning that the produced arrangements will always be repetitive. We show (see Figure 10) that this simple set of operators is sufficient to reach a much broader variety of arrangements than previous methods. Editing the script leads to predictable changes in the synthesized arrangement, which allows an easy iterative design of complex structures. Finally, our operator set is extensible and can be adapted to application-dependent needs. This work is available as a research report [11]

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

Participants: Léo Allemand-Giorgis, Georges-Pierre Bonneau [contact].

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 paper 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. We present the benefit and limitations of using polynomials for reconstruction surfaces from topological data. This research was published in a book chapter [8].

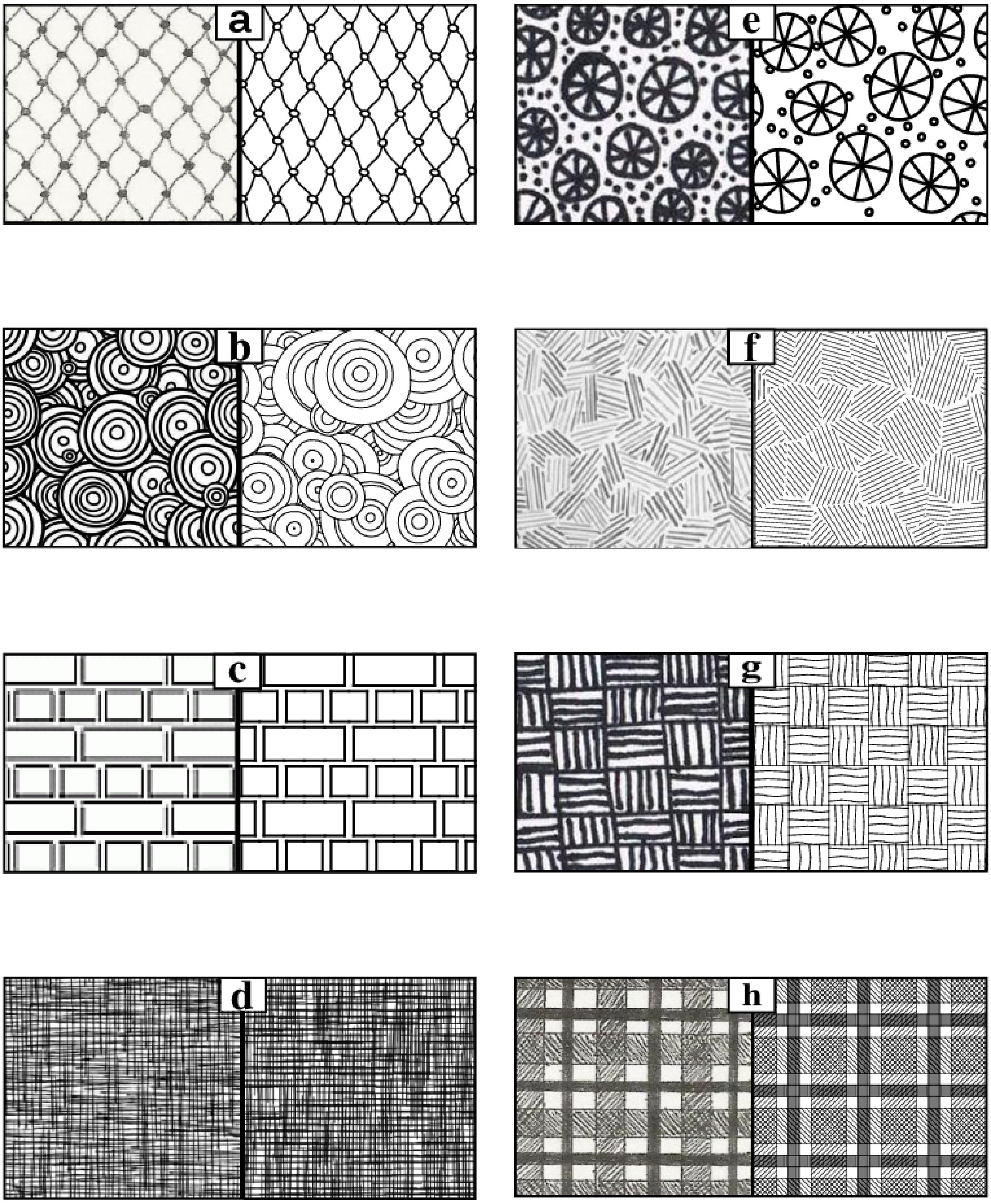


*Figure 6. **Element textures commonly used.** These textures can be found in professional art (d,g,h), casual art (a,e,f), technical productions such as Computer-Assisted Design illustration tools (c), and textile industry (b). For each example, we show a hand-drawn image (left), and our synthesized reproduction of its geometric arrangement (right). (a,b,c) Classic regular distributions with contact, overlap and no adjacency between elements respectively. (d) Overlap of two textures creating cross hatching. (e) Non overlapping combination of two textures. (f,g,h) Complex element textures with clusters of elements. — Image credit: (d,g,h) "Rendering in Pen and Ink: The Classic Book On Pen and Ink Techniques for Artists, Illustrators, Architects, and Designers" ; (a,e) Profusion Art [profusionart.blogspot.com]; (f) Hayes' Art Classes [hayesartclasses.blogspot.com]; (c) CompugraphX [www.compugraphx.com]; (b) 123Stitch [www.123stitch.com].*

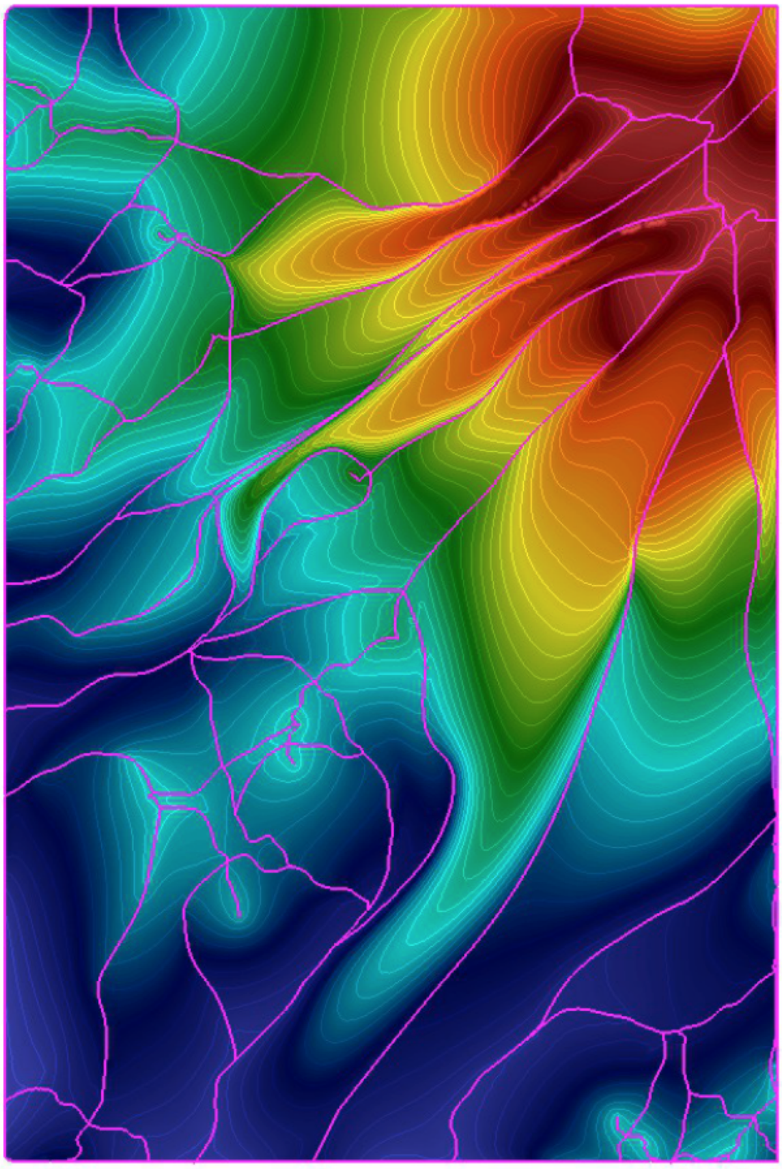


Figure 7. A function is reconstructed from its Morse-Smale complex (in purple). Inside the cells the function is monotonic so that no critical points are inserted, as can be seen from the isocontours in white. This technique is useful in Visualization whenever critical points in the data are important.

MIMETIC Project-Team

7. New Results

7.1. Biomechanics for motion analysis-synthesis

Participants: Charles Pontonnier, Georges Dumont, Steve Tonneau, Franck Multon, Julien Pettré, Ana Lucia Cruz Ruiz, Antoine Muller.

Ana-Lucia Cruz-Ruiz has been recruited as a PhD student since november 2013. The goal of this thesis is to define and evaluate muscle-based controllers for avatar animation. We developed an original control approach to reduce the redudancy 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 [9], [24].

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 [15], Calibrate those models thanks to experimental data [30] and run inverse dynamics to get joint torques from exeperimental motion capture data [14].

7.2. 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 genezalize the use of such devices for the design of workstations.

We deeply assessed the propensity of a virtual reality immersive room and classical interaction devices to evaluate properly the physical risk factors associated to assembly tasks. For this purpose, we compared tasks realized in real and virtual environment in terms of shoulder kinematics and muscular activity [20] and in terms of controlled kinematical variables, on the basis of the uncontrolled manifold theory [31]. Results show that there is less difference between real and virtual conditions than between indviduals, that make us think that such a virtual environment can be used to assess this type of task.

7.3. Interactions between walkers

Participants: Anne-Hélène Olivier, Armel Crétual, Julien Bruneau, Richard Kulpa, Sean Lynch, 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. This year, we developped new experiments in our immersive platform. First, we investigated obstacle avoidance behavior during real walking in a large immersive projection setup [22]. We analyze the walking behavior of users when avoiding real and virtual static obstacles. Indeed, CAVE-like immersive projection environments enable users to see both virtual and real objects, including the user's own body. With recent advances in VR technologies it becomes possible to build large-scale tracked immersive projection environments, which enable users to control their position in a large region of interest by real walking. In such environments virtual and real objects as well as multiple users or avatars may coexist in the same interaction space. Hence, it becomes important to gain an understanding of how the user's behavior is affected by the differences in perception and affordances of such real and virtual obstacles. We consider both anthropomorphic and inanimate objects, each having his virtual and real counterpart. The results showed that users exhibit different locomotion behaviors in the presence of real and virtual obstacles, and in the presence of anthropomorphic and inanimate objects. Precisely, the results showed a decrease of walking speed as well as an increase of the clearance distance (i. e., the minimal distance between the walker and the obstacle) when facing virtual obstacles compared to real ones. Moreover, users act differently due to their perception of the obstacle: users keep more distance when the obstacle is anthropomorphic compared to an inanimate object and when the orientation of anthropomorphic obstacle is from the profile compared to a front position. However, although we observed differences in collision avoidance behavior between real and virtual obstacles, which indicate biases of natural locomotion introduced by the setup, their magnitude seem lower compared to typical results found in HMD environments.This suggests that although the user's behavior in mixed environments varies depending on the nature of the stimulus, the user's locomotion behavior and the management of his/her interaction space is comparable with the ones in real life. Considering these findings, our results open promising vistas for using large CAVE-like setups for socio-physical experiments, in particular in the fields of locomotion and behavioral dynamics.

Second, we studied interactions between an individual and a crowd [7]. When avoiding a group, a walker has two possibilities: either he goes through it or around it. Going through very dense group or around huge one would not seem natural and could break any sense of presence in a virtual environment. The aim of this work was to enable crowd simulators to correctly handle such situations. To this end, we need understanding how real humans decide to go through or around groups. As a first hypothesis, we apply the Principle of Minimum Energy (PME) on different group sizes and density. According to it, a walker should go around small and dense groups while he should go through large and sparse groups. We quantified decision thresholds. However, PME left some inconclusive situations for which the two solutions paths have similar energetic cost. In a second part, we proposed an experiment to corroborate PME decisions thresholds with real observations. We proposed using Virtual Reality to enable accurately controlling experimental factors. We considered as well the role of secondary factors in inconclusive situations. We showed the influence of the group appearance and direction of relative motion in the decision process. Finally, we draw some guidelines to integrate our conclusions to existing crowd simulators and demonstrate that spectators can perceive some improvement in the crowd animation.

This year, we also developed new experiments in real conditions by considering the interaction between a walker and a moving robot. This worked 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 study 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 is 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. Humans preferentially give way to the robot, even if this choice is not optimal with regard to motion adaptation to avoid the collision. We proposed to interpret this behavior based on the notion of perceived danger and safety. Given the difficulty to understand how a robot behaves, and the lack of experience of interactions with the robot, humans apply a conservative avoidance strategy and prefer giving way to the robot. However, it is important to note that human participants succeed in perceiving the motion of the robot (anticipation was observed, no aberrant reaction occurred). One main conclusion is that, if we control robots to move like humans, we have a risk facing unexpected situations where robot compensates and cancels humans adaptations to the robot. A robot programmed to be cooperative could be perceived as hostile. The conclusion of this study opens paths for future research. A first direction is to better understand the possible effect of this notion of danger during interactions. We believe that this notion is of even higher importance when studying interactions with vehicles: a risk of collision with a fast vehicle obviously raises higher danger. A second direction is about the design of safe robots moving among human walkers. How the robot should adapt to others? Should it be collaborative with the risk of compensating human avoidance strategies? Should it be passive? We believe that robots should first be equipped with the ability to early detect humans avoidance strategy and adapt to it. In the near future, we want to continue our study of interactions between a robot and a human. In a first step, we plan to equip the robot with collision avoidance system which imitates real human strategies, and investigate how participants adapt to this new situation in comparison with a passive robot.

Finally, Sean Dean Lynch has been recruited as a PhD student since september 2015. This thesis concerns the visual perception of human motion during interactions in locmotor tasks. From the visual perception of someone's motion, we are able to predict the future course of this motion, interpret and anticipate his/her intentions and adapt our own motion to allow interactions. The main objective of the thesis is to identify the underlying perceptual mechanisms, i.e., the human motion cues which are necessary for an accurate understanding of others' intentions. It would allow to make significant progress in the understanding of human social behaviors. To reach these objectives, the thesis will be based on an experimental approach in virtual reality.

7.4. Motion Sensing

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 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.

Firstly, in ergonomics, we explored the use of low-cost mottion capture systems, 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. To tackle this problem, we propose [16] a new evaluation method based on a virtual mannequin. Thanks to this evaluation method, more than 500,000 configurations have been automatically tested, which is almost impossible to evaluate with classical protocols. The results show that the kinematic information obtained by the Kinect system is generally accurate enough to fill-in ergonomic assessment grids. However inaccuracy strongly increases for some specific poses and sensor positions. Using this evaluation method enabled us to report configurations that could lead to these high inaccuracies. Results obtained with the virtual mannequin are in accordance with those obtained with a real subject for a limited set of poses and sensor configuration. This knowledge can help to anticipate potential problems using a Kinect in given scenarios, and to propose methods to tackle these expected problems.

Secondly, 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 leads 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 [5] . 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 enables the clinician to have a rapid state of the quality of the gait. We proposed [4] a new method to asses 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 artifically 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 ($p < 0.001$), 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.

7.5. Virtual Human Animation

Participants: Julien Pettré, Franck Multon, Steve Tonneau.

Multiped locomotion in cluttered environments is addressed as the problem of planning acyclic sequences of contacts, that characterize the motion. In order

to overcome the inherent combinatorial difficulty of the problem, we separate it in two subproblems [34]: first, planning a guide trajectory for the root of the robot and then, generating relevant contacts along this trajectory. This paper proposes theoretical contributions to these two subproblems. We propose a theoretical characterization of the guide trajectory, named "true feasibility", which guarantee that a guide can be mapped into the contact manifold of the robot. As opposed to previous approaches, this property makes it possible to assert the relevance of a guide trajectory without explicitly computing contact configurations, as proposed in our previous works. This property can be efficiently checked by a sample-based planner (e.g. we implemented a visibility PRM). Since the guide trajectories that we characterized are easily mapped to a valid sequence of contacts, we then focused on how to select a particular sequence with desirable properties, such as robustness, efficiency and naturalness, only considered for cyclic locomotion so far. Based on these novel theoretical developments, we implemented a complete acyclic contact planner and demonstrate its efficiency by producing a large variety of movements with three very different robots (humanoid, insectoid, dexterous hand) in five challenging scenarios. The planner is very efficient in quality of the produced movements and in computation time: given a computed RB-PRM, a legged figure or a dexterous hand can generate its motion in real time. This result outperforms any previous acyclic contact planner.

7.6. VR and sports

Participants: Richard Kulpa, Benoit Bideau, Franck Multon, Anne-Hélène Olivier.

Athletes' performances are influenced by internal and external factors, including their psychological state and environmental factors, especially during competition. As a consequence, current training programs include stress management. In this work [3], we explore whether highly immersive systems can be used for such training programs. First, we propose methodological guidelines to design sport training scenarios both on considering the elements that a training routine must have and how external factors might influence the participant. The proposed guidelines are based on Flow and social-evaluative threat theories. Second, to illustrate and validate our methodology, we designed an experimental setup reproducing a 10 m Olympic pistol shooting. We analyzed whether changes in the environment are able to induce changes in user performance, physiological responses, and the subjective perception of the task. The simulation included stressors in order to raise a social-evaluative threat, such as aggressive public behavior or unforced errors, increasing the pressure while performing the task. The results showed significant differences in their subjective impressions, trends in the behavioral and physiological data were also observed. Taken together, our results suggest that highly immersive systems could be further used for training in sports.

Among the stimuli, visual information uptake is a fundamental element of sports involving interceptive tasks. Several methodologies, like video and methods based on virtual environments, are currently employed to analyze visual perception during sport situations. Both techniques have advantages and drawbacks. We made an experiment to determine which of these technologies may be preferentially used to analyze visual information uptake during a sport situation [21]. To this aim, we compared a handball goalkeeper's performance using two standardized methodologies: video clip and virtual environment. We examined this performance for two response tasks: an uncoupled task (goalkeepers show where the ball ends) and a coupled task (goalkeepers try to intercept the virtual ball). Variables investigated in this study were percentage of correct zones, percentage of correct responses, radial error and response time. The results showed that handball goalkeepers were more effective, more accurate and started to intercept earlier when facing a virtual handball thrower than when facing the video clip. These findings suggested that the analysis of visual information uptake for handball goalkeepers was better performed by using a 'virtual reality'-based methodology.

In a previous work, we analyzed the performance of beginners as they shot basketball free throws using various immersive conditions. Our results supported the assumption that natural complex motor behavior is possible in a VE, with little motor adaptation. The ultimate goal of our work is to design a VE training system for basketball free throws, so in this article we compare the performance of beginners making free throws in various visual conditions (first- versus third-person views using a large-screen immersive display) with that of expert players in the real world [8]. The key idea is to analyze how different visual conditions affect the performance of novices and to what extent it enables them to match the experts' performance.

Distance underestimation or any other perceptual disturbance in VR makes people adapt to the task at hand. The users in our study reached the same success rate by finding a new way to throw the ball, despite this incongruity between perception and action. The main observations reported in this article reinforce the conclusions in previous work, stating that 3PP is more efficient for certain tasks, but further work is required to test this result against other types of training conditions.

Finally, we worked on a transportable virtual reality system to analyse sports situations [6]. We proposed an original methodology to study the action of a goalkeeper facing a free kick. This methodology is based on a virtual reality setup in which a real goalkeeper is facing a virtual player and a virtual defensive wall. The setup has been improved to provide a total freedom of movement to the goalkeeper in order to have a realistic interaction between the goalkeeper and the player. The goalkeeper's movements are captured in real-time to accurately analyze his reactions. Such a methodology not only represents a valuable research tool but also provides a relevant training tool. Using this setup, this paper shows that goalkeepers are more performant during free kick with a wall composed of 5 defenders whatever its position.

7.7. Scheduling activities under spatial and temporal constraints

Participants: Fabrice Lamarche, Carl-Johan Jorgensen.

This work focusses on generating statistically consistent behaviors that can be used to pilot crowd simulation models over long periods of time, up to multiple days [1]. In real crowds, people's behaviors mainly depend on the activities they intend to perform. The way this activity is scheduled rely on the close interaction between the environment, space and time constraints associated with the activity and personal characteristics of individuals. Compared to the state of the art, our model better handle this interaction.

Our main contributions lie in the cdomain of activity scheduling and path planning. First, we proposed an individual activity scheduling process and its extension to cooperative activity scheduling. Based on descriptions of the environment, of intended activities and of agents' characteristics, these processes generate a task schedule for each agent. Locations where the tasks should be performed are selected and a relaxed agenda is produced. This task schedule is compatible with spatial and temporal constraints associated with the environment and with the intended activity of the agent and of other cooperating agents. It also takes into account the agents personal characteristics, inducing diversity in produced schedules. We showed that this model produces schedules statistically coherent with the ones produced by humans in the same situations. Second, we proposed a hierarchical path-planning process. It relies on an automatic environment analysis process that produces a semantically coherent hierarchical representation of virtual cities. The hierarchical nature of this representation is used to model different levels of decision making related to path planning. A coarse path is first computed, then refined during navigation when relevant information is available. It enable the agent to seamlessly adapt its path to unexpected events. Finally, those models have been included in a simulation platform that is able to simulate several thousand of pedestrians performing their daily activities in real-time. In order to deal with unexpected events, a process enabling adaptations of the pedestrian behavior have been designed. Those adaptations range from path modification to schedule adaptation according to the observed situation.

The proposed model handles long term rational decisions driving the navigation of agents in virtual cities. It considers the strong relationship between time, space and activity to produce more credible agents' behaviors. It can be used to easily populate virtual cities in which observable crowd phenomena emerge from individual activities.

7.8. Shoulder biomechanics

Participant: Armel Crétual [contact].

Shoulder hyperlaxity (SHL) is considered a main risk factor for shoulder instability and can be associated with different clinical shoulder instability presentations, such a multidirectional instability or unstable painful shoulder. Interestingly, quantification of shoulder laxity and hyperlaxity, particularly during physical examination, still remains an unsolved problem. Indeed, it is still frequently evaluated only through mono-axial amplitude, in particular using external rotation of the arm whilst at the side (ER1). We previously showed that this parameter is sensitive to inter-operator variability.

Therefore, we proposed a novel way to account for global shoulder mobility, the Shoulder Configuration Space Volume (SCSV) corresponding to the reachable volume in the configuration space of the shoulder joint [10]. In mechanics and robotics, the configuration space is the set of all reachable combination of coordinates. Considering the shoulder as the single joint between thorax and humerus instead of a combination of 4 actual joints (gleno-humeral, thoraco-humeral, scapulo-thoracic and sterno-clavicular), these coordinates are based upon the three joint angles defined by the International Society of Biomechanics (ISB) recommendations as plane of elevation orientation, elevation and axial rotation.

Then, this new index was examined through correlation to shoulder signs of hyperlaxity [19] for which we have shown a link with instability in patients who received a surgical procedure [18].

7.9. The Toric Space: a novel representation for camera control applications

Participants: Marc Christie, Christophe Lino, Quentin Galvane.

Many types of computer graphics applications such as data visualization or virtual movie production require users to position and move viewpoints in 3D scenes to effectively convey visual information or tell stories. The desired viewpoints and camera paths need to satisfy a number of visual properties (e.g. size, vantage angle, visibility, and on-screen position of targets). Yet, existing camera manipulation tools only provide limited interaction methods and automated techniques remain computationally expensive.

We introduce the *Toric space*, a novel and compact representation for intuitive and efficient virtual camera control. We first show how visual properties are expressed in this Toric space and propose an efficient interval-based search technique for automated viewpoint computation. We then derive a novel screen-space manipulation technique that provides intuitive and real-time control of visual properties. Finally, we propose an effective viewpoint interpolation technique which ensures the continuity of visual properties along the generated paths. The proposed approach (i) performs better than existing automated viewpoint computation techniques in terms of speed and precision, (ii) provides a screen-space manipulation tool that is more efficient than classical manipulators and easier to use for beginners, and (iii) enables the creation of complex camera motions such as long takes in a very short time and in a controllable way. As a result, the approach should quickly find its place in a number of applications that require interactive or automated camera control such as 3D modelers, navigation tools or games. The paper has been presented at SIGGRAPH 2015 (see [12] for more details).

We then rely on this Toric Space representation to contruct optimal camera paths (optimal in the satisfaction of visual properties along the path). Indeed, when creating real or computer graphics movies, the questions of how to layout elements on the screen, together with how to move the cameras in the scene are crucial to properly conveying the events composing a narrative. Though there is a range of techniques to automatically compute camera paths in virtual environments, none have seriously considered the problem of generating realistic camera motions even for simple scenes. Among possible cinematographic devices, real cinematographers often rely on camera rails to create smooth camera motions which viewers are familiar with. Following this practice, we have proposed a method for generating virtual camera rails and computing smooth camera motions on these rails. Our technique analyzes characters motion and user-defined framing properties to compute rough camera motions which are further refined using constrained-optimization techniques. Comparisons with recent techniques demonstrate the benefits of our approach and opens interesting perspectives in terms of creative support tools for animators and cinematographers. See [25] for more details.

TO address the more general problem of solving contradicting visual properties, novel ways of aggregating functions has also been proposed [33].

7.10. Data-driven Virtual Cinematography

Participant: Marc Christie.

Our propelling motivation here is to rely on existing data from real movies (automatically extracted or manually annotated), to propose better better and better framing techniques.

We first contributed to the problem of automated editing, by reproducing elements of cinematographic style. Automatically computing a cinematographic consistent sequence of shots over a set of actions occurring in a 3D world is a complex task which requires not only the computation of appropriate shots (viewpoints) and appropriate transitions between shots (cuts), but the ability to encode and reproduce elements of cinematographic style. Models proposed in the literature, generally based on finite state machine or idiom-based representations, provide limited functionalities to build sequences of shots. These approaches are not designed in mind to easily learn elements of cinematographic style, nor do they allow to perform significant variations in style over the same sequence of actions. We have proposed a model for automated cinematography that can compute significant variations in terms of cinematographic style, with the ability to control the duration of shots and the possibility to add specific constraints to the desired sequence. The model is parameterized in a way that facilitates the application of learning techniques. By using a Hidden Markov Model representation of the editing process, we have demonstrated the possibility of easily reproducing elements of style extracted from real movies. Results comparing our model with state-of-the-art first order Markovian representations illustrate these features, and robustness of the learning technique is demonstrated through cross-validation. See [13] for more details.

We also proposed a tool to ease the process of annotating cinematographic content, for the purposes of both film analysis, and film synthesis [29]. The work relies on the proposition of a film language that extends previous representations such as PSL (Prose Storyboard Language) by integrating the editing aspects, through the notion of cinematographic “techniques” described as patterns of shots.

The proposed language, named “Patterns”, is described in more details in [35]. Our language can express the aesthetic properties of framing and shot sequencing, and of camera techniques used by real directors. Patterns can be seen as the semantics of camera transitions from one frame to another. The language takes an editors view of on-screen aesthetic properties: the size, orientation, relative position, and movement of actors and objects across a number of shots. We have illustrated this language through a number of examples and demonstrations. Combined with camera placement algorithms, we demonstrated the language’s capacity to create complex shot sequences in data-driven generative systems for 3D storytelling applications.

7.11. Logic control in interactive storytelling

Participants: Marc Christie, Hui-Yin Wu.

With the rising popularity of engaging storytelling experiences in gaming arises the challenge of designing logic control mechanisms that can adapt to increasingly interactive, immersive, and dynamic 3D gaming environments. Currently, branching story structures are a popular choice for game narratives, but can be rigid, and authoring mistakes may result in dead ends at runtime. This calls for automated tools and algorithms for logic control over flexible story graph structures that can check and maintain authoring logic at a reduced cost while managing user interactions at runtime. In this work we introduce a graph traversal method for logic control over branching story structures which allow embedded plot lines. The mechanisms are designed to assist the author in specifying global authorial goals, evaluating the sequence of events, and automatically managing story logic during runtime. Furthermore, we showed how our method can be easily linked to 3D interactive game environments through a simple example involving a detective story with a flashback. See [36] for more details.

7.12. Automatic Continuity Editing for 3-D Animation

Participants: Marc Christie, Quentin Galvane, Christophe Lino.

We have proposed an optimization-based approach for automatically creating movies from 3-D animation. The method nicely separates the work of the virtual cinematographer (placing cameras and lights to produce nice-looking views of the action) from the work of the virtual film editor (cutting and pasting shots from all available cameras). While previous work has mostly focused on the first problem, the second problem has never been addressed in full details. We have reviewed the main causes of editing errors and built a cost function for minimizing them. We made a plausible semi-Markov assumption, which results in a computationally efficient dynamic programming solution. We showed that our method generates movies that avoid many common errors in film editing, including jump cuts, continuity errors and non-motivated cuts. We also show that our method can generate movies with different pacings. Combined with state-of-the-art cinematography, our approach therefore promises to significantly extend the expressiveness and naturalness of virtual movie-making. The work has been published at AAAI [27]. More details comparisons have been performed in [26].

MINT Project-Team

6. New Results

6.1. Physical and Perceptual Independence of Ultrasonic Vibration and Electrovibration for Friction Modulation

Eric Vezzoli, Wael Ben Messaoud, Michel Amberg, Betty Lemaire-Semail, Frédéric Giraud, Marie-Ange Bueno

Two different principles are available to modulate the user perceived roughness of a surface: electrovibration and ultrasonic vibration of a plate. The former enhances the perceived friction coefficient and the latter reduces it. In this work, we highlighted the independence of the two effects on the physical and perceptual point of view to confirm the increased range of sensation that can be supplied by the two coupled techniques. Firstly, a tribometric analysis of the induced lateral force on the finger by the two coupled effects has been achieved, then a study on the dynamic of the two effects is reported. In the end, a psychophysical experiment on the perception of the two coupled techniques confirms the approach.

6.2. Preliminary design of a multi-touch ultrasonic tactile stimulator

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

Currently there is no solution able to provide a multitouch tactile stimulation based on friction reduction tactile devices. The main objective of this work is to achieve a control method which allows to have a differentiated tactile stimulation on two fingers simultaneously, by superimposing two vibration modes. The proof of concept has been established on a 1D beam, where the tactile stimulation could be differentiated on two selected positions. We have presented the key design rules, as well as the control method. Finally, a psychophysical evaluation has shown that users can detect the location of nodes and antinodes of vibration with an average success rate of 78%

6.3. Generalised modal analysis for closed-loop piezoelectric devices

Christophe Giraud-Audine, Frédéric Giraud, Michel Amberg, Betty Lemaire-Semail.

Stress in piezoelectric material can be controlled by imposing the electrical field. Thanks to a feedback, this electrical field can be a function of some strain related measurement so as to confer to the piezoelectric device a closed loop macroscopic behaviour.We address the modelling of such system by extending the modal decomposition methods to account for the closed loop. To do so the boundary conditions are modified to include the electrical feedback circuit, hence allowing a closed-loop modal analysis. A case study is used to illustrate the theory and to validate it. The main advantage of the method is that design issue such as coupling factor of the device and closed loop stability are simultaneously captured.

6.4. Pressure dependence of friction modulation in ultrasonic devices

Wael Ben Messaoud, Eric Vezzoli, Frédéric Giraud, Betty Lemaire-Semail

Ultrasonic vibrating devices are able to modulate the friction of a finger sliding on them. The underlying principles of the friction reduction are still unclear, and this work is carried out to investigate the influence of the ambient pressure on the friction modulation. A specific tactile stimulator has been used for this purpose and the friction between the finger sliding on the device has been recorded for an ambient pressure of 0.5 and 1 atm showing a significant difference for comparable experimental conditions. A comparison with the model proposed in literature is performed underlying that the squeeze film interaction can be present but not the only responsible of the friction modulation in this kind of devices.

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. HCI models, theories, and frameworks

Participants: Géry Casiez, Alix Goguey, Stéphane Huot.

Pointing is one of the most common and frequent action made with any interactive system whether it be a desktop computer, a mobile device or a wall-size display. Although it has been extensively studied in HCI, current pointing techniques provide no adequate way to select very small objects whose movements are fast and unpredictable, and theoretical tools –such as Fitts' law–- do not model unpredictable motion. To inform the design of appropriate selection techniques, we studied human performance (speed and accuracy) when selecting moving objects in a 2D environment with a standard mouse. We characterized selection performance as a function of the predictability of the moving targets, based on three parameters: the speed (S) of the target, the frequency (F) at which the target changes direction, and the amplitude (A) of those direction changes. Our results show that for a given speed, selection is relatively easy when A and F are both low or high, and difficult otherwise [22].

In spite of previous work showing the importance of understanding users' strategies when performing tasks, HCI researchers comparing interaction techniques remain mainly focused on performance. This can be explained to some extent by the difficulty to characterize users' strategies. To alleviate this problem, we introduced new metrics to quantify if an interaction technique introduces an object or command-oriented strategy, i.e. if users favor completing the actions on an object before moving to the next one, or in contrast, if they are reluctant to switch between commands [21]. Through a study comparing two novel interaction techniques with two from the literature, we showed that our metrics allow to replicate previous findings on users' strategies concerning the latter.

To our knowledge, there are no general design and evaluation methodologies available for the development of digital musical instruments (DMI). One reason is the large diversity of design and evaluation contexts possible in musical interaction, e.g. is this evaluation done from the perspective of the DMI designer/manufacturer, the musician playing it, or the audience watching it be performed? With our collaborators of the MIDWAY associate team, we have analyzed all papers and posters published in the proceedings of the NIME conference from 2012 to 2014 [16]. For each publication that explicitly mentioned the term "evaluation", we looked for: a) What targets and stakeholders were considered? b) What goals were set? c) What criteria were used? d) What methods were used? e) How long did the evaluation last? Results show different understandings of evaluation, with little consistency regarding the usage of the word. Surprisingly in some cases, not even basic information such as goal, criteria and methods were provided. Beyond the attempt to provide an idea of what "evaluation" means for the NIME community, we pushed the discussion towards how we could make a better use of it and what criteria should be used regarding each goal.

7.3. Transfer functions and latency

Participants: Géry Casiez, Alix Goguey, Stéphane Huot, Sylvain Malacria, Nicolas Roussel.

Our work on transfer functions mainly focused this year on *edge-scrolling*, which allows users to scroll a viewport while simultaneously dragging near or beyond its edge. Common implementations rely on rate control, mapping the distance between the pointer and the edge of the viewport to the scrolling velocity. While ubiquitous in operating systems, edge-scrolling had received little attention, though previous works suggested that rate control may be suboptimal for isotonic pointing devices (e.g. mice and touchpads) and space beyond the window's edge might be scarce, limiting scrolling control. To address these problems, we developed *Push-Edge* and *Slide-Edge* two position-based techniques that allow scrolling by "pushing" against the viewport edge [23]. A controlled experiment shows that our techniques reduce overshoots and offer performance improvements up to 13% over traditional edge-scrolling.

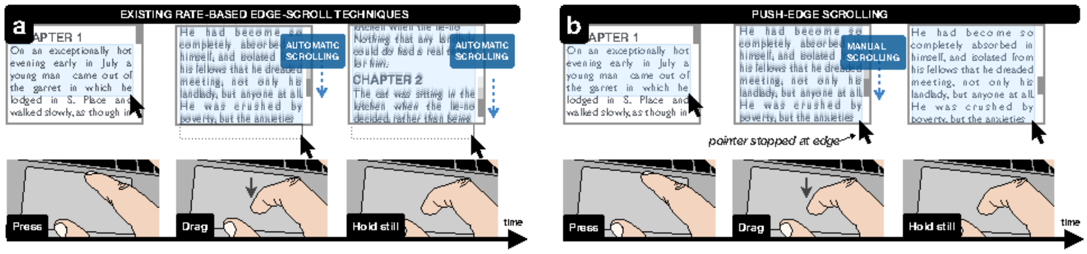


Figure 1. When selecting text with a touchpad, downward movements after crossing the viewport edge will (a) change the rate of automatic scrolling with existing techniques (rate control); or (b) manually scroll the document, stopping the pointer at the edge with push-edge scrolling (position control).

Our work on latency focused on its measurement in existing graphical user interfaces, a problem for which we developed a simple method [18], [27]. Our method consists in positioning an unmodified optical mouse on the screen while displaying and translating a particular texture to fake mouse displacements, which results in controlled mouse events. This works with most optical mice and allows accurate and real-time latency measures up to 5 times per second. The method also allows easy insertion of probes at different places in the system to investigate the sources of latency. Measurements performed on different systems, toolkits and applications notably showed that latency is affected by the operating system and system load. Substantial differences were also found between C++/GLUT and C++/Qt or Java/Swing implementations, as well as between Web browsers.

7.4. Interaction techniques

Participants: Géry Casiez, Stéphane Huot, Sylvain Malacria.

While touchpads are both widespread and expressive input devices, there has been surprisingly little research regarding how they could be used for more than simple pointer movements. In [17], we explore the design space of gesture shortcuts on touchpads and introduce four novel interaction techniques. *SpotPad* and *LociPad* rely on one-finger static gestures, but differ in their graphical representation. *ChordPad* relies on two-finger static gestures with a hierarchical representation. Finally, *InOutPad* relies on dynamic gestures crossing the edges of the touchpad. We compare the properties of these four interaction techniques and describe how they can be deployed on OS X.

The hands of virtual characters are highly complex 3D models that can be tedious and time-consuming to animate with current methods. In [14], we introduce *THING*, a novel tablet-based approach that leverages multi-touch interaction for the quick and precise control of a 3D hand's pose. The flexion/extension and abduction/adduction of the virtual fingers can be controlled individually for each finger or for several fingers in parallel through sliding motions on the tablet. We describe two variants of THING: MobileTHING, which maps the spatial location and orientation of the tablet to that of the virtual hand, and DesktopTHING, which combines multi-touch controls of fingers with traditional mouse controls for the hand's global position and orientation. We also report on two usability studies in which we compared THING to mouse-only controls and a data glove.

Figure 2. THING enables the control of 3D hand models (in blue) by sliding fingers along sliders arranged in a morphologically-consistent pattern on the tablet's screen.

Interactive technologies have radically changed the way visual artists work, and a large portion of the artistic production has now moved from paper to the computer. However, many artists still work on paper and keep using traditional painting and drawing tools. This is not only due to resistance to progress or due to the well-known usability properties of physical tools: despite the use of pen displays, the progress of artistic stroke-rendering techniques, and the powerful and advanced functionalities of existing computer tools, they fail to fully capture the richness and variety of artistic styles supported by physical media. We interviewed four professional illustrators in their work environment. We also followed the work of an artist for a two-year period. We observed that artists mix a variety of techniques that involve specialized computer software and hardware such as Adobe Photoshop, a graphics tablet and a scanner, and traditional physical tools such as pencils, paper, and customized light tables. Our findings inspired *BricoSketch* [25], an augmented paper interface that enables illustrators to zoom into parts of their drawings and work at different levels of detail on paper. Our early results show that BricoSketch supports real tasks, improving productivity on paper while enhancing illustrators' creative ways of working.

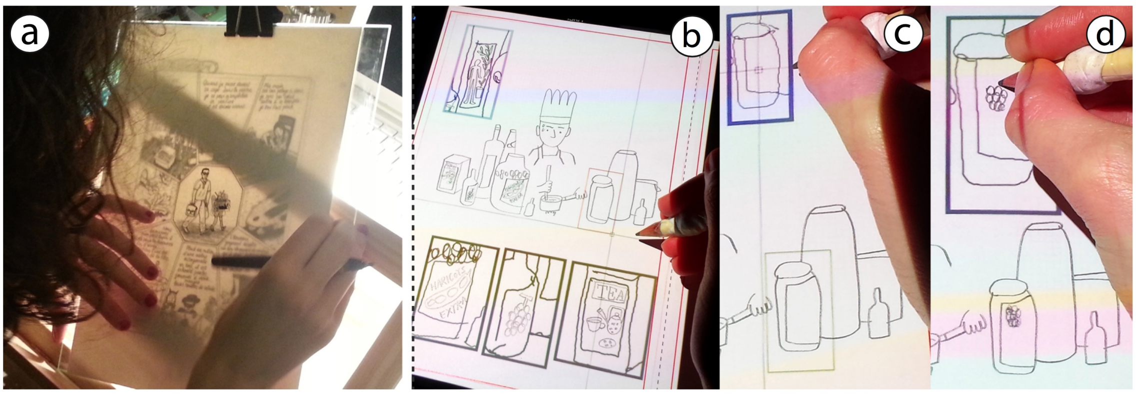


Figure 3. BricoSketch. (a) An artist works on layers of physical paper with a home-made light table: she draws the panels of a page for a graphic novel by using earlier sketches as guides. (b-d) The same artist uses our system to add details to different parts of her illustration through partial scaled views.

7.5. Interactive visualization

Participant: Fanny Chevalier.

The differential diagnosis of hereditary disorders is a challenging task for clinicians due to the heterogeneity of phenotypes that can be observed in patients. Existing clinical tools are often text-based and do not emphasize consistency, completeness, or granularity of phenotype reporting, which can impede clinical diagnosis and limit their utility to genetics researchers. The *PhenoBlocks* tool described in [13] is a novel visual analytics platform designed to support clinical differential diagnosis. It supports the comparison of phenotypes between patients, or between a patient and the hallmark features of a disorder. An informal evaluation with expert clinicians suggests that the visualization effectively guides the process of differential diagnosis and could reinforce the importance of complete, granular phenotypic reporting.

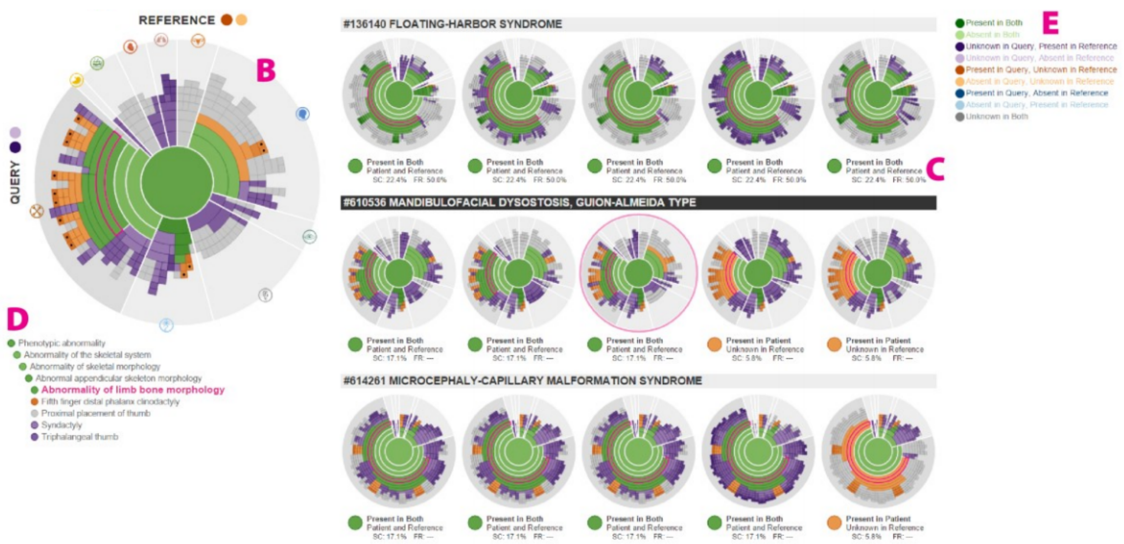


Figure 4. PhenoBlocks allows to compare the phenotype hierarchies of an undiagnosed query patient to a diagnosed reference patient. During differential diagnosis, clinicians use shared (green) phenotypes to gauge confidence in their diagnostic hypothesis and missing (purple) phenotypes to identify candidates for subsequent analysis.

7.6. Adaptive interfaces

Participant: Sylvain Malacria.

As news is increasingly accessed on smartphones and tablets, the need for personalizing news applications is apparent. In [19], [20], we report on a series of studies addressing key issues in the development of adaptive news interfaces. We first surveyed users' news reading preferences and behaviors. We then implemented and deployed an Android application that logs users' interactions with the application. We used the logs to train a classifier and showed that it is able to reliably recognize a user according to their reader type. Finally we evaluated alternative, adaptive user interfaces for each reader type. The evaluation demonstrates the differential benefit of the adaptation for different users and the feasibility of adaptive interfaces for news applications.

In [12], we investigate the use of a companion application on a tablet to augment viewing of information-rich television programs. The application displays a synchronized graphical abstraction of the program's content in the form of a concept map. Two experiments were conducted involving participants watching an astronomy documentary. Results show that the companion application improved participants' understanding and recall of the program. Participants were found to manage their visual attention systematically when using the companion application, and correlations were found in the way they shifted their gaze from TV screen to tablet and back in response to changes in the program content. Increasing interaction with the application disrupted understanding of the television program and visual attention. Participants were positive about the value of companion applications for understanding and recall of programs, but distraction and "knowing where to look" were significant concerns.

POTIOC Project-Team

7. New Results

7.1. Pointing in Spatial Augmented Reality from 2D Pointing Devices

Participants: Renaud Gervais, Jérémy Frey, Martin Hachet.

Figure 3. A user reaches a target displayed on a spatially augmented object with an indirect input device

Spatial Augmented Reality (SAR) opens interesting perspectives for new generations of mixed reality applications. Compared to traditional human-computer interaction contexts, there is little work that studies user performance in SAR. In this project, we present an experiment that compares pointing in SAR versus pointing in front of a screen, from standard pointing devices (mouse and graphics tablet). The results showed that the participants tend to interact in SAR in a way that is similar to the screen condition, without a big loss of performance [30] (See Figure 3).

7.2. Tangible Viewports

Participants: Renaud Gervais, Joan Sol Roo, Martin Hachet.

Spatial augmented reality and tangible interaction enrich the standard computer I/O space. Systems based on such modalities offer new user experiences and open up interesting perspectives in various fields. On the other hand, such systems tend to live outside the standard desktop paradigm and, as a consequence, they do not benefit from the richness and versatility of desktop environments. In this work, we propose to join together physical visualization and tangible interaction within a standard desktop environment. We introduce the concept of Tangible Viewport, an on-screen window that creates a dynamic link between augmented objects and computer screens, allowing a screen-based cursor to move onto the object in a seamless manner (Figure 4). We describe an implementation of this concept and explore the interaction space around it. A preliminary evaluation shows that the metaphor is transparent to the users while providing the benefits of tangibility [31].

7.3. Tobe

Participants: Renaud Gervais, Jérémy Frey, Alexis Gay, Fabien Lotte, Martin Hachet.

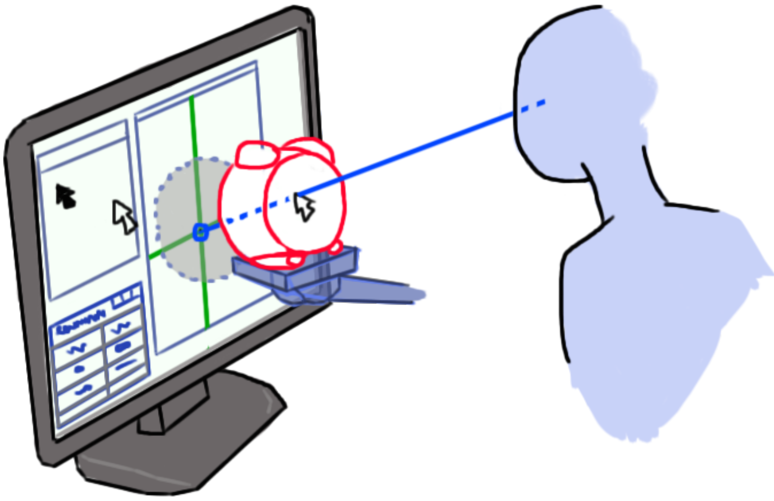


Figure 4. A user interacts with an object located in front of the screen as if the object was rendered on screen

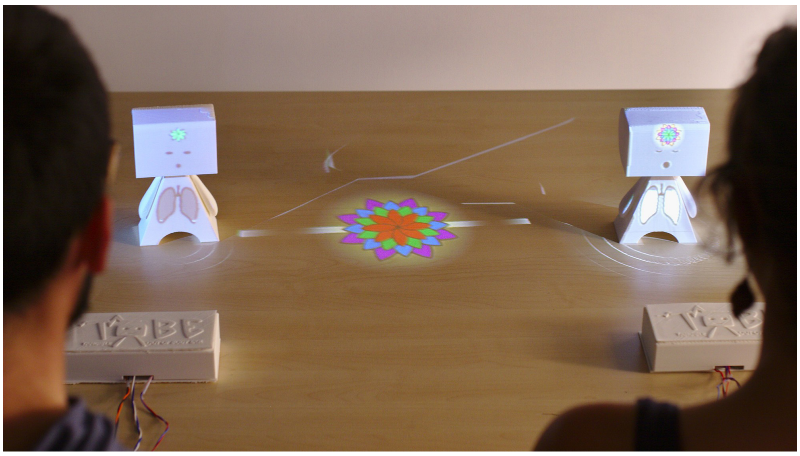


Figure 5. Two users are relaxing together using Tobe as a biofeedback for heartrate and breathing

We propose Tobe, a toolkit for creating Tangible Out-of-Body Experiences: exposing the inner states of users using physiological signals such as heart rate or brain activity. Tobe can take the form of a tangible avatar displaying live physiological readings to reflect on ourselves and others. Such a toolkit could be used by researchers and designers to create a multitude of potential tangible applications, including (but not limited to) educational tools about Science Technologies Engineering and Mathematics (STEM) and cognitive science, medical applications or entertainment and social experiences with one or several users or Tobes involved. Through a co-design approach, we investigated how everyday people picture their physiology and we validated the acceptability of Tobe in a scientific museum. We also give a practical example where two users relax together, with insights on how Tobe helped them to synchronize their signals and share a moment, as illustrated in Figure 5 [29].

7.4. Inner Garden

Participants: Joan Sol Roo, Renaud Gervais, Martin Hachet.

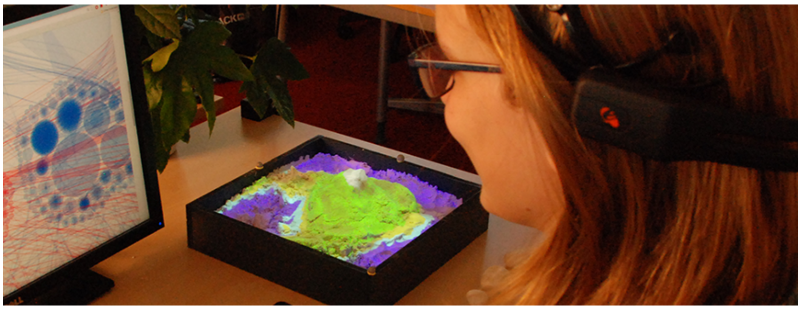


Figure 6. Inner garden is an augmented sandbox which depicts an evolving world reflecting the inner state of the user

We present a prototype of an augmented sandbox where the sand is used to create a miniature living world, designed as an ambient display for contemplation and self-reflection. The landscape can be reshaped at any time. Once the sand is left still for a moment, the world starts evolving – vegetation grows, water flows and creatures move around – according to the user's internal state. We use a consumer-grade EEG and breathing sensors to reflect on frustration and meditative states of users, which they can monitor by looking at the sandbox (Figure 6) [49].

7.5. Augmented geographic maps

Participants: Julia Chatain, Marie Demangeat, Anke Brock, Martin Hachet.

Interactive geographic maps are today widely available, but remain mostly limited to standard interaction contexts. We introduce SyMAPse [48], a spatial augmented reality map, based on the PapARt framework. In our prototype, we use augmented reality to display a virtual map on a physical piece of paper, thus keeping features of both media. Thanks to the digital map base, users can pan, zoom and even change the basemap. At the same time, the paper base allows users to manipulate the map physically and so to interact in a more "natural" way, as well as to draw on the paper using regular pens. In a preliminary study with visitors of the "Cap Sciences" science center, we compared interaction techniques based on touch, tangible and spatial modalities for these three common map functions: zooming, panning, and changing the basemap. Our results suggest that object-based and spatial interaction may be advantageous over touch in our augmented reality setup.

Figure 7. Interacting with an augmented geographic map using tangible, spatial and multitouch interaction

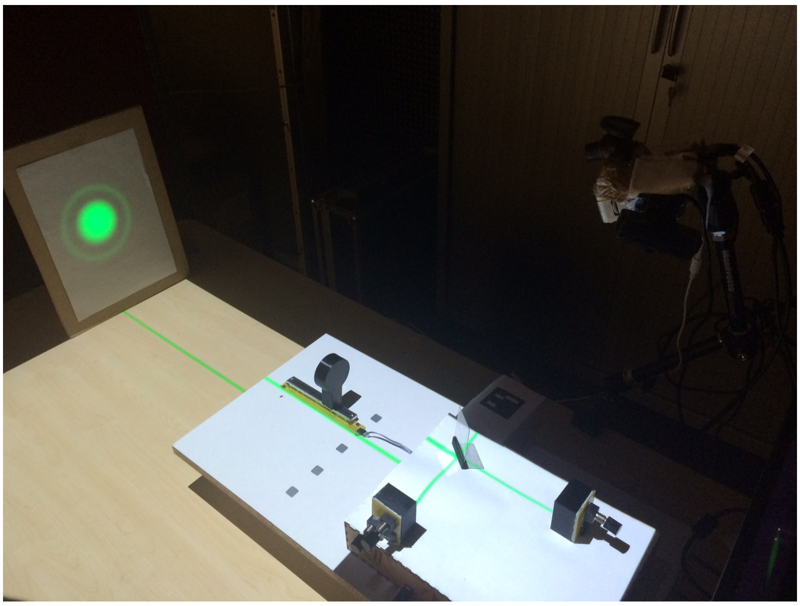


Figure 8. HOBIT mixes physical and virtual elements to teach optics.

7.6. HOBIT - Hybrid Optical Bench for Innovative Teaching

Participants: David Furio, Benoit Coulais, Martin Hachet.

Experiments in optics are essential for learning and understanding physical phenomena. The problem with these experiments is that they are generally time consuming for both their construction and their maintenance, potentially dangerous through the use of laser sources, and often expensive due to high technology optical components. We propose to simulate such experiments by using hybrid systems that exploit both spatial augmented reality and tangible interaction (See Figure 8). In particular, we focus on one of the most popular optical experiments: Michelson interferometer. In our approach, we target a highly interactive system where students are able to interact in real time with the Augmented Michelson Interferometer (AMI) to observe, test hypotheses and then to enhance their comprehension. Compared to a fully digital simulation, we are investigating an approach that benefits from both physical and virtual elements, and where the students experiment by manipulating 3D-printed physical replicas of optical components (e.g. lenses and mirrors). Our objective is twofold. First, we want to ensure that the students will learn with our simulator the same concepts and skills that they learn with traditional methods. Second, we hypothesis that such a system opens new opportunities to teach optics in a way that was not possible before, by manipulating concepts beyond the limits of observable physical phenomena. To reach this goal, we have built a complementary team composed of experts in the field of optics, human-computer interaction, computer graphics, sensors and actuators, and education science. HOBIT is a joint project between Inria and Université de Bordeaux (Idex CPU – LAPHIA), in collaboration with Université de Lorraine (team PErSEUs). [28]

7.7. Mixed Reality to improve children's interaction with astronomical concepts

Participant: Martin Hachet.

This project stands on a collaboration with Stéphanie Fleck from Université de Lorraine. To make astronomical learning more efficient for young pupils, we have designed an Augmented Inquiry-Based Learning Environment (AIBLE): HELIOS. Because manipulations in astronomy are intrinsically not possible, we propose to manipulate the underlying model. With HELIOS, virtual replicas of the sun, moon and earth are directly manipulated from tangible manipulations. This digital support combines the possibilities of Augmented Reality (AR) while maintaining intuitive interactions following the principles of didactic of sciences. Light properties are taken into account and shadows of Earth and Moon are directly produced by an omnidirectional light source associated to the virtual Sun. This AR environment provides users with experiences they would otherwise not be able to experiment in the physical world. Our main goal is that students can take active control of their learning, express and support their ideas, make predictions and hypotheses, and test them by conducting investigations. [24][23]

7.8. Combining and Revealing Spaces for Musical Performances

Participant: Martin Hachet.

In collaboration with University of Bristol (Florent Berthaut, Diego Martinez, and Sriram Subramanian) we have designed a mixed-reality environment for musical performances that allows for freely displaying virtual content on stage, such as 3D virtual musical interfaces or visual augmentations of instruments and performers. This environment, called Reflets, relies on spectators and performers revealing virtual objects by slicing through them with body parts or objects, and on planar slightly reflective transparent panels that combine the stage and audience spaces. It allows for placing virtual content anywhere on large stages, even overlapping with physical elements and provides a consistent rendering of this content for large numbers of spectators. It also preserves non-verbal communication between the audience and the performers, and is inherently engaging for the spectators. Reflets opens musical performance opportunities such as augmented interaction between musicians and novel techniques for 3D sound shapes manipulation [20].

7.9. Improving User-Training for Brain-Computer Interfaces

Participants: Martin Hachet, Emilie Jahanpour, Camille Jeunet, Fabien Lotte, Boris Masencal, Julia Schumacher.

While Mental Imagery based Brain-Computer Interfaces (MI-BCIs) are promising for many applications, ranging from assistive technologies for motor disabled patients to video games, their usability"out-of-the-lab" has been questioned due to their lack of reliability: literature reports that 15% to 30% of users cannot control such a technology, while most of the remaining users obtain only modest performances. As controlling an MI-BCI requires the acquisition of specific skills (i.e., producing stable and distinct brain-activity patterns), an adapted training is necessary. Thus, the main objective of our project is to improve the user training to facilitate the acquisition of MI-BCI related skills. In order to do so, we focused on two axes [18]: (1) the impact of the user-profile and (2) the impact of the protocol on MI-BCI performance.

Concerning the impact of the user-profile, our results ([40], [14]) suggested an important impact of some aspects of the personality (such as the tension and autonomy levels) as the spatial abilities (i.e., the ability to produce, interpret and transform mental imageries). On the one hand, we are working on learning companions, whose goal would be to provide the learners with a specific emotional support, based on their profile and on their cognitive state. On the other hand, we are currently implementing and testing a spatial ability training in order to test the hypothesis of a causal effect of the spatial abilities on MI-BCI performance [39]. In other words, we would like to know if increasing spatial abilities would result in better MI-BCI performance. One application of such a research is stroke rehabilitation. Indeed, motor after-effects are usual following a stroke. MI-BCI have been shown very useful to facilitate the rehabilitation process, which consists in enhancing brain plasticity through motor-imagery, as they enable to visualise the BCI activity while the patients perform MI-tasks. However, MI-tasks tend to increase the depressive state of the patients as they remind them they lost the ability to move their limb. Thus, as spatial ability exercises (e.g., mental rotation) activate the motor cortex, they could be used as more transparent rehabilitation exercises to trigger brain plasticity.

Second, concerning the impact of the protocol, we completed a study (see activity report 2014) in which we asked the participants to use the standard MI-BCI training protocol to learn to perform simple motor tasks: drawing circles and triangles on a graphic tablet. As it would have been the case for an MI-BCI experiment, they had to find the right strategy so that the system recognises the task they were performing. Seventeen percent of the participants (N=54) showed difficulties in performing these tasks. Also, when we selected the 10 best and 10 worst performers of this experiment and asked them to use an MI-BCI (by imagining left and right-han movements), it appeared that the ones who had difficulties in performing the simple motor tasks improved in terms of performance during the MI-BCI experiment, while the participants who performed well during the motor experiment did not progress during the second. Furthermore, we have shown that tactile feedback was more efficient than an equivalent visual feedback in a multitasking context [32]. Based on a literature review, this could be due to an increased sense of agency (i.e., the feeling to be in control). We are thus currently exploring the impact of the sense of agency on MI-BCI performance. Finally, still regarding the feedback, we explored what kind of information could help the user to perform better mental imagery tasks. As such, we look for physiological features that could predict whether a mental task will be correctly recognized by the BCI, and that could be understood by the user. Among the different features we explored, it appears that the user's relaxation (from a muscular point of view), as measured in EMG activity collected by EEG channels, is one of such features. We are currently building and exploring new BCI training protocols that provide additional information about the user's muscular relaxation as complementary feedback [34].

7.10. EEG Signal Processing

Participant: Fabien Lotte.

To make BCI practical and useful, we need to make them reliable, i.e., able to recognize the users' mental commands, despite noise and non-stationarities [42]. We also need to reduce their calibration time, as current systems needs many examples from each user to calibrate the system for this specific user. This year we addressed these two issues with two different studies.

In order to reduce BCI calibration times, we first surveyed existing approaches, these approaches being notably based on regularization, user-to-user transfer, semi-supervised learning and a-priori physiological information. We then proposed new tools to reduce BCI calibration time. In particular, we proposed to generate artificial EEG trials from the few EEG trials initially available, in order to augment the training set size. These artificial EEG trials are obtained by relevant combinations and distortions of the original trials available. We proposed 3 different methods to do so. We also proposed a new, fast and simple approach to perform user-to-user transfer for BCI. Finally, we studied and compared offline different approaches, both old and new ones, on the data of 50 users from 3 different BCI data sets. This enabled us to identify guidelines about how to reduce or suppress calibration time for BCI [16].

In order to increased BCI robustness, we performed an empirical comparison of covariance matrix averaging methods for EEG signal classification. Indeed, averaging EEG signal covariance matrices is a key step in designing brain-computer interfaces (BCI) based on the popular common spatial pattern (CSP) algorithm. BCI paradigms are typically structured into trials and we argue that this structure should be taken into account. Moreover, the non-Euclidean structure of covariance matrices should be taken into consideration as well. We reviewed several approaches from the literature for averaging covariance matrices in CSP and compared them empirically on three publicly available data sets. Our results showed that using Riemannian geometry for averaging covariance matrices improves performances for small dimensional problems, but also the limits of this approach when the dimensionality increases [36].

7.11. ECoG-based analysis of Speech processes

Participant: Fabien Lotte.

Acoustic speech output results from coordinated articulation of dozens of muscles, bones and cartilages of the vocal mechanism. While we commonly take the fluency and speed of our speech productions for granted, the neural mechanisms facilitating the requisite muscular control are not completely understood. Previous neuroimaging and electrophysiology studies of speech sensorimotor control has typically concentrated on speech sounds (i.e. phonemes, syllables and words) in isolation; sentence-length investigations have largely been used to inform coincident linguistic processing. In this study, we examined the neural representations of segmental features in the context of fluent, continuous speech production. We used recordings from the cortical surface (electrocorticography (ECoG)) to simultaneously evaluate the spatial topography and temporal dynamics of the neural correlates of speech articulation that may mediate the generation of hypothesized gestural or articulatory scores. We found some aspects of speech production (place of articulation) involved broad networks of brain regions during all phases of speech production: preparation, execution and monitoring. Other aspects (manner of articulation and voicing status) were dominated by auditory cortical responses after speech had been initiated. These results provide a new insight into the articulatory and auditory processes underlying speech production in terms of their motor requirements and acoustic correlates (see Figure 9 , [15]).

7.12. Toward a portable tangible EEG interface

Participants: Maxime Duluc, Thibault Laine, Jérémy Frey, Renaud Gervais, Fabien Lotte, Martin Hachet.

Last year we presented Teegi, the first interface 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.

Yet, the gap between research projects and the field is not often bridged. While our past prototype used an external projector and a supplementary tracking device to display information onto the head of the puppet, over the course of the year we developed a semi-spherical display based on LEDs (see Figure 10). By embedding all the electronics into the puppet, Teegi will be easier to bring outside the laboratory. Thanks to these technological advances, real-life applications of the system are finally within reach.

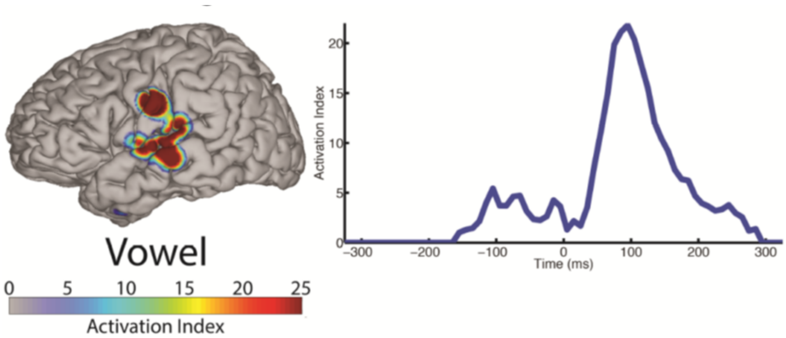


Figure 9. Exemple of the ECoG signature of vowel phonemes.

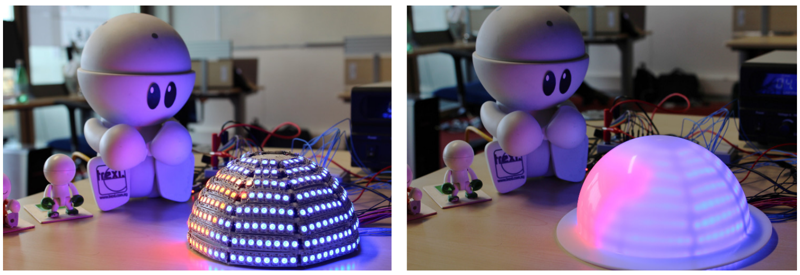


Figure 10. Development version of a portable version of Teegi, a Tangible EEG Interface. An array of LEDs and a diffuser replace the use of spatial augmented reality.

7.13. Electroencephalography-based evaluation of user experience

Participants: Jérémy Frey, Maxime Daniel, Dennis Wobrock, Julien Castet, Martin Hachet, Fabien Lotte.

Designing user interfaces requires adequate evaluation tools to ensure good usability and user experience. While many evaluation tools are already available and widely used, existing approaches generally cannot provide continuous and objective measures of usability qualities during interaction without interrupting the user. On the other hand, the measure of brain activity by the mean of electroencephalography (EEG) is mature enough to assess mental states. Combined with existing methods, such tool can be used to strengthen the understanding of user experience.

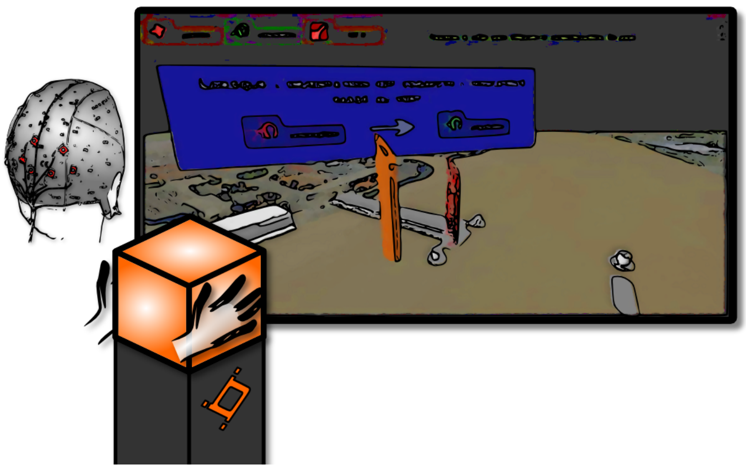


Figure 11. Schematic view of a user performing 3D manipulations tasks with the CubTile input device. His/her mental effort are monitored based on brain signals (electroencephalography).

In [35] we studied 3D object manipulation tasks. We showed how mental workload can be estimated from EEG, and then measured it on 8 participants during an actual 3D object manipulation task with an input device known as the CubTile (see figure 11). These first results suggested that we could continuously assess the 3DUI and/or interaction technique ease-of-use.

Figure 12. A keyboard (left) can be compared with a touch interface (middle) using a continuous measure of mental workload through electroencephalography (right).

We pushed further these finding in a second study [26], where we have developed a set of methods to continuously estimate the user's mental workload, attention level and recognition of interaction errors during different interaction tasks. We validated these measures in a controlled virtual environment and showed how they can be used to compare different interaction techniques – for instance a keyboard and a touch-based interface (see Figure 12).

Thanks to such framework, EEG becomes a useful addition to the repertoire of available evaluation tools, enabling a finer grain assessment of the ergonomic qualities of computer systems.

7.14. Classifying EEG Signals during Stereoscopic Visualization to Estimate Visual Comfort

Participants: Jérémy Frey, Aurélien Appriou, Fabien Lotte, Martin Hachet.

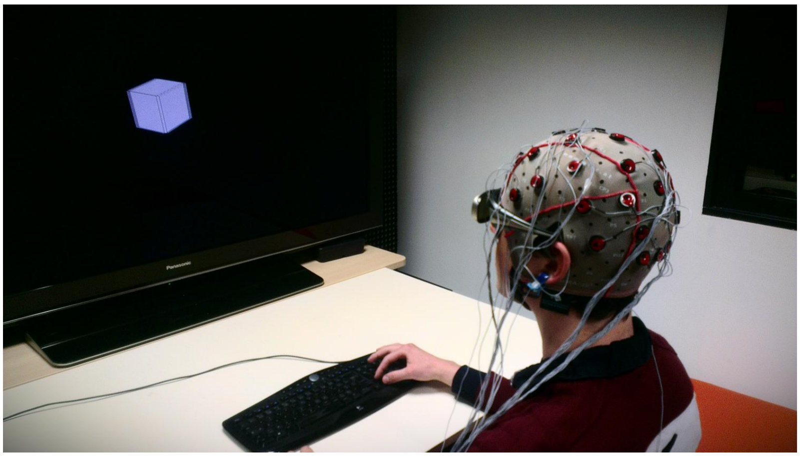


Figure 13. Setup of the experiment, with a subject being presented with stereoscopic images while his EEG signals are being recorded.

With stereoscopic displays a sensation of depth that is too strong can impede visual comfort and may result in fatigue or pain. We used Electroencephalography (EEG) to develop a novel brain-computer interface that monitors users' states in order to reduce visual strain. We present the first system that discriminates comfortable conditions from uncomfortable ones during stereoscopic vision of still images using EEG [13], [25] – see Figure 13). In particular, we show that changes in event-related potentials' (ERPs) amplitudes following stereoscopic objects presentation can be used to estimate visual comfort. Our system reacts within 1 second to depth variations, achieving 63% accuracy on average (up to 76%) and 74% on average when 7 consecutive variations are measured (up to 93%). Performances are stable (≈62.5%) when a simplified signal processing is used to simulate online analyses or when the number of EEG channels is lessened. This study could lead to adaptive systems that automatically suit stereoscopic displays to users and viewing conditions. For example, it could be possible to match the stereoscopic effect with users' state by modifying the overlap of left and rAight images according to the classifier output

TITANE Project-Team

7. New Results

7.1. Analysis

7.1.1. Planar Shape Detection and Regularization in Tandem

Participants: Sven Oesau, Florent Lafarge, Pierre Alliez.

In collaboration with EADS ASTRIUM

We contributed a method for planar shape detection and regularization from raw point sets. The geometric modeling and processing of man-made environments from measurement data often relies upon robust detection of planar primitive shapes. In addition, the detection and reinforcement of regularities between planar parts is a means to increase resilience to missing or defect-laden data as well as to reduce the complexity of models and algorithms down the modeling pipeline. The main novelty behind our method is to perform detection and regularization in tandem. We first sample a sparse set of seeds uniformly on the input point set, then perform in parallel shape detection through region growing, interleaved with regularization through detection and reinforcement of regular relationships (coplanar, parallel and orthogonal). In addition to addressing the end goal of regularization, such reinforcement also improves data fitting and provides guidance for clustering small parts into larger planar parts (Figure 1). We evaluate our approach against a wide range of inputs and under four criteria: geometric fidelity, coverage, regularity and running times. Our approach compares well with available implementations such as the efficient RANSAC-based approach proposed by Schnabel and co-authors in 2007 [8]. This work has been published in the Computer Graphics Forum journal.

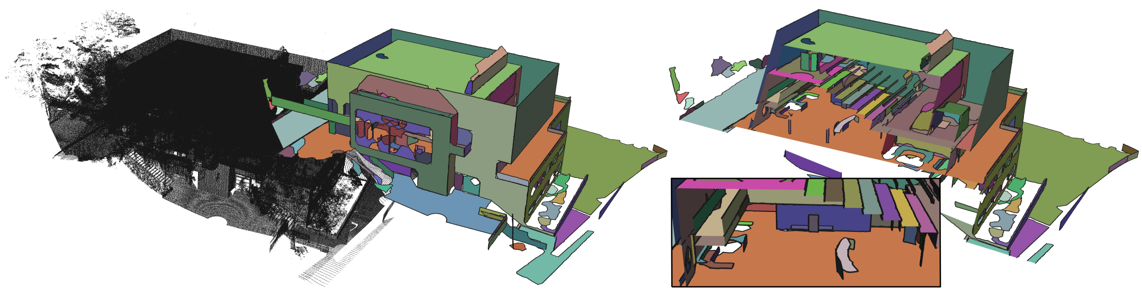


Figure 1. Shape detection and regularization. The input point set (5.2M points) has been acquired via a LIDAR scanner, from the inside and outside of a physical building. 200 shapes have been detected, aligned with 12 different directions in 179 different planes. The cross section depicts the auditorium in the upper floor and the entrance hall in the lower floor. The closeup highlights the steps of the auditorium which are made up of perfectly parallel and orthogonal planes.

7.1.2. Image partitioning into convex polygons

Participants: Liuyun Duan, Florent Lafarge.

In collaboration with Geoimage

The over-segmentation of images into atomic regions has become a standard and powerful tool in Vision. Traditional superpixel methods, that operate at the pixel level, cannot directly capture the geometric information disseminated into the images. We propose an alternative to these methods by operating at the level of geometric shapes. Our algorithm partitions images into convex polygons. It presents several interesting properties in terms of geometric guarantees, region compactness and scalability. The overall strategy consists in building a Voronoi diagram that conforms to preliminarily detected line-segments, before homogenizing the partition by spatial point process distributed over the image gradient. Our method is particularly adapted to images with strong geometric signatures, typically man-made objects and environments (Figure 2). We show the potential of our approach with experiments on large-scale images and comparisons with state-of-the-art superpixel methods [17]. This work has been published in the Computer Graphics Forum journal. Published in the proceedings of CVPR (IEEE conference on Computer Vision and Pattern Recognition).

Figure 2. Image partitioning into convex polygons.

7.1.3. 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 [22] (see Figure 3).

7.1.4. *Optimizing partition trees for multi-object segmentation with shape prior*

Participants: Emmanuel Maggiori, Yuliya Tarabalka.

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

Partition trees, multi-class segmentation, shape priors, graph cut.

A partition tree is a hierarchical representation of an image. Once constructed, it can be repeatedly processed to extract information. Multi-object multi-class image segmentation with shape priors is one of the tasks that can be efficiently done upon an available tree. The traditional construction approach is a greedy clustering based on color similarities. However, not considering higher level cues during the construction phase leads to trees that might not accurately represent the underlying objects in the scene, inducing mistakes in the later segmentation. We proposed a method to optimize a tree based both on color distributions and shape priors [15]. It consists in pruning and regrafting tree branches in order to minimize the energy of the best segmentation that can be extracted from the tree. Theoretical guarantees help reduce the search space and make the optimization efficient. Our experiments (see Figure 4) show that we succeed in incorporating shape information to restructure a tree, which in turn enables to extract from it good quality multi-object segmentations with shape priors. Published in the proceedings of BMVC (British Machine Vision Conference).

7.2. Reconstruction

7.2.1. *LOD Generation for Urban Scenes*

Participants: Florent Lafarge, Pierre Alliez.

We contributed a novel approach that reconstructs 3D urban scenes in the form of levels of detail (LODs). Starting from raw data sets such as surface meshes generated by multi-view stereo systems, our algorithm proceeds in three main steps: classification, abstraction and reconstruction (Figure 5). From geometric attributes and a set of semantic rules combined with a Markov random field, we classify the scene into four meaningful classes. The abstraction step detects and regularizes planar structures on buildings, fits icons on trees, roofs and facades, and performs filtering and simplification for LOD generation. The abstracted data are then provided as input to the reconstruction step which generates watertight buildings through a min-cut formulation on a set of 3D arrangements. Our experiments on complex buildings and large scale urban scenes show that our approach generates meaningful LODs while being robust and scalable. By combining semantic segmentation and abstraction it also outperforms general mesh approximation approaches at preserving urban structures [10]. Published in the ACM Transactions on Graphics journal.

7.2.2. *A Surface Reconstruction Method for In-Detail Underwater 3D Optical Mapping*

Participant: Pierre Alliez.

In collaboration with Ricard Campos and Rafael Garcia from the Computer Vision and Robotics Group from University of Girona, and Mariette Yvinec from the GEOMETRICA Inria project-team.

(a)	Bottle	Chair	Couch	Lamp	Mug	Shelf	Table	Vase
Bottle	2	0	0	0	1	0	0	2
Chair	0	4	1	0	0	0	0	0
Couch	0	0	5	0	0	0	0	0
Lamp	0	1	0	3	1	0	0	0
Mug	0	0	0	0	4	0	0	1
Shelf	0	0	0	0	0	5	0	0
Table	0	0	0	0	0	0	5	0
Vase	0	0	0	0	0	0	0	5

(b)	Bottle	Chair	Couch	Lamp	Mug	Shelf	Table	Vase
Bottle	2	0	0	0	1	0	0	
Chair	1	4	0	0	0	0	0	0
Couch	0	1	3	0	0	0	1	0
Lamp	0	0	0	4	1	0	0	0
Mug	0	0	0	0	4	0	0	1
Shelf	0	0	0	0	0	5	0	0
Table	0	0	0	0	0	0	5	0
Vase	1	0	0	0	0	0	0	4

(c)	Bottle	Chair	Couch	Lamp	Mug	Shelf	Table	Vase
Bottle	4	0	0	0	0	0	0	1
Chair	0	2	1	1	0	1	0	0
Couch	0	1	4	0	0	0	1	0
Lamp	0	0	1	3	0	0	0	0
Mug	2	0	0	0	1	0	0	2
Shelf	0	0	0	0	0	5	0	0
Table	0	0	0	0	0	0	5	0
Vase	0	0	0	1	0	0	0	4

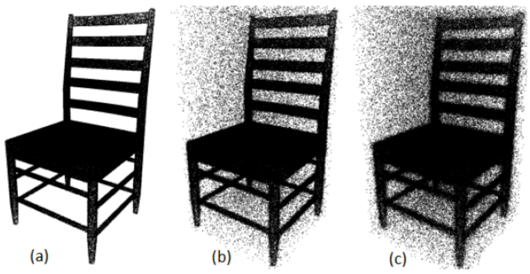
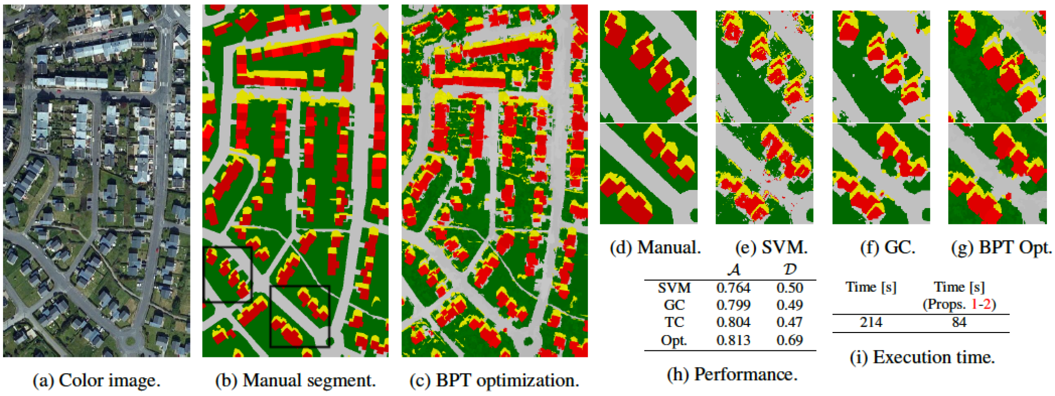


Figure 3. Classification. Left: We used four tabletop object classes from the Princeton Shape Benchmark: Bottle, Lamp, Mug and Vase. We also select four furniture object classes common to indoor scenes: Chair, Couch, Shelf and Table. Right: We evaluate our approach through computing a confusion matrix, for an increasing amount of noise and outliers. (a): Without noise and outliers. The precision of the class prediction is $82,5\%$. The classifier is not reliable for classifying the bottles, which get mislabeled as vases. (b): Added 10% outliers and 0.5% noise. Compared to the noise-free version the precision slightly dropped to 77.5%. (c): Added 20% outliers and 1% noise. The method maintains a precision of 70% for this level of noise.



	$\mathcal{A}$	$\mathcal{D}$
SVM	0.764	0.50
GC	0.799	0.49
TC	0.804	0.47
Opt.	0.813	0.69

(h) Performance.

Time [s]	Time [s] (Props. 1-2)
214	84

(i) Execution time.

Figure 4. Classification results for the satellite image over Brest. $\mathcal{A}$ denotes overall classification accuracy, and $\mathcal{D}$ denotes average buildings overlap. The performance of the proposed binary partition tree (BPT) optimization method is compared with the following methods: 1) support vector machines (SVM) classification; 2) graph cut (GC) with α-expansion; 3) cut on the BPT, regularized by the number of regions without using shape priors (TC).

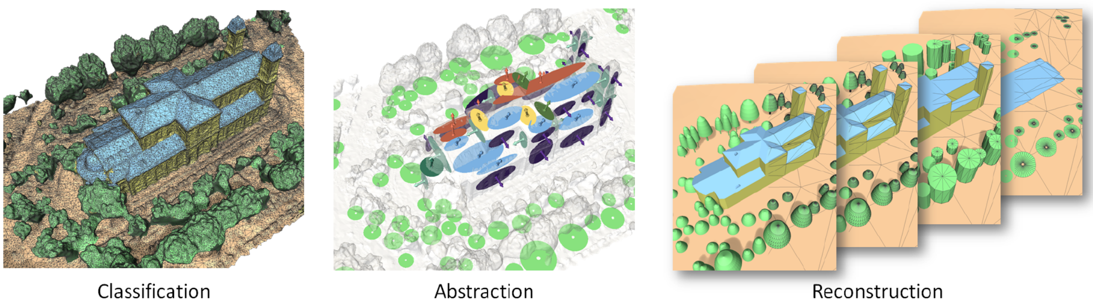


Figure 5. LOD Generation for Urban Scenes. Main steps of our algorithm.

Underwater range scanning techniques are starting to gain interest in underwater exploration, providing new tools to represent the seafloor. These scans (often) acquired by underwater robots usually result in an unstructured point cloud, but given the common downward-looking or forward-looking configuration of these sensors with respect to the scene, the problem of recovering a piecewise linear approximation representing the scene is normally solved by approximating these 3D points using a heightmap (2.5D). Nevertheless, this representation is not able to correctly represent complex structures, especially those presenting arbitrary concavities normally exhibited in underwater objects. We present a method devoted to full 3D surface reconstruction that does not assume any specific sensor configuration. The method presented is robust to common defects in raw scanned data such as outliers and noise often present in extreme environments such as underwater, both for sonar and optical surveys (Figure 6). Moreover, the proposed method does not need a manual preprocessing step. It is also generic as it does not need any information other than the points themselves to work. This property leads to its wide application to any kind of range scanning technologies and we demonstrate its versatility by using it on synthetic data, controlled laser-scans, and multibeam sonar surveys. Finally, and given the unbeatable level of detail that optical methods can provide, we analyze the application of this method on optical datasets related to biology, geology and archeology [4]. Published in the International Journal of Robotics Research.

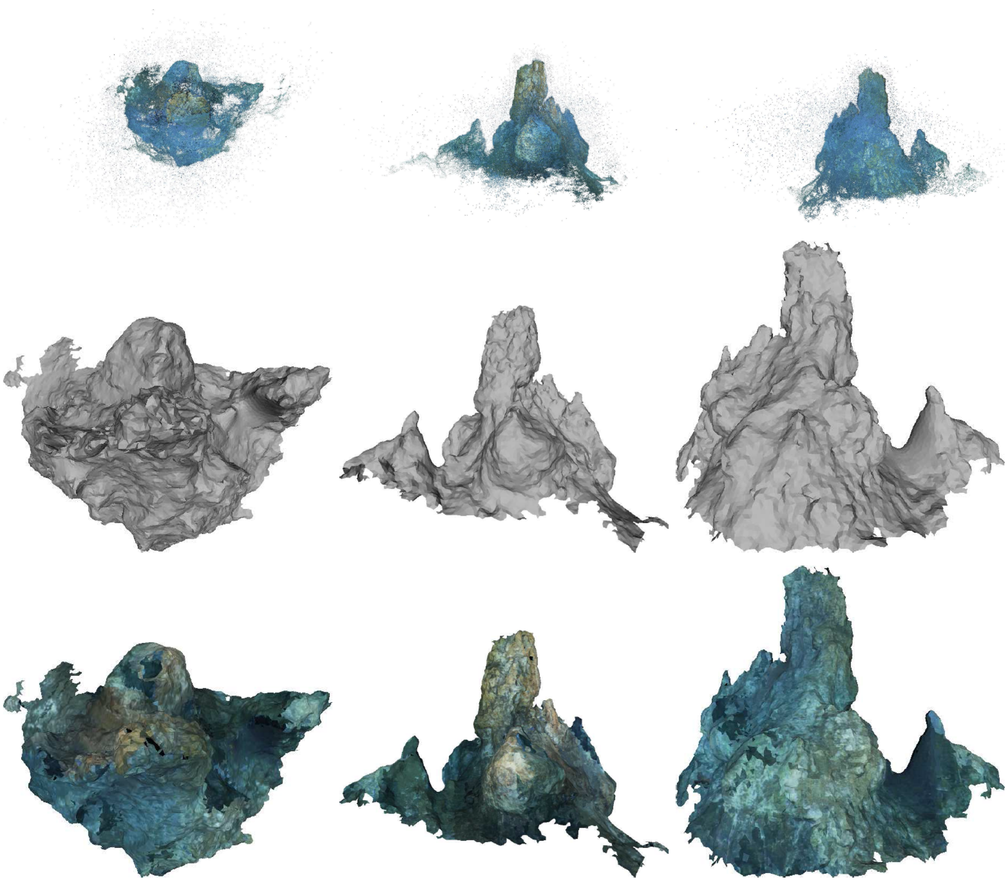


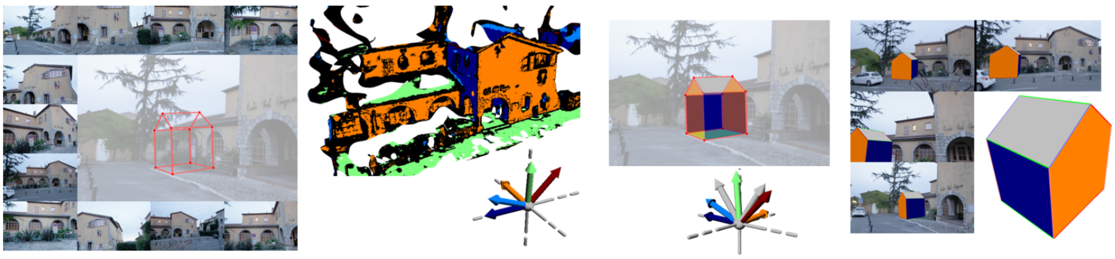
Figure 6. Underwater reconstruction.

7.2.3. Line Drawing Interpretation in a Multi-View Context

Participants: Jean-Dominique Favreau, Florent Lafarge.

In collaboration with Adrien Bousseau from the Inria project-team GraphDeco.

Many design tasks involve the creation of new objects in the context of an existing scene. Existing work in computer vision only provides partial support for such tasks. On the one hand, multi-view stereo algorithms allow the reconstruction of real-world scenes, while on the other hand algorithms for line-drawing interpretation do not take context into account. Our work combines the strength of these two domains to interpret line drawings of imaginary objects drawn over photographs of an existing scene. The main challenge we face is to identify the existing 3D structure that correlates with the line drawing while also allowing the creation of new structure that is not present in the real world. We propose a labeling algorithm to tackle this problem, where some of the labels capture dominant orientations of the real scene while a free label allows the discovery of new orientations in the imaginary scene (Figure 7). We illustrate our algorithm by interpreting line drawings for urban planing, home remodeling, furniture design and cultural heritage [18]. Published in the proceedings of CVPR (IEEE conference on Computer Vision and Pattern Recognition).



(a) Input (b) Dominant 3D orientations (c) Orientation labeling (d) Output

Figure 7. Line Drawing Interpretation. (a) Our algorithm takes as input multiple images of a scene along with a line-drawing traced over one of these images. (b) We first detect the dominant orientations of the existing scene from its multi-view stereo reconstruction. (c) Our labeling algorithm estimates the orientation of each facet of the drawing, favoring orientations already present in the scene. We visualize each dominant orientation with a random color, gray denotes new orientations. (d) We finally solve for the 3D model corresponding to the estimated orientations.

7.2.4. Marked point process model for curvilinear structures extraction

Participant: Yuliya Tarabalka [contact].

In collaboration with Seong-Gyun Jeong and Dr. Josiane Zerubia (AYIN team, Inria-SAM).

In this work, we proposed a new marked point process (MPP) model and the associated optimization technique to extract curvilinear structures [12]. Given an image, we compute the intensity variance and rotated gradient magnitude along the line segment. We constrain high level shape priors of the line segments to obtain smoothly connected line configuration. The optimization technique consists of two steps to reduce the significance of the parameter selection in our MPP model. We employ a Monte Carlo sampler with delayed rejection to collect line hypotheses over different parameter spaces. Then, we maximize the consensus among line detection results to reconstruct the most plausible curvilinear structures without parameter estimation process. Experimental results (see Figure 8) show that the algorithm effectively localizes curvilinear structures on a wide range of datasets.

7.2.5. Inference of curvilinear structure based on learning a ranking function and graph theory

Participant: Yuliya Tarabalka [contact].

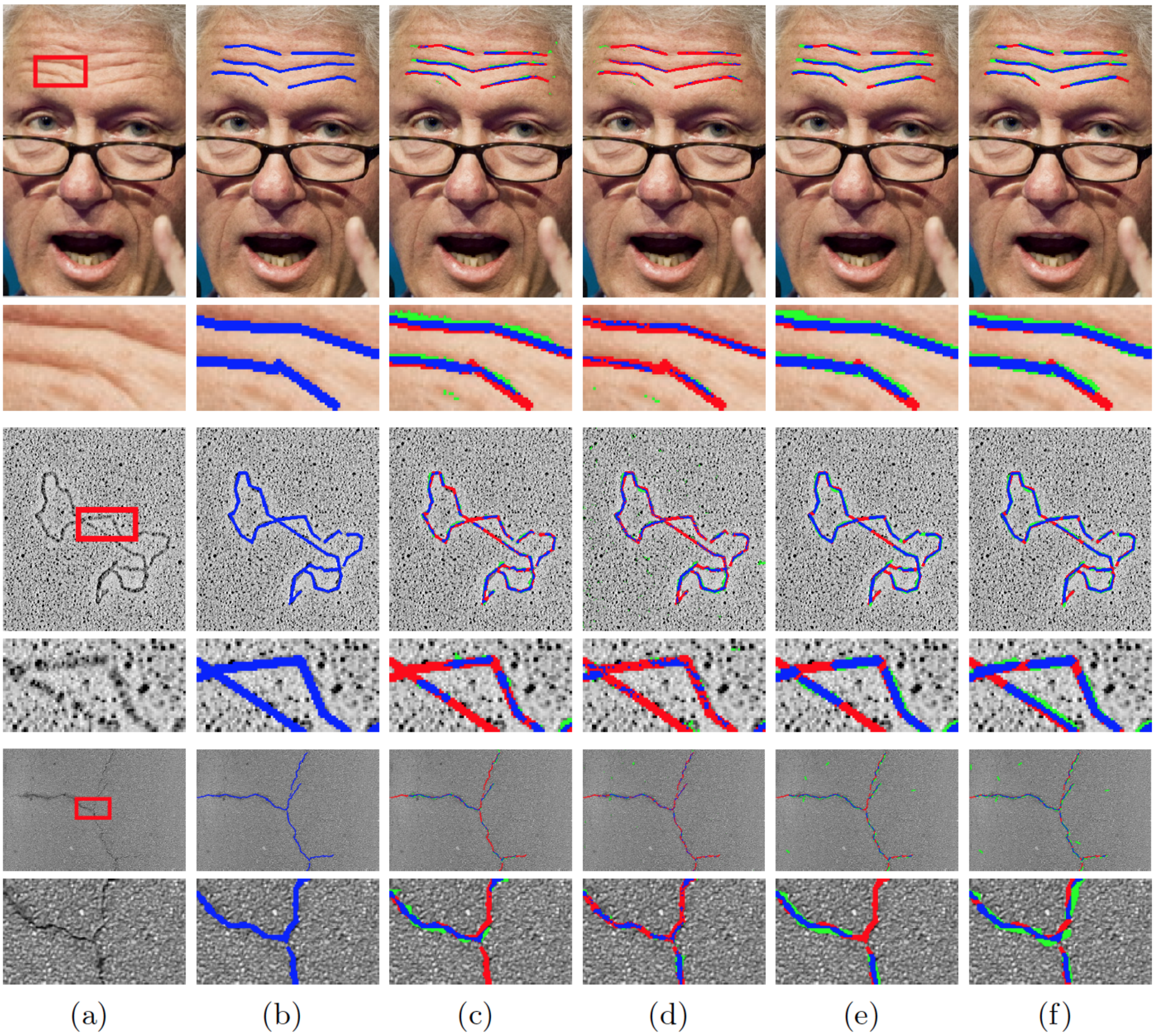
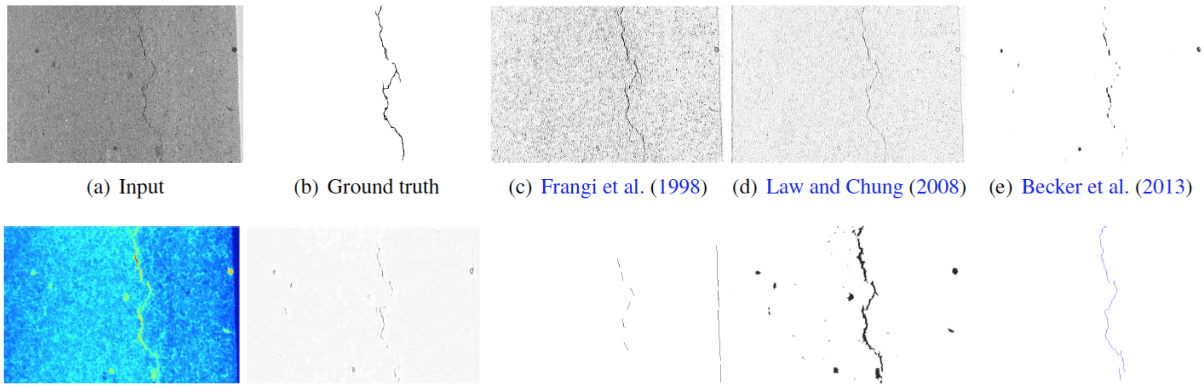


Figure 8. We visualize the localization of the curvilinear structures on input images (a). We compare with the results of a manually labeled image by a human expert (b), morphological filtering [Talbot 2007] (c), supervised feature learning [Becker 2013] (d), baseline MPP (e), and the proposed algorithm (f). Threshold values of (c) and (d) are chosen to achieve the closest recall scores to the proposed method. We use blue pixels to indicate areas which are completely corresponding to (b). Green and red pixels denote over-detected and under-detected areas, respectively, as compared with ground-truth. The name of the test images is from top to bottom: WRINKLE, DNA, *and* CRACK.

In collaboration with Seong-Gyun Jeong and Dr. Josiane Zerubia from the AYIN team and Dr. Nicolas Nisse from the COATI project-team.

Curvilinear structure extraction, inference of structured data, ranking learning, graphical model, shape simplification.

To detect curvilinear structures in natural images, we proposed a novel ranking learning system and an abstract curvilinear shape inference algorithm based on graph theory. We analyze the curvilinear structures as a set of small line segments. In this work, the rankings of the line segments are exploited to systematize the topological feature of the curvilinear structures. A Structured Support Vector Machine is employed to learn the ranking function that predicts the correspondence of the given line segments and the latent curvilinear structures. We first extract curvilinear features using morphological profiles and steerable filter responses. Also, we propose an orientation-aware feature descriptor and a feature grouping operator to improve the structural integrity during the learning process. To infer the curvilinear structure, we build a graph based on the output rankings of the line segments. We progressively reconstruct the curvilinear structure by looking for paths between remote vertices in the graph. Experimental results (see Figure 9 for an example of the experimental results' comparison on the CRACK dataset) show that the proposed algorithm faithfully detects the curvilinear structures within various datasets.



(a) Input (b) Ground truth (c) Frangi et al. (1998) (d) Law and Chung (2008) (e) Becker et al. (2013)

(f) Output rankings (g) Sironi et al. (2014) (h) Jeong et al. (2015) (i) Dissimilarity scores (j) Structure inference

Figure 9. Inference of curvilinear structure on the CRACK dataset. Our approach is depicted by (j).

7.3. Approximation

7.3.1. Isotopic approximation within a tolerance volume

Participants: Manish Mandad, Pierre Alliez.

In collaboration with David Cohen-Steiner from the GEOMETRICA project-team.

We introduce an algorithm that generates a surface triangle mesh given an input tolerance volume. The mesh is guaranteed to be within the tolerance, intersection free and topologically correct. A pliant meshing algorithm is used to capture the topology and discover the anisotropy in the input tolerance volume in order to generate a concise output. We first refine a 3D Delaunay triangulation over the tolerance volume while maintaining a piecewise-linear function on this triangulation, until an isosurface of this function matches the topology sought after. We then embed the isosurface into the 3D triangulation via mutual tessellation, and simplify it while preserving the topology. Our approach extends to surfaces with boundaries and to non-manifold surfaces. We demonstrate the versatility of our approach on a variety of data sets and tolerance volumes [7]. Figure 10 illustrates the robustness of our approach on defect-laden inputs.

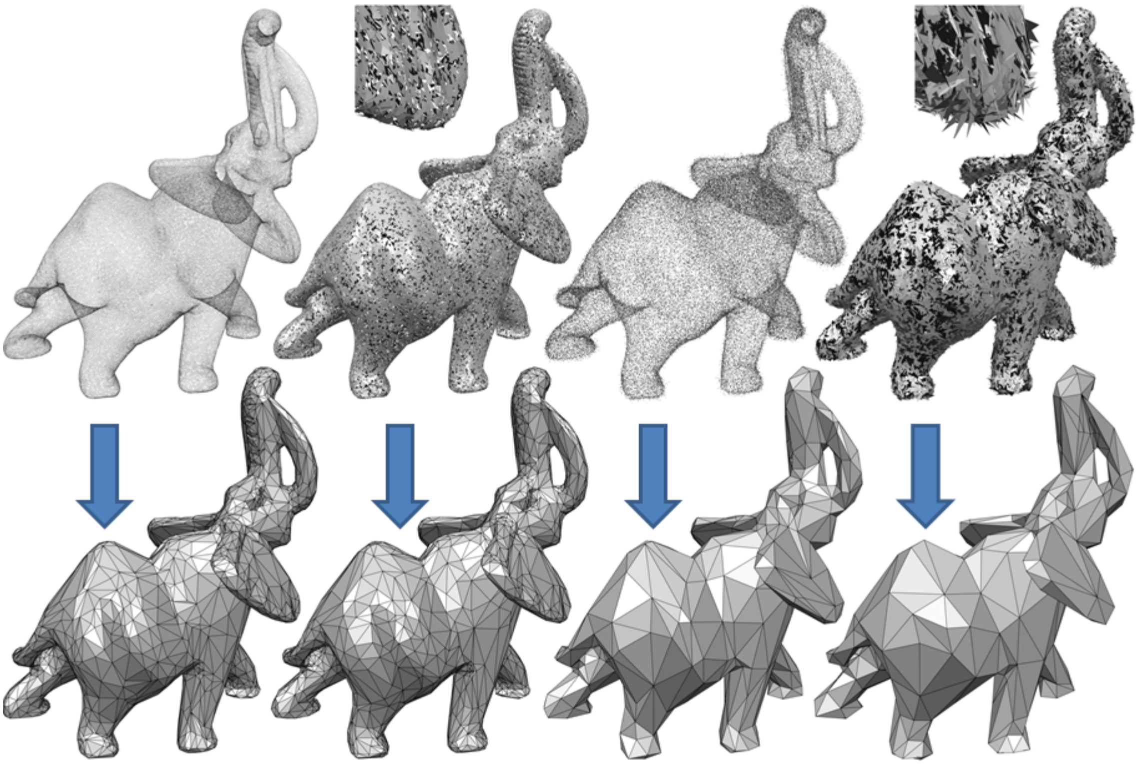


Figure 10. Isotopic approximation within a tolerance volume.

7.3.2. Structure-Aware Mesh Decimation

Participants: David Salinas, Florent Lafarge, Pierre Alliez.

We contributed to a novel approach for the decimation of triangle surface meshes. Our algorithm takes as input a triangle surface mesh and a set of planar proxies detected in a pre-processing analysis step, and structured via an adjacency graph. It then performs greedy mesh decimation through a series of edge collapse operators, designed to approximate the local mesh geometry as well as the geometry and structure of proxies (Figure 11). Such structure-preserving approach is well suited to planar abstraction, i.e., extreme decimation approximating well the planar parts while filtering out the others. Our experiments on a variety of inputs illustrate the potential of our approach in terms of improved accuracy and preservation of structure [9].

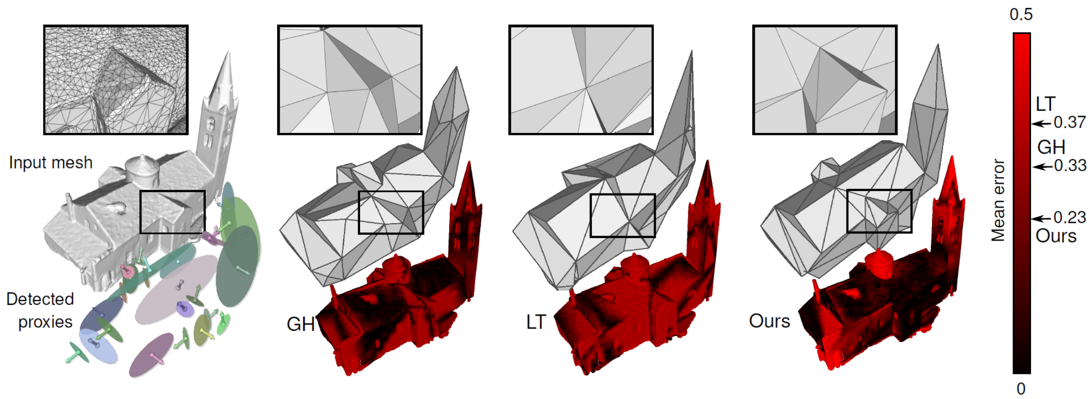


Figure 11. Structure-aware mesh decimation. Our algorithm simplifies dense triangle surface meshes via a structured set of planar proxies (left) that guides the decimation process while preserving the structure. At coarse complexity (here 50 vertices), common mesh decimation approaches (middle, quadric error metrics from Garland-Heckbert, and volume-preserving from Lindstrom-Turk) fail to reach low approximation error (see colored meshes) while preserving structure (see closeups).

7.3.3. CGALmesh: a Generic Framework for Delaunay Mesh Generation

Participants: Clément Jamin, Pierre Alliez.

In collaboration with Mariette Yvinec and Jean-Daniel Boissonnat from the GEOMETRICA project-team.

CGALmesh is the mesh generation software package of the Computational Geometry Algorithm Library (CGAL). It generates isotropic simplicial meshes – surface triangular meshes or volume tetrahedral meshes – from input surfaces, 3D domains as well as 3D multi-domains, with or without sharp features (see Figure 12). The underlying meshing algorithm relies on restricted Delaunay triangulations to approximate domains and surfaces, and on Delaunay refinement to ensure both approximation accuracy and mesh quality. CGALmesh provides guarantees on approximation quality as well as on the size and shape of the mesh elements. It provides four optional mesh optimization algorithms to further improve the mesh quality. A distinctive property of CGALmesh is its high flexibility with respect to the input domain representation. Such a flexibility is achieved through a careful software design, gathering into a single abstract concept, denoted by the oracle, all required interface features between the meshing engine and the input domain. We already provide oracles for domains defined by polyhedral and implicit surfaces [6].

7.4. Watermarking

7.4.1. Anti-Cropping Blind Resynchronization for 3D Watermarking

Participants: Xavier Rolland-Nevière, Pierre Alliez, Gwenaël Doërr.

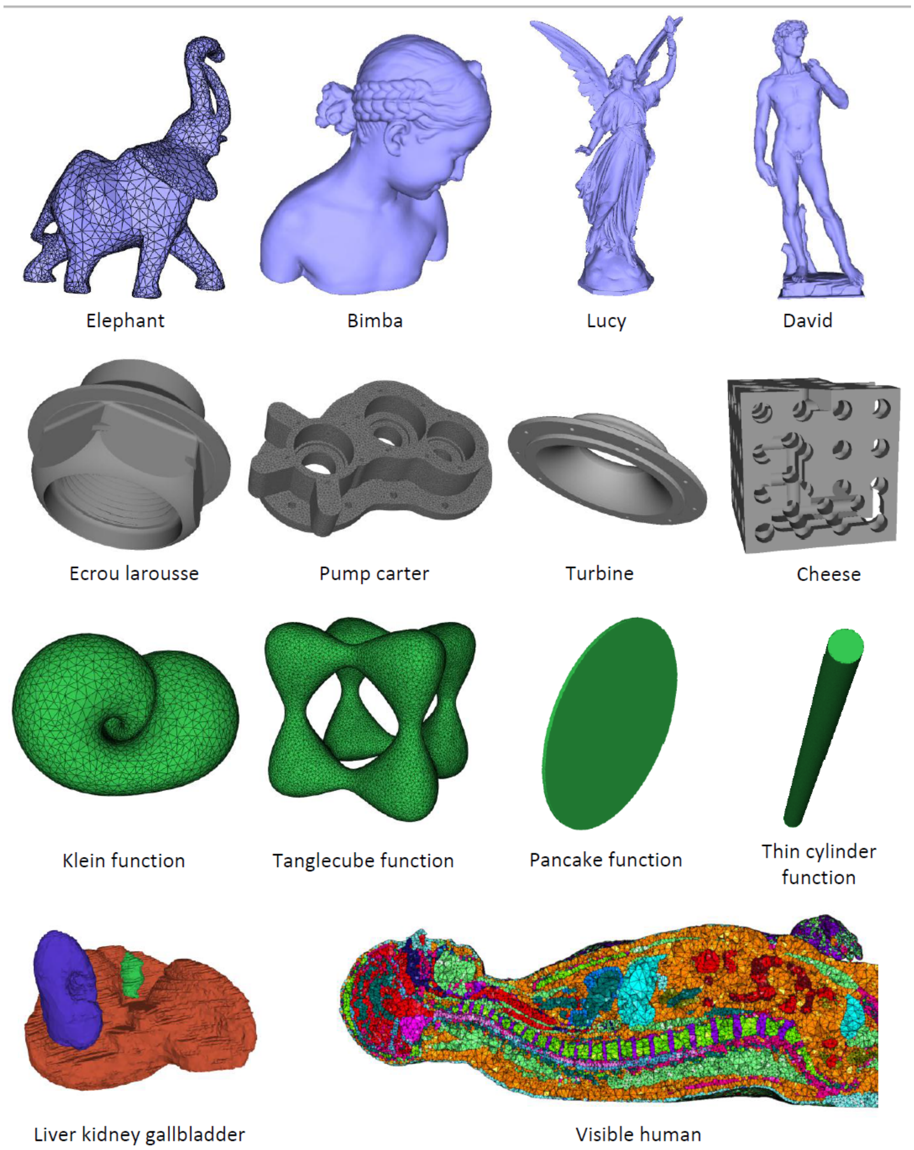


Figure 12. Input domains: domains bounded by smooth surfaces (top row, in blue), CAD models with sharp features (second row, in grey), implicit functions (third row, in green), 3D images (bottom row, multicolored).

Radial-based 3D watermarking alters the distances between the center of mass of the 3D mesh and its vertices. These watermarking systems are inherently sensitive to cropping. To address this limitation, this paper introduces a complementary blind resynchronization module to transmit critical synchronization information to the watermark decoder. Spherical patterns formed by several secret landmark vertices are embedded alongside the payload and blindly retrieved by the decoder, thereby conveying the synchronization information needed. Experimental results showcase significant improvement against cropping, while preserving performances against valumetric attacks thanks to a control parameter that automatically switches between alternate resynchronization modes [16].

ALPAGE Project-Team

7. New Results

7.1. Playing with DyALog-based parsers

Participants: Éric Villemonte de La Clergerie, Nicholas Parslow.

Éric de la Clergerie has continued the development of two DyALog-based parsers, namely DYALOG-SR, a transition-based dependency parser, and FRMG, a wide-coverage French TAG based on an underlying metagrammar.

The coverage of FRMG has been extended to cover more (rare) syntactic phenomena. A new conversion scheme has been added for the French version of the Universal Dependency Scheme. Preliminary evaluation experiments have been conducted on the French UD corpus, with both FRMG and DYALOG-SR. FRMG has also been evaluated on the French SPMRL corpus, alone and with coupling with DYALOG-sr

A new notion of secondary edges has been investigated in FRMG metagrammar and parser to provide additional dependency edges, helpful for understanding parsing outputs. In particular, secondary edges are used to denote controls between a verb and its hidden subject.

FRMG's disambiguation tuning is learned from CONLL-like treebanks using supervised learning method. We have conducted preliminary experiments to use unsupervised learning methods with observed accuracy gains between 1 to 1.5 points w.r.t. the no tuning case. However, trying to mix supervised and unsupervised methods have shown no significant gain w.r.t. the supervised case.

The hybridation of FRMG and DYALOG-SR have been tried on a larger spectrum of treebanks.

FRMG has also been exploited during the Master internship of Nicholas Parslow about the use of NLP tools to provide feedback information and correlations on essays written by non-native French learners. In particular, the correction mechanism of FRMG has been extended to cover more cases of frequent errors and provide more explicit messages.

7.2. Linear-time discriminant syntactico-semantic parsing

Participants: Benoit Crabbé, Maximin Coavoux, Rachel Bawden.

In this module we study efficient and accurate models of statistical phrase structure parsing. We focus on linear time lexicalized parsing algorithms (shift reduce) with approximations entailing linear time processing. The existing prototype involves a global discriminant parsing model of the large margin family (Perceptron,Mira, SVM) able to parse user defined structured input tokens [62]. Thus the model can take into account various sources of information for taking decisions such as word form, part of speech, morphology or semantic classes inter alia.

Our model has been generalized in a multilingual setting where we are among the state of the art systems and state of the art on some languages [23]. To our knowledge the parser is one of the fastest existing multilingual phrase structure parser. In order to ease model design for multilingual settings, we currently study efficient feature selection procedures for automating model adaptation to new languages.

We have also extended our model to continuous representations by means of deep learning methods. We currently have a neural network based decision procedure for parsing [22]. It involves both greedy search and beam based search techniques. Current work focuses on the design of dynamic oracles for improving greedy search procedures. This framework is currently tested in the multilingual setting too.

Further work involves to tackle the knowledge acquisition bottleneck problem by integrating either symbolic knowledge such as dictionaries or semi-supervised procedures for improving the formal representation of lexical dependencies in order to leverage data sparsity and estimation issues recurrent in lexicalized parsing.

7.3. French Deep Syntactic Dependency Parsing

Participants: Corentin Ribeyre, Djamé Seddah, Éric Villemonte de La Clergerie, Marie Candito.

At Alpage, we used two distinct but complementary approaches to parse and produce deep syntactic dependency graphs from the DeepSequoia and the DeepFTB (crossref here). The first one was developped by using OGRE [87], [86], a graph rewriting system (crossref here). We developed a set of rewriting rules to transform surfacic syntactic dependency trees into deep syntactic dependency graphs, then we applied this set of rules on previously parsed surfacic trees. Those trees were produced using up to three different surfacic syntactic parsers: FRMG [109], DyALog-SR [109] and Mate [47]. The results were convincing and on par with what we got on English.

The second approach was based on the work made last year regarding the English broad-coverage semantic dependency parsing. We reused our two graph parsers (the first one is based on a previous work on DAG parsing [89] and the second one on the FRMG surfacic syntactic parser [109]) to parse the same graphs. As we previously have shown on English, the use of a mix of syntactic features (tree fragments from a constituent syntactic parser [80], dependencies from a syntactic parser [47], elementary spinal trees using a spine grammar [102], etc.) improve our results. Our intuition is that syntax and semantic are not independent of each other and using syntax could improve semantic parsing. Finally, we extended a dual-decomposition third-order graph parser [76] to incorporate our syntactic feature set and we were able to reach the best performances to this day on the task for both English [28] and French (Ribeyre et al, to appear).

7.4. Towards a French FrameNet

Participants: Marie Candito, Marianne Djemaa, Benoît Sagot.

The ASFALDA project [0] is an ANR project coordinated by Marie Candito. 5 partners collaborate on the project, on top of Alpage : the Laboratoire d'Informatique Fondamentale de Marseille(LIF), the Laboratoire de Linguistique Formelle (LLF), the MELODI team (IRIT - Toulouse) and the CEA-List. The project started in October 2012, and will end in march 2016. Its objective is to build semantic resources (generalizations over predicates and over the semantic arguments of predicates) and a corresponding semantic analyzer for French. We chose to build on the work resulting from the FrameNet project [45], [0] which provides a structured set of prototypical situations, called *frames*, along with a semantic characterization of the participants of these situations (called *frame elements*). The resulting resources will consist of :

1. a French lexicon in which lexical units are associated to FrameNet frames,
2. a semantic annotation layer added on top of existing syntactic French treebanks
3. and a frame-based semantic analyzer, focused on joint models for syntactic and semantic analysis.

In 2015, we continued the corpus annotation phase, which started in 2014. We currently have about 90 frames and 790 lexical units with at least one annotated occurrence, totalizing about $11,000$ annotated occurrences. We also set up :

- procedures for checking the coherence of the annotations
- a procedure for extracting the "annotated lexicon", namely extract quantitative information about the annotated lexical units, and syntax/semantics interface information (in terms of the probabilistic distributions of the syntactic paths used to express a given semantic role)
- the graphical vizualization of the annotated corpus

We also just started a collaboration with the LIF laboratory for using deep syntactic representations for predicting semantic frames and roles.

7.5. Development of Verb∋net

Participants: Laurence Danlos, Quentin Pradet, Lucie Barque.

[0] https://sites.google.com/site/anrasfalda/
[0] https://framenet.icsi.berkeley.edu/

VerbNet is an English lexical resources for verbs, which is internationally known and widely used in numerous NLP applications [74]. Verb∋net is a French adaptation of this resource. It is semi-automatically developed thanks to the use of two French existing resources created in the 70's: LG, Lexique-Grammaire developed at LADL under the supervision of Maurice Gross, and LVF, Lexique des verbs du français by Dubois and Dubois-Charlier. The idea is to map English classes, which gather verbs with a common syntactic and semantic behavior, into classes of LG and LVF, then to manually adapt the syntactic frames according to French grammar while keeping the thematic roles and the semantic information, [84], [68] [14]. A first version of this work has been achieved in June 2015 in collaboration with Takuya Nakamura (Institut Gaspard Monge) [33].

The next step was to verify the coherence of the resource. A particular focus has been to check the way alternations have been encoded and to document this encoding. A journal article extracted from this documentation has been submitted to the *TAL* journal and Verb∋net will be released after getting the feedback of the editorial board.

7.6. Development of the French Discourse TreeBank (FDTB)

Participants: Laurence Danlos, Margot Colinet, Jacques Steinlin, Pierre Magistry.

FDTB1 is the first step towards the creation of the French Discourse Tree Bank (FDTB) with a discourse layer on top of the syntactic one which is available in the French Tree Bank (FTB). In this first step, we have identified all the words or phrases in the corpus that are used as "discourse connectives". The methodology was the following: first, we highlighted all the items in the corpus that are recorded in LexConn [88], a lexicon of French connectives with 350 items, next we eliminated some of these items with the following criteria:

1. first, we filtered out the LexConn items that are annotated in FTB with parts of speech incompatible with a connective use, e.g. *bref* annotated as *Adj* instead of *Adv*, *en fait* annotated as *Pro V* instead of (compound) *Adv*;
2. second, as we lay down for theoretical and pratical reasons that elementary arguments of connectives must be clauses or VPs, we filtered out e.g. LexConn prepositions that introduce NPs;
3. last, we filtered out LexConn prepositions and adverbials with a non-discursive function.

The last criterion requires a manual work contrarily to the two others. For example the preposition *pour (to)*, is ambiguous between a connective use *(Fred s'est dépeché pour être à la gare à 17h (Fred hurried to be at the station at 17h))* and a preposition introducing a complement *(Fred s'est dépeché pour aller à la gare (Fred hurried to go to the station))*, and the disambiguation between the two uses is subtle and so the topic of a long paper [58], whose results have been used to enhance Le*fff* [93].

FDTB1 identifies 9 833 discourse connectives (among 18 535 sentences). This ressource is freely available and has been released in May 2015 [36].

FDTB2 is the next step in the creation of the FDTB. It consists in annotating the arguments of the discourse connectives identified in FDTB1 as well as the senses of these connectives (senses expressed through a set of discourse relations). This resource is still worked on.

7.7. Discourse Parsing

Participants: Chloé Braud, Laurence Danlos.

Discourse parsing goal is to reflect the rhetorical structure of a document, how pieces of text are linked in order to form a coherent document. Understanding such links could benefits to several other natural language applications (summarization, language generation, information extraction...).

A discourse parser corresponds to two major subtasks: a segmentation step wherein discourse units (DUs) are extracted, and a parsing step wherein these DUs are (recursively) related through "discourse (rhetorical) relations". The most difficult task in discourse parsing is the labeling of the relations between DUs, especially when no so-called connective overtly marks the relation (we then talk about implicit relations as opposed to explicit ones).

In her PhD, defended in December 2015, Chloé Braud develops a discourse relation classifier, carrying experiments on French and English. Focusing on the problem on implicit relation identification, this work explores ways of using raw data in combination with the available manually annotated data: this work led to systems based on domain adaptation methods exploiting automatically annotated explicit relations – demonstrating improvements on the French corpus Annodis and on the English corpus PDTB –, and to systems using word embeddings built from raw text to efficiently transform a word based representation of the data – leading to state-of-the art performance or above on the English corpus PDTB without the need of hand-crafted resources [21].

7.8. Towards a morpho-semantic resource for French designed for Word Sense Disambiguation

Participant: Lucie Barque.

The most promising WSD methods are those relying on external knowledge resources [78] but semantic resources for French are scarse. Moreover, existing resources offer fine grained sense dictinctions that do not fit to WSD. Our aim is to provide the NLP community with a broad-coverage morpho-semantic lexicon for French that relies on coarse-grained sense distinctions for polysemic units. Preliminary results concern nouns, on which we have first focused because their semantic description, compared to verbs, crucially lacks (for information retrieval, for instance) and because the regular polysemy phenomenon (recurring cases of polysemy within semantic classes) mainly occurs in nominal semantic classes:

- We proposed a linguistically motivated description of general semantic labels for nouns, that will allow for coarse-grained sense distinctions [107]
- Regular polysemy of nouns that can denote an event or a participant of this event has also been described for a large number of French nouns in [46]
- From a morphological point of view, nouns denoting events in French are mostly deverbal nouns (eg. *conversation* 'conversation', *promenade* 'stroll'), but there are also underived event nouns (eg. *guerre* 'war', *séisme* 'earthquake'). We compared their semantic properties in [35].
- Some lexical meanings are not easily captured by ontological semantic classes and a closer look has to be taken at them. Relational meanings in relational nouns are one of them [15].

7.9. Development of the Corpus de Référence du Français

Participants: Stéphane Riou, Benoît Sagot.

The 'Initiative Corpus de Référence du Français' (ICRF) is a project of Institut de Linguistique Française (ILF-FR2393 CNRS), coordinated by its director Franck Neveu and by Benoît Sagot.

The purpose of the ICRF is the development of a first prototype of the future French Reference Corpus, so as to assess the feasibility of this project and evaluate its potential impact. ICRF reuses existing freely-available French corpora, supplemented by additional data in an opportunistic fashion (e.g. a French media critic corpus and the corpus of talks given at an workshop on ethics and neurodegenerative diseases). ICRF preserves copyright and authorship of all corpora used. These corpora have been or will be part-of-speech tagged with MElt, converted to XML (TEI-P5-compliant) and made accessible via a web interface. The aim of ICRF is not to replace individual corpora and the interface will therefore allow, whenever possible, to easily recover access to each individual corpus. ICRF adds 5 metadata tags to categorize each individual corpus: spoken/written, text type and genre, linguistic competence level, date and linguistic area.

In 2015, the normalisation, tagging and conversion to XML of individual corpora has started, following the design of format specifications. The development of the web interface has already started, and a prototype is now available. Users can perform queries (search by tokens and/or POS) and use basic linguistic tools on the corpora (e.g. a concordancer). It is therefore more than a simple search interface or a download site: it improves research and selection of corpus.

7.10. Word order variation in Old French

Participants: Benoit Crabbé, Alexandra Simonenko.

As participant of the strand *Experimental Grammar* of the Labex EFL project *Empirical Foundations of Linguistics*[0] we study word order issues on Old French and more specifically the relative ordering of complements of ditransitive verbs. The inquiry seeks to identify several factors influencing the ordering of Old French complementation in different texts (varying in dates and genres) by carrying quantitative and statistical work from annotated Old French data. [0]

The first quantitative results [29] will be compared with what is known from corpus studies on the relative ordering of subject and complement in Old French [75]. It will also be compared to the quantitative results obtained on the relative ordering of complements of ditransitive verbs in Modern French [8] and modern English [53]. This comparative perspective is expected to provide new insights on French language evolution.

7.11. Cross linguistic factors governing word order

Participant: Benoit Crabbé.

In many languages, flexible word order often has a pragmatic role and marks the introduction of new information, a focus or a topic shift. Other cases of language- internal word order variation are alternations between two options such as *Mary gave John a book* and *Mary gave a book to John*, which are conditioned on syntactic and semantic factors such as the complexity of the constituents (as in *Mary gave John a book she had read ten times*), their animacy or the meaning of the verb [52].

One of the goals of this module is to investigate the connection between the quantitative aspects of word order variation across languages and the quantitative aspects of word order variation within a language. We study the corresponding patterns in language-internal variation by looking at the syntactically annotated corpora of various languages. Focusing on the variation of the internal word order of the noun-phrase as a case study [25], we explore, in collaboration with Kristina Gulordava (PhD at the University of Geneva, former international visitor at Alpage), to which extent a computational corpus-based analysis can provide new evidence not only for empirical, but also for theoretical linguistic research.

[0] www.labex-efl.org
[0] SRCMF corpus: http://srcmf.org/; MCVF: http://www.voies.uottawa.ca

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, Emmanuel Vincent, Denis Jouvet, Julie Busset, Benjamin Elie, Andrea Bandini, Ilef Ben Farhat, Sara Dahmani, Valérian Girard.

7.1.1. Articulatory modeling

7.1.1.1. Acoustic simulations

The acoustic simulation plays a key role in articulatory synthesis since it generates the acoustic signal from the instantaneous geometry of the vocal tract. This year we extended the single-matrix formulation to enable self-oscillation models of vocal folds, including glottal chinks, to be connected to the vocal tract. It also integrates the case of a local division of the main air path into two lateral channels, as it may occur during the production of lateral approximants. Extensions give rise to a reformulation of the acoustic conditions at the glottis, and at the upstream connection of bilateral channels. Numerical simulations validate the simulation framework. In particular the presence of a zero around 4 kHz due to the presence of bilateral channels around both sides of the tongue for the sound /l/ is confirmed by the simulations. These results agree with those obtained via independent techniques. Simulations of static vowels reveal that the behavior of the vocal folds is qualitatively similar whether they are connected to the single-matrix formulation or to the classic reflection type line analog model.

7.1.1.2. Acquisition of articulatory data

Magnetic resonance imaging (MRI) is a technique which provides very good static images of the vocal tract. However, it cannot be used directly to acquire dynamic images of the vocal tract which would enable a better comprehension of articulatory phenomena and the development of better coarticulation models. We thus have a cooperation with the IADI (Imagerie Adaptatative Diagnostique et Interventionnelle) INSERM laboratory in Nancy Hospital intended to develop cineMRI [86], [87] (see. 6.8).

7.1.1.3. Articulatory models

An articulatory model of the velum [66], [65] was developed in order to complete an articulatory model already comprising other articulators. The velum contour was delineated and extracted from a thousand of X-ray images corresponding to short sentences in French. A principal component analysis was applied in order to derive the main deformation modes. The first component corresponds to the opening and comes with a shape modification linked to the apparition of a bulb in the upper part of the velum when it rises. The area function of the oral tract is modified so as to incorporate the velum movements. This model was connected with acoustic simulations in order to synthesize sentences containing French nasal vowels and consonants.

7.1.2. Expressive acoustic-visual synthesis

During this year, we have focused on the development of the acquisition infrastructure necessary to acquire audiovisual data. Mainly, we have developed several methods that allow acquiring acoustic and visual data synchronously. The visual data can originate from the Articulograph, Vicon or Intel RealSense devices. This heterogeneity of the data needs developing techniques to merge precisely the data in one unique reference. Synchronization techniques have also been developed for this purpose. We have evaluated the precision of the acquisition of such systems [61]. The combination of more than one motion capture technique aims to use the best quality data for each part of the face: (1) EMA (articulograph) for the lips, to have high precise measurement of the shape of the mouth that is related to speech and (2) kinect-like or Vicon system for the upper part of the face, that model mainly expressions.

We have acquired a small expressive audiovisual speech corpus of two actors: based on motion capture data (Vicon) and acoustic data. The content of the corpus is composed of six basic emotions (joy, sadness, anger, surprise, disgust and fear). This corpus will be used to investigate the characterization of emotions in audiovisual speech in the visual space and in the acoustic space.

We have also developed an algorithm to animate the 3D model of human face from a limited number of markers. The animation is very efficient and provides realistic animation results [82]. The 3D face will be used with the audiovisual system.

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 investigated the schooling of a population of 166 students from primary to intermediate and secondary schools. These children and teenagers had specific language impairment: SLI (severe language impairment), dyslexia, dysorthographia. Since their childhood, they faced phonemic discrimination, phonological and phonemic analysis difficulties. We observed that they had trouble learning to read and more generally they experienced learning difficulties. Consequently, this lead them to repeat one or more grades, whereas in France, repetition is prohibited within each cycle and very limited between cycles.

7.1.3.2. Analysis of non-native pronunciations

Thanks to the detailed manual annotation of the French-German learner corpus that was carried out at the phonetic level in the IFCASL project (cf. 9.1.2), it was possible to investigate non-native pronunciation variants. The analysis revealed that German learners of French have most problems with obstruents in word-final position, whereas French learners of German show complex interferences with the vowel contrasts for length and quality [41]. Also, the correct pronunciation rate of the sounds, for several phonetic classes, was analyzed with respect to the learner's level, and compared to native pronunciations. One outcome is that different sound classes show different correct rates over the proficiency levels; and, for the German data, the frequently occurring syllabic [=n] is a prime indicator of the proficiency level.

We analyzed the realizations of French voiced fricatives by German non-native and French native speakers, in final position of an accentual group, a position where German fricatives are devoiced [27], [28]. Three speaker levels (from beginners to advanced) and different boundary types (depending on whether the fricative is followed by a pause, a schwa, or is directly followed by the first phoneme of the subsequent group) were considered. A set of cues, among which periodicity and fricative duration, have been analyzed. Results argue in favor of an influence of L1 (German) final devoicing on non-native realizations and show a strong interdependence between voicing, speakers' level, prosodic boundaries. The influence of orthography also strongly influenced voicing results.

We also investigated the realization of the short/long German contrast by French learners through three methods [60]. All these methods - phonetic annotation, perceptual experiment and acoustic analysis - used the same database (the IFCASL corpus). Depending on the method the results shed light on slightly different aspects of the same process, the interference of the French phonetic and phonological systems on the production of the German L2 vowels. Whereas the first method (phonetic annotation) revealed that especially rounded vowels are problematic in the long/short distinction, we could show with the second method (a perceptual experiment) that particularly the [o:]/[O] distinction seems to be hard to produce for French learners. The third method (an acoustical analysis) corroborated this finding and added acoustic details on duration and formants. The results of the studies can be used to create individualized training and feedback for foreign language learners, aimed at reducing their accent in L2.

7.2. Statistical Modeling of Speech

Participants: Antoine Liutkus, Emmanuel Vincent, Irina Illina, Dominique Fohr, Denis Jouvet, Joseph Di Martino, Vincent Colotte, Ken Deguernel, Amal Houidhek, Xabier Jaureguiberry, Aditya Nugraha, Luiza Orosanu, Imran Sheikh, Nathan Souviraà-Labastie, Dung Tran, Imene Zangar, Mohamed Bouallegue, Thibaut Fux, Emad Girgis, Juan Andres Morales Cordovilla, Sunit Sivasankaran, Freha Boumazza.

7.2.1. Source separation

Audio source separation is an inverse problem, which requires the user to guide the separation process using prior models for the source spectra and their spatial covariance matrices. We studied the impact of deterministic subspace constraints [14] over the spatial covariance matrices and pursued our work on the separation of multichannel mixtures guided by multiple, deformed reference signals such as repeated excerpts of the same music or repeated versions of the same sentence uttered by different speakers [17], [56]. Other models we have been working on include those based on local regularities of the spectral representations of musical sources (KAM, [52], [43], [51]). We also validated the positive impact of speech enhancement based on the FASST toolbox on speaker recognition [53].

As a new research direction, we extended the Gaussian framework for source separation to the family of α-stable stochastic processes [42]. This extension notably opens the path to new and robust parameters estimation algorithms for source separation [16], [67], that should be less prone to local minima. Current research notably comprises multichannel stable processes.

In parallel, we started yet another research track on the use of deep learning for source separation [24]. We proposed a new multichannel enhancement technique that 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 [75]. The model parameters are alternately estimated in an expectation-maximization (EM) fashion. We used this technique for music separation and speech enhancement in the context of the 2015 Signal Separation Evaluation Campaign (SiSEC) and the 3rd CHiME Speech Separation and Recognition Challenge, respectively [55]. 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 [81].

Finally, we pursued our long-lasting efforts on the evaluation of audio source separation by co-organizing the 2015 Signal Separation Evaluation Campaign (SiSEC) [69] and writing a position paper about the scaling up of dataset sizes [21].

The ANR young researcher project KAMoulox (2016-2019 - cf. 9.1.5), that has just been accepted will deal with large audio archives, and more precisely with the "Archives du CNRS — Musée de l'homme" that gather a large set of old and noisy audio recordings (cf. 4.4). The work on source separation can lead to the design of semi automatic denoising and enhancement features, that would allow these researchers to significantly enhance their investigation capabilities, even without expert knowledge in sound engineering.

7.2.2. Acoustic modeling

We explored the use of an auxiliary function technique for fast training of neural networks [58]. We did not apply this technique to deep neural network acoustic models yet.

In the framework of using speech recognition for helping communication with deaf or hard-of-hearing people, robustness of the acoustic modeling was investigated. Studies were related to improving robustness with respect to speech signal level and environment noise through multicondition training and enhanced set of acoustic features (noise robust features or standard features after spectral noise subtraction) [37].

7.2.3. Linguistic modeling

7.2.3.1. Out-of-vocabulary proper name retrieval

Recognition of proper names (PN) is a challenging task in information retrieval in large audio/video databases. Proper names are semantically rich and are usually key to understanding the information contained in a document. Within the ContNomina project (cf. 9.1.3), we focus on increasing the vocabulary coverage of a speech transcription system by automatically retrieving proper names from contemporary text documents. We proposed methods that dynamically augment the automatic speech recognition system vocabulary, using lexical and temporal features in diachronic documents (documents that evolve over the time). Our work uses temporal context modeling to capture the lexical information surrounding proper names so as to retrieve out-of-vocabulary (OOV) proper names and increase the automatic speech recognition vocabulary.

We proposed new methods to retrieve OOV PNs relevant to an audio news document by using probabilistic topic models. We addressed retrieval of rare OOV PNs, which further improves the recall. Our proposed lexical context model improves the mean average precision of OOV PN retrieval [62]. We also proposed a two step approach for recognition of OOV PNs in an audio document. The first step retrieves OOV PNs relevant to an audio document using probabilistic topic models; and the second step uses a phonetic search for the target OOV PNs using a k-differences approximate string matching algorithm [63]. In [64], we discuss two specific phenomena, word frequency bias and loss of specificity, which affect the retrieval of OOV PNs using Latent Dirichlet Allocation (LDA) topic models. We studied different entity-topic models, which are extensions of LDA designed to learn relations between words, topics and PNs. We showed that our proposed methods of rare OOV PN and lexical context re-ranking improve the recall and the mean average precision for the LDA and the entity-topic models.

For OOV retrieval, we proposed the continuous space word representation using neural networks. This continuous vector representation (word embeddings) is learned from large amounts of unstructured text data. To model semantic and lexical context of proper names, different strategies of local context modeling were proposed [34], [33]. We studied OOV PN retrieval using temporal versus topic context modeling, different word representation spaces for word-level and document-level context modeling, and combinations of retrieval results [38]. We extended the previously proposed neural networks for word embedding models: the word vector representation proposed by Mikolov is enriched by an additional non-linear transformation. This model allows to better take into account lexical and semantic word relationships [39].

7.2.3.2. Adding words in a language model

A novel approach was proposed to add some new words in an existing n-gram language model, based on a similarity measure between the new words to be added and words already present in the language model [47]. Based on a small set of sentences containing the new words and on a set of n-gram counts containing the known words (known for the current language model), we search for known words which have the most similar neighbor distribution (of the few preceding and few following neighbor words) to the new words. The similar words are determined through the computation of KL divergences on the distribution of neighbor words. The n-gram parameter values associated to the similar words are then used to define the n-gram parameter values of the new words.

7.2.3.3. Selecting data for training a language model

Large vocabulary language models for speech transcription are usually trained from large amounts of textual data collected from various sources, which are more or less related to the target task. Selecting data that matches the target task was investigated in this context [46], this leads to a small reduction of the perplexity, and a smaller size of the resulting language model.

7.2.3.4. 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 pursued our pioneering work on music language modeling, with a particular focus on the modeling of long-term structure [12]. We also assessed the applicability of our prior work on joint modeling of note and chord sequences to new corpora of improvised jazz music, with the difficulty that these corpora are very small.

7.2.4. Speech generation by statistical methods

7.2.4.1. Pathological voice transformation

With respect to pathological voice processing, a competing approach to signal processing techniques consists in recognizing the pathological voice in order to transform it in a text version that can be re-synthesized. Such an approach is currently being experimented, and preliminary results are quite encouraging [15].

7.2.4.2. HMM-based synthesis

This year, we started working on HMM-based synthesis in the framework of a CMCU PHC project with ENIT (Engineer school at Tunis-Tunisia; cf. 9.3.2.2). Two topics will be explored by two PhD students. The first topic deals with the building of an Arabic corpora along with the analysis of linguistic features which are relevant for the HMM-based synthesis of the Arabic language. The second topic deals with improving the quality of the HMM-based synthesis system. In parallel, we started applying the HTS system (HMM-based Speech Synthesis System) to the French language.

7.3. Uncertainty Estimation and Exploitation in Speech Processing

Participants: Emmanuel Vincent, Odile Mella, Dominique Fohr, Denis Jouvet, Agnès Piquard-Kipffer, Baldwin Dumortier, Luiza Orosanu, Dung Tran, Sucheta Ghosh, Antoine Chemardin, Aghilas Sini.

7.3.1. Uncertainty and acoustic modeling

7.3.1.1. Noise-robust speech recognition

In many real-world conditions, the target speech signal overlaps with noise and some distortion remains after speech enhancement. 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 two successful editions in 2011 and 2013, we organized the third edition in 2015 [25].

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 made a conceptual breakthrough by proposing to learn the estimator from data using a non-parametric estimator and discriminative training [18], [59]. With GMM-HMM acoustic models, we obtained on the order of 30% relative word error rate reduction with respect to conventional decoding (without uncertainty), that is about twice as much as the reduction achieved by the best single uncertainty estimator. We also started working on the propagation of uncertainty in deep neural network acoustic models [19] and on its use for noise-robust speaker recognition [54].

7.3.1.2. Other applications

Besides the above applications, we started exploring applications of uncertainty modeling to robot audition [23] and control of wind turbines [31]. 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. 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.

7.3.2. Uncertainty and speech recognition

In the framework of using speech recognition for helping communication with deaf or hard-of-hearing people in the FUI project RAPSODIE (cf. 9.1.7), the best way for displaying the speech transcription results has been investigated. To our knowledge there is no suitable, validated and currently available display of the output of automatic speech recognizer for hard-of-hearing persons, in terms of size, colors and choice of the written symbols. The difficulty comes from the fact that speech transcription results contain recognition errors, which may impact the understanding process. Although the speech recognition system does not know the errors it makes, through the computation of confidence measures, the speech recognizer estimates if a word or a syllable is rather correctly recognized or not; hence such information can be used to adjust the display of the transcription results. Different ways were investigated for displaying the speech recognition results which take also into account the reliability of the recognized items. In this qualitative study, 10 persons have been interviewed to find the best way of displaying the speech transcription results. All the participants are deaf with different levels of hearing loss and various modes of communication [50].

7.3.3. Uncertainty and phonetic segmentation

Within the framework of the IFCASL project (cf. 9.1.2), a speech corpus of native and non-native speech for the French-German language pair was designed and recorded. Besides beeing used for analyzing non-native phenomena (cf. 7.1.3.2), this corpus will be used for developing and assessing automatic algorithms that will provide diagnosis on the learner mispronunciations [78]. Therefore, the automatic alignments of the audio files corresponding to the French and German speakers uttering French sentences (4100 audio files) were manually checked and corrected by a group of seven French annotators (the German data were handled by the German partner). We analyzed with CoALT the inter-annotator agreement with respect to an expert annotator for boundary shifts, insertions and deletions as well as devoicing diacritic [45]. The accuracy of the phone boundaries on non-native speech were investigated with respect to the HMM acoustic models used. The best performance (smallest amount of non-native phone segments whose boundaries are shifted by more than 20 ms compared to the manual boundaries) was obtained by combining each French native HMM model with an automatically selected German native HMM model [35].

Within the ANR ORFEO project (cf. 9.1.6), we addressed the problem of the alignment of spontaneous speech. The audio files processed in the ORFEO project were recorded under various conditions with a large SNR range and contain extra speech phenomena and overlapping speech. We trained several sets of acoustic models and tested different methods to adapt them to the various audio files [36]. Moreover in the framework of the EQUIPEX ORTOLANG (cf. 9.1.1), a web application, ASTALI (cf. 6.2), was developped in order to align a speech signal with its corresponding orthographic transcription (given in simple text file for short audio signals or in .trs files as generated by transcriber for longer speech signals).

In conventional speech-text alignments, a 10 ms frame shift is usually used for the acoustic analysis which leads to a minimum duration of 30 ms for each phone segment. Such duration constraint may not fit with actual sound duration in fast speaking rate. To overcome such contraint, a 5 ms frame shift can be used. Statistics on pronunciations variants estimated on large speech corpora have shown that when the conventional 10 ms frame shift is used, the frequency of the longest pronunciation variants gets underestimated [26]. Moreover, the analysis of some pronunciation variant frequencies have shown that some final consonantal cluster completely disappear at high speaking rates [40].

7.3.4. Uncertainty and prosody

Detection of sentence modality (question vs. affirmation) has been investigated using linguistic and prosodic features. Best results are achieved when the classifier uses all the available information [48], that is both linguistic and prosodic features. A detailed analysis has also shown that small errors in the determination of the sentence boundaries are not critical [49].

Speech-text alignments have been used to extract speech segments containing words and expressions that can be used either as normal lexical words or as discourse particles (as for example *quoi*, *voilà*, ...). The prosodic features for these words and expressions were extracted and analyzed [30]; automatic identification of the word function (discourse particle or not) from these prosodic features was also investigated.

In the context of the EQUIPEX ORTOLANG (cf. 9.1.1), several algorithms for computing the fundamental frequency have been implemented in the JSnoori software. These features can be computed directly from the GUI interface or through Python scripts. Future work will focus on improving the quality and robustness of the fundamental frequency estimation, and on determining the reliability of the estimations.

PANAMA Project-Team

7. New Results

7.1. Recent results on sparse representations

Sparse approximation, high dimension, scalable algorithms, dictionary design, sample complexity

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the field of sparse representations, 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),

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). This year, 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. A journal paper is currently under revision [57].

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 [28], and a journal paper has been submitted [55].

Random sampling of bandlimited signals on graphs: 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 submitted [56].

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 show that it reaches equal performance while being faster by a factor at least two for large datasets. A conference paper has been accepted at ICASSP 2016 [43] and a long version is in preparation.

7.1.2. Algorithmic and theoretical results on dictionary learning

Participants: Rémi Gribonval, Luc Le Magoarou, Nicolas Bellot, Thomas Gautrais, Nancy Bertin, Srdan Kitic.

Main collaboration (theory for dictionary learning): Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (Inria, Paris)), Martin Kleinsteuber, Matthias Seibert (TU-Munich),

Theoretical guarantees for dictionary learning : 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 litterature to learn a dictionary from a collection of training samples, there are little existing results which provide an adequate mathematical understanding of the behaviour of these techniques and their ability to recover an ideal dictionary from which the training samples may have been generated.

Beyond our pioneering work [86], [109] [5] on this topic, which concentrated on the noiseless case for non-overcomplete dictionaries, we showed the relevance of an ℓ^1 penalized cost function for the locally stable identification of overcomplete incoherent dictionaries, in the presence of noise and outliers [19]. Moreover, we established sample complexity bounds of dictionary learning and other related matrix factorization schemes (including PCA, NMF, structured sparsity ...) [20].

Learning computationally efficient dictionaries 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 –and their fast implementation. The principle and its application to image denoising appeared at ICASSP 2015 [33] and an application to speedup linear inverse problems was published at EUSIPCO 2015 [32]. A journal paper is currently under revision [51].

We further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms – a conference paper on this latter topic has been accepted for publication in ICASSP 2016 [41]. A C++ software library is in preparation to release the resulting algorithms.

Operator learning for cosparse representations: Besides standard dictionary learning, we also considered learning in the context of the cosparse model. The overall problem is to learn a low-dimensional signal model from a collection of training samples. The mainstream approach is to learn an overcomplete dictionary to provide good approximations of the training samples using sparse synthesis coefficients. This famous sparse model has a less well known counterpart, in analysis form, called the cosparse analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete analysis operator.

This year we obtained an upper bound of the sample complexity of the learning process for analysis operators, and designed a stochastic gradient descent (SGD) method to efficiently learn analysis operators with separable structures. Numerical experiments were provided that link the sample complexity to the convergence speed of the SGD algorithm. A journal paper has been published [24].

7.1.3. An alternative framework for sparse representations: analysis sparse models

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic, Laurent Albera.

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 [101], [85], [102] successful applications of this approach to sound source localization, audio declipping and brain imaging have been developed this year.

Versatile co-sparse regularization: Digging the groove of last year results (comparison of the performance of several cosparse recovery algorithms in the context of sound source localization [94], demonstration of its efficiency in situations where usual methods fail ([96], see paragraph 7.5.2), applicability to the hard declipping problem [95], application to EEG brain imaging [60] (see paragraph 7.5.3), a journal paper embedding the latest algorithms and results in sound source localization and brain source localization in a unified fashion was published in IEEE Transactions on Signal Processing [23]. Other communications were made in conferences and workshops [50], [31] and Srdan Kitic defended his PhD thesis [12]. 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)

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 showed that the unknown speed of sound can be learned jointly in the process of cosparse recovery, under mild conditions (work presented last year at iTwist'14 workshop [66]). This year, improved and extended results were obtained: first with a new algorithm for sound source localization with unknown speed of sound [12], then by extending 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 accepted for publication in ICASSP 2016 [38], see also Section 7.3.2).

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)

7.2.1. Characterizing 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 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 [77] and several conference communications [75], [76]. The work this year concentrated on characterizing waveforms with optimum PAPR [30], [48].

7.3. Emerging activities on compressive learning and inverse problems

Compressive sensing, compressive learning, audio inpainting,

7.3.1. Audio inpainting

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 [59]), new results were obtained last year and this year to address the case of audio declipping with the competitive cosparse approach. Last year, its promising results, especially when the clipping level is low, were confirmed experimentally by the formulation and use of a new algorithm named Cosparse Iterative Hard Tresholding [95], which is a counterpart of the sparse Consistent Iterative Hard Thresholding.

This year, we proposed a new algorithmic framework called SPADE, based on non-convex heuristics and which can accomodate both the sparse and cosparse prior. We studied their performance numerically and observed in particular that its cosparse version offers a very appealing trade-off between reconstruction performance and computational time [31], making it suitable for practical applications, even in real-time. We could also confirm our results by subjective listening tests conducted this year [12].

The work on cosparse audio declipping was awarded the Conexant best paper award at the LVA/ICA conference [31] and draw the attention of a world leading company in professional audio signal processing, with which some transfer has been negotiated.

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, time-frequency inpainting, joint source separation and declipping).

7.3.2. Blind Calibration of Impedance and Geometry

Participants: Rémi Gribonval, Nancy Bertin, Srdan Kitic.

Main collaborations: Laurent Daudet, Thibault Nowakowski, Julien de Rosny (Institut Langevin)

This 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 [23] (see section 7.1.3), we managed to preserve the sound source localization performance when the speed of sound is unkonwn, 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 were accepted for publication in ICASSP 2016 [38], [35].

7.3.3. 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, now at GIPSA-Lab)

When fitting a probability model to voluminous data, memory and computational time can become prohibitive. In this work, we propose 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 show that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. Our algorithm provides good reconstruction and scales to higher dimensions than previous probability mixture estimation algorithms, while consuming less memory in the case of numerous data. It also provides a privacy-preserving data analysis tool, since the sketch does not disclose information about individual datum it is based on [70], [68], [69]. This year, we consolidated our extensions to non-isotropic Gaussians, with new algorithms [49] and conducted large-scale experiments demonstrating its potential for speaker verification. A conference paper has been accepted to ICASSP 2016 [40] and a journal version is being finalized.

7.4. Recent results on tensor decompositions

tensor, multiway array, canonical polyadic decomposition, nonnegative tensor factorization

Multi-linear algebra is defined as the algebra of q-way arrays ($q > 2$), that is, the arrays whose elements are addressed by more than two indices. The first works dates back to Jordan who was interested in simultaneously diagonalizing two matrices at a time [93]. It is noteworthy that such two matrices can be interpreted as both slices of a three-way array and their joint diagonalization can be viewed as Hitchcock's polyadic decomposition [89] of the associated three-way array. Other works followed discussing rank problems related to multi-way structures and properties of multi-way arrays. However, these exercices in multilinear algebra were not linked to real data analysis but stayed within the realm of mathematics. Studying three-way data really started with Tucker's seminal work, which gave birth to the three-mode factor analysis [112]. His model is now often referred to as the Tucker3 model. At the same moment, other authors focused on a particular case of the Tucker3 model, calling it PARAFAC for PARAllel FACtor analysis [88], and on the means to achieve such a decomposition, which will become the famous canonical decomposition [73]. In honor to Hitchcock's pionneer work, we will call it the Canonical Polyadic (CP) decomposition.

Achieving a CP decomposition has been seen first as a mere non-linear least squares problem, with a simple objective criterion. In fact, the objective is a polynomial function of many variables, where some separate. One could think that this kind of objective is easy because smooth, and even infinitely differentiable. But it turns out that things are much more complicated than they may appear to be at first glance. Nevertheless, the Alternating Least Squares (ALS) algorithm has been mostly utilized to address this minimization problem, because of its programming simplicity. This should not hide the inherently complicated theory that lies behind the optimization problem. Moreover, in most of the applications, actual tensors may not exactly satisfy the expected model, so that the problem is eventually an approximation rather than an exact decomposition. This may result in a slow convergence (or lack of convergence) of iterative algorithms such as the ALS one [97]. Consequently, a new class of efficient algorithms able to take into account the properties of tensors to be decomposed is needed.

7.4.1. CP decomposition of semi-symmetric three-way arrays subject to arbitrary convex constraints

Participant: Laurent Albera.

Main collaborations : Lu Wang (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France), Jean-Christophe Pesquet (LIGM, France)

We addressed the problem of canonical polyadic decomposition of semi-symmetric 3rd order tensors (i.e. joint diagonalization by congruence) subject to arbitrary convex constraints. Sufficient conditions for the existence of a solution were proved. An efficient algorithm based on the Alternating Direction Method of Multipliers (ADMM) was then designed. ADMM provides an elegant approach for handling the additional constraint terms, while taking advantage of the structure of the objective function. Numerical tests on simulated matrices showed the benefits of the proposed method for low signal to noise ratios. Simulations in the context of nuclear magnetic resonance spectroscopy were also provided. This work was presented at the IEEE CAMSAP'15 conference [29].

7.4.2. Joint eigenvalue decomposition of non-defective matrices for the CP decomposition of tensors

Participant: Laurent Albera.

We proposed a fast and efficient Jacobi-like approach named JET (Joint Eigenvalue decomposition based on Triangular matrices) for the Joint EigenValue Decomposition (JEVD) of a set of real or complex non-defective matrices based on the $\mathbf{LU}$ factorization of the matrix of eigenvectors [98]. The JEVD can be useful in several contexts such as CP decomposition of tensors [99] and more particularly in Independent Component Analysis (ICA) based on higher order cumulants where it allows us to blindly compute the mixing matrix of sources with kurtosis of different signs. Regarding the proposed JET approach, contrary to classical Jacobi-like JEVD methods, its iterative procedure can be reduced to the search for only one of the two triangular matrices involved in the factorization of the matrix of eigenvectors, hence decreasing the numerical complexity. Two variants of the JET technique, namely JET-U and JET-O, which correspond to the optimization of two different cost functions were described in detail and these were extended to the complex case. Numerical simulations showed that in many practical cases the JET approach provides more accurate estimation of the matrix of eigenvectors than its competitors and that the lowest numerical complexity is consistently achieved by the JET-U algorithm.

7.5. Source separation and localization

Source separation, sparse representations, tensor decompositions, semi-nonnegative independent component analysis, 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 [106], which generalizes a number of existing techniques including our former study on spectral GMMs [61]. 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 [83], [92], [91], knowledge on the source positions and room characteristics [84], or a better initialization of parameters thanks to specific source localization techniques [67]. This accumulated progress lead to two main achievements last year: a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released [108] and we published an overview paper on recent and going research along the path of *guided* separation, *i.e.*, techniques and models allowing to incorporate knowledge in the process towards efficient and robust solutions to the audio source separation problem, in a special issue of IEEE Signal Processing Magazine devoted to source separation and its applications [113].

7.5.1. Towards real-world separation and remixing applications

Participants: Nancy Bertin, Frédéric Bimbot, Nathan Souviraà-Labastie, Ewen Camberlein, Romain Lebarbenchon.

Main collaboration: Emmanuel Vincent (EPI PAROLE, Inria Nancy)

While some challenges remain, work from previous years and our review paper on guided source separation [113] highlighted that progress has been made and that audio source separation is closer than ever to successful industrial applications, especially when some knowledge can be incorporated. This was exemplified by the contract with MAIA Studio, which reached its end in December 2014 and showed in particular how user input or side information could raise source separation tools to efficient solutions in real-world applications.

In some applicative contexts of source separation, several mixtures are available which contain similar instances of a given source. We have designed a general multi-channel source separation framework where additional audio references are available for one (or more) source(s) of a given mixture. Each audio reference is another mixture which is supposed to contain at least one source similar to one of the target sources. Deformations between the sources of interest and their references are modeled in a linear manner using a generic formulation. This is done by adding transformation matrices to an excitation-filter model, hence affecting different axes, namely frequency, dictionary component or time. A nonnegative matrix co-factorization algorithm and a generalized expectation-maximization algorithm are used to estimate the parameters of the model. Different model parameterizations and different combinations of algorithms have been tested on music plus voice mixtures guided by music and/or voice references and on professionally-produced music recordings guided by cover references. Our algorithms has provided improvement to the signal-to-distortion ratio (SDR) of the sources with the lowest intensity by 9 to 15 decibels (dB) with respect to the original mixtures [25]. Combining these techniques, with automatic audio motif spotting, we have proposed a new concept called SPORES (for SPOtted Reference based Separation) and applied it to guided separation of audio tracks [13] .

This year saw the beginning of 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 2015, we participated in a data collection campaign, and in benchmarking and adaptation of existing localization and separation tools to the particular context of this application.

7.5.2. 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*) [67]. 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 7.1.3).

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 started investigating last year 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 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 [78].

From now on, we have established an analytical model linking 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. This work is mainly lead by Aly Magassouba, whose Ph.D. is co-supervised by Nancy Bertin and François Chaumette. A conference paper presenting these first results was published this year [34] and another was submitted to ICRA 2016. Future work will include additional real-world experiments with the robot Romeo from Aldebaran Robotics, investigation of new tasks with active sensing strategies, explicit use of echoes and reverberation to increase robustness, and exploration of dense methods (control from raw acoustic signals rather than from acoustic features).

7.5.3. Brain source localization

Participants: Laurent Albera, Srdan Kitic, Nancy Bertin, Rémi Gribonval.

Main collaborations : Hanna Becker (GIPSA & LTSI, France), Pierre Comon (GIPSA, France), Isabelle Merlet (LTSI, France), Fabrice Wendling (LTSI, France)

From tensor to sparse models

The brain source imaging problem has been widely studied during the last decades, giving rise to an impressive number of methods using different priors. Nevertheless, a thorough study of the latter, including especially sparse and tensor-based approaches, is still missing. Consequently, we proposed i) a taxonomy of the methods based on a priori assumptions, ii) a detailed description of representative algorithms, iii) a review of identifiability results and convergence properties of different techniques, and iv) a performance comparison of the selected methods on identical data sets. Our aim was to provide a reference study in the biomedical engineering domain which may also be of interest for other areas such as wireless communications, audio source localization, and image processing where ill-posed linear inverse problems are encountered and to identify promising directions for future research in this area. This work was published in the IEEE Signal Processing Magazine [14].

A sparsity-based approach

Identifying the location and spatial extent of several highly correlated and simultaneously active brain sources from EEG recordings and extracting the corresponding brain signals is a challenging problem. In our comparison of source imaging techniques presented at ICASSP'14 [65], the VB-SCCD algorithm [81], which exploits the sparsity of the variational map of the sources, proved to be a promising approach. We proposed several ways to improve this method. In order to adjust the size of the estimated sources, we added a regularization term that imposes sparsity in the original source domain. Furthermore, we demonstrated the application of ADMM, which permitted to efficiently solve the optimization problem. Finally, we also considered the exploitation of the temporal structure of the data by employing L1,2-norm regularization. The performance of the resulting algorithm, called SISSY, was evaluated based on realistic simulations in comparison to VB-SCCD and several state-of-the-art techniques for extended source localization. This work was partially presented at EUSIPCO'14 [64] and a journal paper is in preparation.

Tensor- and sparsity-based approaches

The separation of EEG sources is a typical application of tensor decompositions in biomedical engineering. The objective of most approaches studied in the literature consists in providing separate spatial maps and time signatures for the identified sources. However, for some applications, a precise localization of each source is required.

To achieve this, a two-step approach was presented at the IEEE EMBC conference [26]. The idea of this approach is to separate the sources using the canonical polyadic decomposition in the first step and to employ the results of the tensor decomposition to estimate distributed sources in the second step, using the SISSY algorithm [64].

Next, we proposed to combine the tensor decomposition and the source localization in a single step [27]. To this end, we directly imposed structural constraints, which are based on a priori information on the possible source locations, on the factor matrix of spatial characteristics. The resulting optimization problem was solved using the alternating direction method of multipliers (ADMM), which was incorporated in the alternating least squares tensor decomposition algorithm. Realistic simulations with epileptic EEG data confirmed that the proposed single-step source localization approach outperformed the previously developed two-step approach.

7.5.4. Independent component analysis

Participant: Laurent Albera.

Main collaboration: Sepideh Hajipour (LTSI & BiSIPL), Isabelle Merlet (LTSI, France), Mohammad Bagher Shamsollahi (BiSIPL, Iran)

Independent Component Analysis (ICA) is a very useful tool to process biomedical signals including EEG data.

We proposed a Jacobi-like Deflationary ICA algorithm, named JDICA. More particularly, while a projection-based deflation scheme inspired by Delfosse and Loubaton's ICA technique (DelL$^{\mathbb{R}}$) [80] was used, a Jacobi-like optimization strategy was proposed in order to maximize a fourth order cumulant-based contrast built from whitened observations. Experimental results obtained from simulated epileptic data mixed with a real muscular activity and from the comparison in terms of performance and numerical complexity with the FastICA [90], RobustICA [114] and DelL$^{\mathbb{R}}$ algorithms, show that the proposed algorithm offers the best trade-off between performance and numerical complexity. This work was published in the IEEE Signal Processing Letters journal [21].

In addition, we illustrated in the ICA context the interest of being able to solve efficiently the (non-orthogonal) JEVD problem. More particularly, we showed that, when the noise covariance matrix is unknown and the source kurtoses have different signs, the joint diagonalization problem involved in the ICAR method [58] becomes a non-orthogonal JEVD problem. Consequently, by using our JET-U algorithm [98], giving birth to the MICAR-U (Modified ICAR based on JET-U) technique, we then provided a more robust ICA method. The identifiability of the MICAR-U technique was studied and proved under some conditions. Computer results given in the context of brain interfaces showed the better ability of the MICAR-U approach to denoise electro-cortical data compared to classical ICA techniques for low signal to noise ratio values. These results were presented in [98].

7.5.5. Semi-nonnegative independent component analysis

Participant: Laurent Albera.

Main collaboration: Lu Wang (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France), Huazhong Shu (LIST, China)

ICA plays also an important role in many other areas including speech and audio [62], [63], [74], [71], radiocommunications [79] and document restoration [111] to cite a few.

For instance in [111], the authors use ICA to restore digital document images in order to improve the text legibility. Indeed, under the statistical independence assumption, authors succeed in separating foreground text and bleed-through/show-through in palimpsest images. Furthermore, authors in [82] use ICA to solve the ambiguity in X-ray images due to multi-object overlappings. They presented a novel object decomposition technique based on multi-energy plane radiographs. This technique selectively enhances an object that is characterized by a specific chemical composition ratio of basis materials while suppressing the other overlapping objects. Besides, in the context of classification of tissues and more particularly of brain tumors [107], ICA is very effective. In fact, it allows for feature extraction from Magnetic Resonance Spectroscopy (MRS) signals, representing them as a linear combination of tissue spectra, which are as independent as possible [110]. Moreover, using the JADE algorithm [72] applied to a mixture of sound waves computed by means of the constant-Q transform (Fourier transform with log-frequency) of a temporal waveform broken up into a set of time segments, the authors of [71] describe trills as a set of note pairs described by their spectra and corresponding time envelopes. In this case, pitch and timing of each note present in the trill can be easily deduced.

All the aforementioned applications show the high efficiency of the ICA and its robustness to the presence of noise. Despite this high efficiency in resolving the proposed applicative problems, authors did not fully exploit properties enjoyed by the mixing matrix such as its nonnegativity. For instance in [82], the thickness of each organ, which stands for the mixing coefficient, is real positive. Furthermore, reflectance indices in [111] for the background, the overwriting and the underwriting, which correspond to the mixing coefficients, are also nonnegative. Regarding tissue classification from MRS data, each observation is a linear combination of independent spectra with positive weights representing concentrations [87]; the mixing matrix is again nonnegative.

By imposing the nonnegativity of the mixing matrix within the ICA process, we showed through computer results that the extraction quality can be improved. Exploiting the nonnegativity property of the mixing matrix during the ICA process gives rise to what we call semi-nonnegative ICA. More particularly, we performed the latter by computing a constrained joint CP decomposition of cumulant arrays of different orders [100] having the nonnegative mixing matrix as loading matrices. After merging the entries of the cumulant arrays in the same third order array, the reformulated problem follows the semi-symmetric semi-nonnegative CP model defined in section 7.4.1 . Hence we use the new method described in section 7.4.1 to perform semi-nonnegative ICA. Performance results in biomedical engineering were given in the paper cited in section 7.4.1 .

7.6. Audio and speech content processing

Audio segmentation, speech recognition, motif discovery, audio mining

7.6.1. Audio motif discovery and spotting

Participants: Frédéric Bimbot, Nathan Souviraà-Labastie.

This work was performed in close collaboration with Emmanuel Vincent from Inria Nancy-Grand Est.

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio contents.

Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. The developed algorithms allows discovering and collecting occurrences of repeating patterns in the absence of prior acoustic and linguistic knowledge, or training material. When the audio pattern is determined in a user supervised fashion, the task becomes that of motif spotting.

Investigated in the context of SPORES (SPOtted Reference based Separation) [13], audio motif spotting has been illustrated as a useful way to exploit redundancy in audio contents, for guided source separation purposes.

7.6.2. Mobile device for the assistance of users in potentially dangerous situations

Participants: Romain Lebarbenchon, Ewen Camberlein, Frédéric Bimbot.

The S-Pod project is a cooperative project between industry and academia aiming at the development of mobile systems for the detection of potentially dangerous situations in the immediate environment of a user, without requiring his/her active intervention.

In this context, the PANAMA research group has been involved in the design of algorithms for the analysis and monitoring of the acoustic scene around the user, yielding audio-based information which can be fused with other sensors (physiological, positional, etc.) in order to trigger an alarm (and subsequent appropriate measures) when needed.

The last phase of the project has been dedicated towards robustness improvement of audio scene analysis, with a particular focus on threat vs non-threat detection, on the basis of adaptive training scenarii. Knowledge and know-how transfer has been achieved for the hardware implementation of the designed methods and the efficient integration into an operational prototype.

7.7. Music Content Processing and Music Information Retrieval

Acoustic modeling, non-negative matrix factorisation, music language modeling, music structure

7.7.1. Music structure modeling by System & Contrast

Participants: Frédéric Bimbot, Corentin Louboutin.

The *System & Contrast* (S&C) model aims at describing the inner organization of structural segments within music pieces in terms of : (i) a carrier system, i.e. a sequence of morphological elements forming a multi-dimensional network of self-deducible syntagmatic relationships and (ii) a contrast, i.e. a substitutive element, usually the last one, which partly departs from the logic implied by the rest of the system [16].

With a primary focus on pop music, the S&C model provides a framework to describe internal implication patterns in musical segments by encoding similarities and relations between its constitutive elements so as to minimize the complexity of the resulting description. It is applicable at several timescales and to a wide variety of musical dimensions in a polymorphous way, therefore offering an attractive meta-description of different types of musical contents.

We have established the filiation of the S&C model as an extension of Narmour's Implication-Realization model [104], [105] and Cognitive Rule-Mapping [103].

We have introduced the Minimum Description Length scheme as a productive paradigm that supports the estimation of S&C descriptions and establishes promising connections between Music Data Processing and Information Retrieval on the one hand, and modern theories in Music Perception and Cognition on the other hand, together with interesting perspectives in other areas in Musicology.

The model is currently being investigated for the multi-scale description of chord sequences.

7.7.2. Tree-based representation of music pieces

Participants: Frédéric Bimbot, Corentin Guichaoua.

Modeling music structure, i.e. the organisation of musical elements and their relationships within a piece of music, is an open problem of primary importance in MIR.

To address this challenge, we approach music structure description as the inference of a low complexity generative grammar able to account for the music piece, itself represented as a sequence of symbols.

Originally introduced for the inference of structure in DNA sequences, Straight-Line Grammars (SLG) form a particular subclass of Context-Free Grammars (CFG) which can be used to model symbolic sequences and to represent them as hierarchical trees. However, SLGs appear to be poorly suited to some particularities of musical patterns, such as segmental regularities, closure substitutions and specific style structures.

We are designing and investigating formal and algorithmic extensions of SLGs as SLEGs (Straight-Line Edition Grammars). Based on a more general minimum description criterion, the SLEG extension allows alterations in the generation step and enables the use of priors in the grammar inference process. Current work includes a diagnostic comparison between the various approaches on the structural segmentation of chord sequences from pop songs.

SEMAGRAMME Project-Team

6. New Results

6.1. Syntax-semantics interface

6.1.1. 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 [40], [31], [26], [59]. 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 [58].

Frames can be formalized as extended typed feature structures [71], [50] and specified as models of a suitable logical language. Such a language allows for the composition of lexical frames on the sentential level by means of an explicit syntax-semantics interface [50]. Yet, this logical framework does not provide a direct link between Frames and truth-conditional semantics, where natural language utterances are considered with respect to the conditions under which they are true or false. In particular, it does not provide means for the lexical items to introduce explicit quantification over entities or events.

To overcome these limitations, we proposed use Hybrid Logic (HL) [27], [25]. HL is an extension of modal logic. As such, it is well-suited to the description of graph structures. Moreover, HL introduce *nominals*, that allow the logical formulas to refer to specific nodes of the graph. It is then possible, for example, to specify when two edges should meet. Moreover, it introduces *variables* for nodes, and the associated *quantifiers*, that can appear in the logical formulas. We used this framework to model quantification in Frame Semantics [23], [18]

6.1.2. Compositionality and Modularity

One says that a semantics is compositional when it allows the meaning of a complex expression to be computed from the meaning of its constituents. One also says that a system is modular if it is made of relatively independent components. In the case of a semantic system (e.g, a Montague grammar), we say that it is modular if the ontology on which it is based (including notions such as *truth, entities, events, possible worlds, time intervals, state of knowledge, state of believe, ...*) is obtained by combining relatively independent simple ontologies.

The intensionalization procedure introduced in [4] provides a first step towards modularity. It allows the extensional interpretation of a language to be transformed into an intensionalized interpretation that offers room for accommodating truly intensional phenomena. Moreover, this procedure is conservative in the sense that it preserves the truth conditions of sentences. Another instance of such a procedure is provided by the dynamization procedure described in [57], which allows a static interpretation to be turned into a dynamic one capable of accommodating phenomena related to discourse dynamics.

In [15], we showed that both the intensionalization and dynamization procedures are instances of an abstract general scheme for which conservativity results may be established using the notion of logical relation.

6.1.3. Abstract Categorial Parsing

Kanazawa [53], [54] has shown how parsing and generation may be reduced to datalog queries for a class of grammars that encompasses mildly context-sensitive formalisms. These grammars, which he calls *context-free λ-term grammars*, correspond to second-order abstract categorial grammars.

In [14], we showed how Kanazawa's reduction may be carried out in the case of abstract categorial grammars of a degree higher than two. To this end, we reduced the parsing problem for general Abstract Categorial Grammars to a provability problem in Multiplicative Exponential Linear Logic.

6.2. Discourse dynamics

6.2.1. Discourse Structure Modeling

It is usually assumed that the internal structure of a text, typically characterized by discourse or rhetorical relations, plays an important role in its overall interpretation. In order to build such a structure, some approaches rely on discourse grammars. The key idea is to consider the structural regularities in discourse structure similarly as syntactic regularities. A particular trend relies on tree grammars. This trend has been further developed by integrating the modeling of both clausal syntax and semantics, and discourse syntax and semantics within the frameworks of Tree-Adjoining Grammar (TAG) [48], [49] and TAG for Discourse (D-LTAG) [79], [41], [80], [42].

Two important features characterize these approaches. First, while they use a single grammatical formalism, two different grammars are used for syntactic parsing and then for discourse parsing. In addition to adding an intermediate processing step, this two-tiered treatment both complicates the modeling of connectives that are ambiguous in their syntactic and discourse use, and prevents using standard disambiguation techniques. Second, some discourse structures better represented by directed acyclic graphs (DAG) than by trees are not accounted for.

In order to address the second issue of building DAG structures, [36], [37] have proposed Discourse Synchronous TAG (D-STAG), a TAG based approach together with a higher-order interpretation of sentences using Synchronous Tree-Adjoining Grammar (STAG) [67], [77].

We developed a method to interface a sentential grammar and a discourse grammar without resorting to an intermediate processing step. The method is general enough to build discourse structures that are DAG and not only trees. Our analysis is based on D-STAG. We also use an encoding of TAG into ACG. This encoding allows us to express a higher-order semantic interpretation that enables building DAG discourse structures on the one hand, and to smoothly integrate the sentential and the discourse grammar thanks to the modular capability of ACG. The results has been published [13] and all the examples of the article have been implemented and may be run and tested with the ACGtk software (see 5.1).

6.2.2. Effects and Handlers

We made the argument that pragmatics are to semantics what side effects are to calculations in a programming language. We demonstrated this parallel on two aspects.

First off, both pragmatics and side effects serve the same function. Side effects in programming languages account for the effects of expressions that reach beyond their scope and for the way a language interacts with the world of its users. Pragmatics is concerned with phenomena that also involve the non-immediate effects of expressions (e.g., discourse anaphora, presupposition accommodation) and with the way language interacts with the world of its users. Secondly, we pointed out that very similar formal theories are being used to treat the both of them (i.e. monads and continuations).

Having established this parallel, we then put forward a preliminary proposal of integrating semantics and pragmatics while keeping them separate by assigning effectful computations of truth values as meanings of linguistic expressions. In this way, we can implement the pragmatics at the level of the side effects and then focus on pure semantics at the level of values.

6.3. Common basic resources

6.3.1. Graph Rewriting

Bruno Guillaume and Guy Perrier have proposed to use Graph Rewriting for parsing syntactic dependencies [17]. It is an application of a Graph Rewriting formalism that they have established with Guillaume Bonfante and Mathieu Morey [32] and implemented in the Grew software [47]. They have developed a system of rewriting rules dedicated to French, which they have evaluated by parsing the Sequoia corpus [33].

6.3.2. Categorial Logic

Elaborating on the work of Grishin [45], Moortgat has introduced the non-associative Lambek-Grishin calculus (**LG**) as the foundations of a new kind of symmetric categorial grammar [63], [64], which allows for the treatment of linguistic phenomena such as displacement or discontinuous dependencies.

In [16], we compared **LG** with the non-associative classical Lambek calculus (**CNL**) introduced by de Groote and Lamarche [81]. We provided a translation of **LG** into **CNL**, which allows **CNL** to be seen as a non-conservative extension of **LG**. We then introduced a bimodal version of **CNL** that we called 2-**CNL**. This allowed us to define a faithful translation of **LG** into 2-**CNL**. Finally, we showed how to accommodate Grishin's interaction principles by using an appropriate notion of polarity. From this, we derived a new one-sided sequent calculus for **LG**.

6.3.3. Deep Syntax Annotation of the Sequoia French Treebank

Deep-sequoia introduces a deep syntactic representation scheme for French, built from the surface annotation scheme of the Sequoia corpus and abstracting away from it [69]. This scheme expresses the grammatical relations between content words. When these grammatical relations take part into verbal diatheses, the diatheses are considered as resulting from redistributions from the canonical diathesis, which is retained in the annotation scheme. The first version of the deep-sequoia corpus was released in 2014.

In November 2015, a new version (7.0) of the corpus was release (see http://deep-sequoia.inria.fr). Most of the modifications were corrections of annotations that improve the overall consistency of the corpus. Marie Candito and Guy Perrier have published the annotation guidelines associated with the corpus in [22].

6.3.4. Large Scale Grammatical Resources

Guy Perrier and Bruno Guillaume have achieved the development of a French grammar FRIGRAM with a large coverage [12] in the formalism of Interaction Grammars [5]. The originality of the formalism lies in its system of polarities which expresses the resource sensitivity of natural languages and which is used to guide syntactic composition. We present the principles underlying grammar design, highlight its modular architecture and show that the lexicon used is independent of the grammar formalism. We also introduce the "companion property", and show that it helps to enforce grammar consistency.

6.3.5. Universal Dependency Treebank

Bruno Guillaume participates with Marie-Catherine de Marneffe to the production of the French sub-corpus of the Universal Dependency Treebank [68]. In November 2015, the version 1.2 was released. On the French sub-corpus, Grew was used to detect inconsistency and to correct automatically systematic errors.

Chroma Team

7. New Results

7.1. Sensor Fusion

7.1.1. Observability properties of the visual-inertial structure

Participant: Agostino Martinelli.

We continued to investigate the visual-inertial structure from motion problem by further addressing the following issues:

1. analytically deriving its observability properties in challenging scenarios (i.e., when some of the system inputs are unknown and act as disturbances);
2. obtaining simple and efficient methods for data matching and localization.

Regarding the first issue, we extended our previous results (published last year on the journal Foundations and Trends in Robotics [43]) by also including the extreme case of a single point feature and when the camera is not extrinsically calibrated. Even if this extension seems to be simple, the analytic computation must be totally changed. Indeed, by including in the state the camera extrinsic parameters, the computation, as carried out in [43] in the case when the camera is calibrated, becomes prohibitive.

The problem of deriving the observability properties of the visual-inertial structure from motion problem, when the number of inertial sensors is reduced, corresponds to solve a problem that in control theory is known as the Unknown Input Observability (UIO). This problem is still unsolved in the nonlinear case. In [43] 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 [43] 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 and consequently prohibitive in our case. Our effort to deal with this fundamental issue, was devoted to separate the information on the original state from the information on its extension. We fully solved this problem in the case of a single unknown input. For the general case, we partially solved this problem and we suggested a technique able to partially perform this separation. Since these results are very general (their validity is not limited to the visual-inertial structure from motion problem) we presented them at two international conferences on automatic control (SIAM on Control and Applications, [18] and MED, [16]). By applying these new methods to the the visual-inertial structure from motion problem, we obtained the following result. Even in the case of a single point feature, the information provided by a sensor suit composed by a monocular camera and two inertial sensors (along two independent axes and where at least one is an accelerometer) is the same as in the case of a complete inertial measurement unit (i.e., when the inertial sensors consist of three orthogonal accelerometers and three orthogonal gyroscopes). This result has been presented at ICRA, [17].

Regarding the second issue, our focus was in the framework of Micro Aerial Vehicle navigation. State of the art approaches for visual-inertial sensor fusion 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. Last year, we published, on the journal of computer vision, a closed-form solution providing such an initialization [42]. This solution determines the velocity (angular and linear) of a monocular camera in metric units by only using inertial measurements and image features acquired during a short time interval. This year, we study the impact of noisy sensors on the performance of this closed-form solution. Additionally, starting from this solution, we proposed new methods for both localization and data matching in the context of micro aerial navigation. These methods have been tested in collaboration with the vision and perception team in Zurich (in the framework of the ANR-VIMAD) and published on the journal of Robotics and Autonomous Systems [4].

7.1.2. Sensing floor for Human & objects localisation and tracking

Participants: Mihai Andries (inria Nancy, Larsen), Olivier Simonin, François Charpillet (inria Nancy, Larsen).

In the context of the PhD of Mihai Andries, co-advised by François Charpillet (Inria Nancy, Larsen) and Olivier Simonin, we investigated a large distributed sensor — a grid of connected sensing tiles on the floor — that was developped by the Maia team, at Nancy, in 2012.

Localization, tracking, and recognition of objects, robots and humans are basic tasks that are of high value in the applications of ambient intelligence. Sensing floors were introduced to address these tasks in a non-intrusive way. To recognize the humans moving on the floor, they are usually first localized, and then a set of gait features are extracted (stride length, cadence, and pressure profile over a footstep). However, recognition generally fails when several people stand or walk together, preventing successful tracking. In the Phd, defended on December 15 [27], we proposed a detection, tracking, and recognition technique which uses objects' weight. It continues working even when tracking individual persons becomes impossible. Inspired by computer vision, this technique processes the floor pressure-image by segmenting the blobs containing objects, tracking them, and recognizing their contents through a mix of inference and combinatorial search. The result lists the probabilities of assignments of known objects to observed blobs. The concept was successfully evaluated in daily life activity scenarii, involving multi-object tracking and recognition on low-resolution sensors, crossing of user trajectories, and weight ambiguity. This model can be used to provide a probabilistic input for multi-modal object tracking and recognition systems. The model and the experimental results have been published in Journal IEEE Sensors [1] and international conference ICRA 2015 [7].

7.2. Bayesian Perception

Participants: Christian Laugier, Lukas Rummelhard, Amaury Nègre, Jean-Alix David, Procópio Silveira-Stein, Jerome Lussereau, Tiana Rakotovao, Nicolas Turro (sed), Jean-François Cuniberto (sed), Diego Puschini (cea Dacle), Julien Mottin (cea Dacle).

7.2.1. Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT)

Participants: Lukas Rummelhard, Amaury Nègre, Christian Laugier.

In 2015, 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. This work exploits the *Bayesian Occupancy Filter (BOF)* paradigm [28], developed and patented by the team several years ago [0]. It also extends the more recent concept of *Hybrid Sampling BOF(HSBOF)* [46], whose purpose was to adapt the concept to highly dynamic scenes and to analyse 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 2015 was to overcome some of the shortcomings of the *HSBOF* approach [0] , and to obtain a better understanding of the observed dynamic scenes through the introduction an additional object level into the model. The new framework, whose development will be continued in 2016, is called *Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT)* [10]. 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.

[0] The *Bayesian programming formalism* developed in e-Motion, pioneered (together with the contemporary work of Thrun, Burgards and Fox [53]) 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 [31], [48], etc), the retrospective object level analysis by dynamic grid segmentation can be computationally expensive and subjected to some data association errors.

[0] Nanoelec Technological Research Institute (Institut de Recherche Technologique Nanoelec)

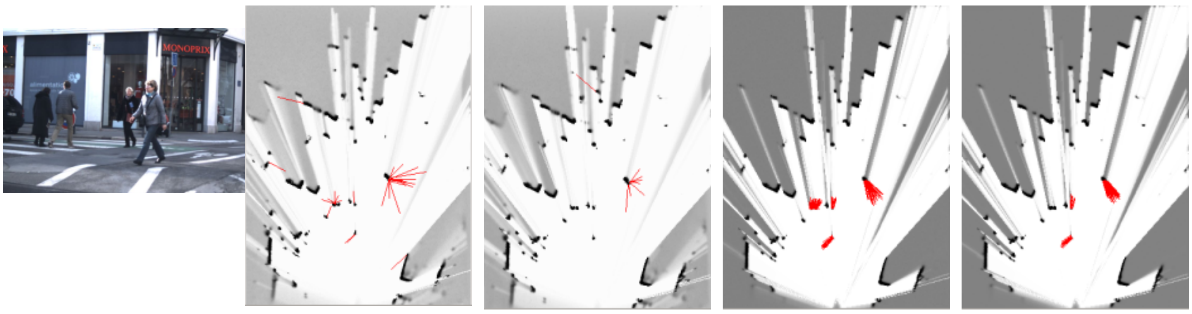
[0] National Research Agency (Agence Nationale de la recherche)

The *CMCDOT* approach introduces an drastic change in the underlying formal expressions: instead of directly filtering the occupancy data, we have added *hidden states* for representing what is currently present in a cell. Then, the occupancy distribution can then be inferred from those hidden states. Besides presenting a clear distinction between static and dynamic parts, the main interest of this new approach is to introduce a specific processing of dataless areas, excluding them from the velocity estimation (and consequently optimizing the processing of the dynamic parts) and disabling their temporal persistence (which is used to generate estimation bias in newly discovered areas). This updated formalism also enables the introduction of an appropriate formal model for the particle initialization and management (which was previously more isolated).

Another important added feature is the automatic segmentation of the dynamic parts of the occupancy grid, according to its shapes and dynamics. While the *CMCDOT* tracks spatial occupancy in the scene without object segmentation, Detection and Tracking of Moving Objects (DATMO) is often required for high level processing. A standard approach would be to analyse the *CMCDOT* outputs, to apply a clustering algorithm on the occupancy grid (enhanced by velocities), and to use those clusters as potential object level targets. This clustering can turn out to be computationally expensive, considering the grid dimensions and the size and complexity of the dynamic particle model. The basic idea of our new approach is to exploit the particle propagation process within the *CMCDOT*: the way particles are resampled can leads to the wanted segmentation after a number of time steps. After initialization, at each step, the particles that correctly fit the motion of a dynamic object are multiplied, those which do not are forgotten. In a few steps, the best particles propagate in the object, and the object motion is fully described by a set of particles deriving from a common particle root. By marking each particle at the initialization step with a unique identification number, all the dynamic areas which are coherent in term of space and motion are marked after few iterations. The convergence of those markers is fastened by additional rules.

Sequence	Grid size (m)	HSBOF	CMCDOT
Highway	20x70	76.9%	23.5%
Semi-Urban	30x60	89.3%	46.7%
City Center	30x60	93.2%	40.1%

Figure 2. Estimation of irrelevant particle allocation ratio.

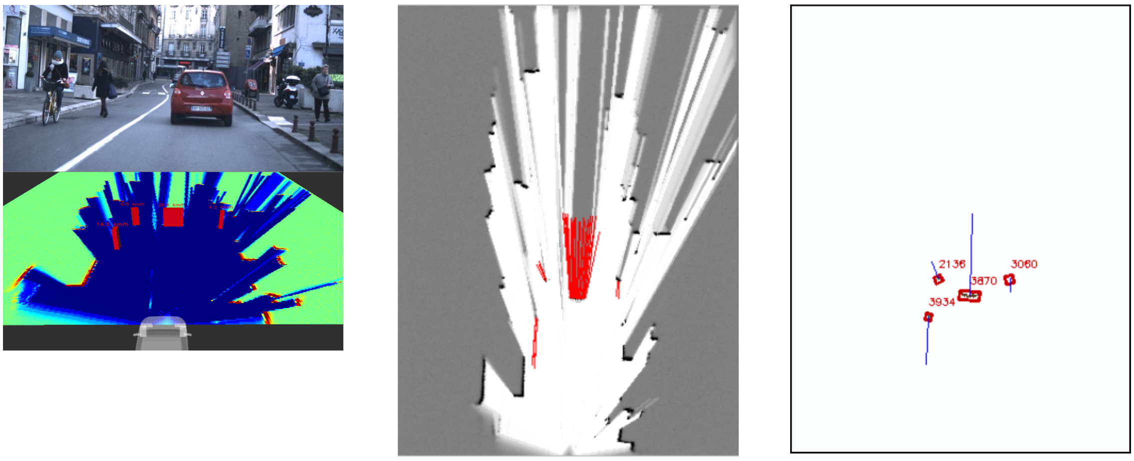


(a) (b) (c) (d) (e)

Figure 3. Results of the HSBOF with 262144 and 32768 particles (b) and (c), and of the CMCDOT with the same number of particles (d) and (e). Red segments represent the average estimated per-cell velocity. They show that the CMCDOT is more accurate and still manages to track most of the moving pedestrians (even with a severely reduced number of samples), whereas the HSBOF loses track of almost all objects.

Experimental results showed that the insertion of an "unknown" state in the model leads to a better distribution of dynamic samples on observed areas (see figure 2) and also allows us to be more reactive and accurate on the velocity distributions, while requiring less computing power (see figure 3).

The intrinsic clustering approach has also been tested on real road data, showing promising results in real-time tracking of moving objects, regardless of their type. The method could be improved by managing split-and-merge events that can occur in complex urban environment (see figure 4).



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

Figure 4. Result of the dynamic objects clustering. (a) Camera image; (b) 3D view of the grid with detected objects; (c) resulting occupancy grid with velocity; (d) extracted dynamic objects (red boxes) with velocity (blue segments) and id.

7.2.2. Multimodal dynamic objects classification

Participants: Amaury Nègre, Jean-Alix David.

The method described in section 7.2.1 allows to obtain a list of dynamic objects and to track each object over time. In order to increase the level of representation of the environment, we have developed a method to classify detected objects using both the camera images and the occupancy grid representation estimated by the *CMCDOT*. For each detected object, the bounding box of the object is projected in the camera image and a local image is extracted from the camera. Jointly, we can extract a patch from the occupancy grid around the dynamic object position. The extracted camera image and the occupancy grid patch can then be used as the input of a Deep Neural Network (DNN) to identify the class of the object. The DNN we designed is a combination on two classic neural networks, the "ImageNet" Convolutional Neural Networks [35] for the camera image input and the "LeNet" [37] for the occupancy grid input (see fig 5).

To train and evaluate the model, a dataset has been created from the data recorded with the Lexus platform. We extracted the camera images and the occupancy grid for each object detected by the *CMCDOT* module, then we manually annotated the object class among "pedestrian", "crowd", "car", "truck", "two-wheelers" and "misc" categories. The resulting dataset contains more than 100000 camera images & occupancy grid pairs. The training process and the classification module has been done by using the open source library caffe [33]. An example of the obtained results is shown on fig 6 . The percentage of good classification is greater than 90% on our evaluation dataset.

7.2.3. Visual Map-Based Localisation with OSM

Participants: Jean-Alix David, Amaury Nègre.

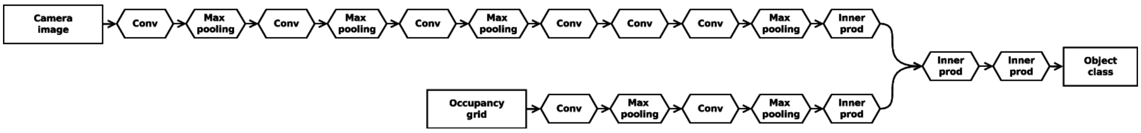


Figure 5. Deep neural network used for obstacle detection.

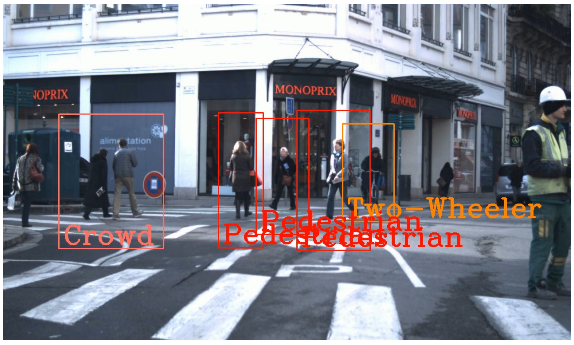


Figure 6. Result of the dynamic obstacle classification.

This module aims to improve both the global localization provided by the GPS [0] and the lane-relative localization information estimated by a lane tracker by combining their mutual strengths. The idea is to detect lane markings on the road using a camera, and then to compare the extracted lines with those stored in the map. This is done using the ICP [0] algorithm. This work is described in a confidential Toyota project report entitled *Real Traffic Data Acquisition and Risk Assessment Experiments*.

7.2.3.1. The map

Our solution is based on a post-processed OSM [0] map shown on figure 7 . Typically, these maps contain information on the roads and lanes, but contain no information about lane markers on the ground. Thus, we ran a semi-manual process to complete the existing maps with information about the number and type of markers.

New data are stocked in a local server. Requests can be sent to this server to fetch map data using HTTP protocol.

7.2.3.2. Line detection

The line extraction is done using ridge detection on a top-down view of the camera image. Only one monocular camera is used, as it is an inexpensive sensor, and needs only to be calibrated once. The line detection is based on an algorithm using Laplacian to extract ridges of the monochrome image. The algorithm is implemented for parallelized calculation using CUDA on a GPU, for an improved performance. Figure 8 shows the results of the ridge detector.

[0]Global Positioning System
[0]Iterative Closest Point
[0]OpenStreetMap

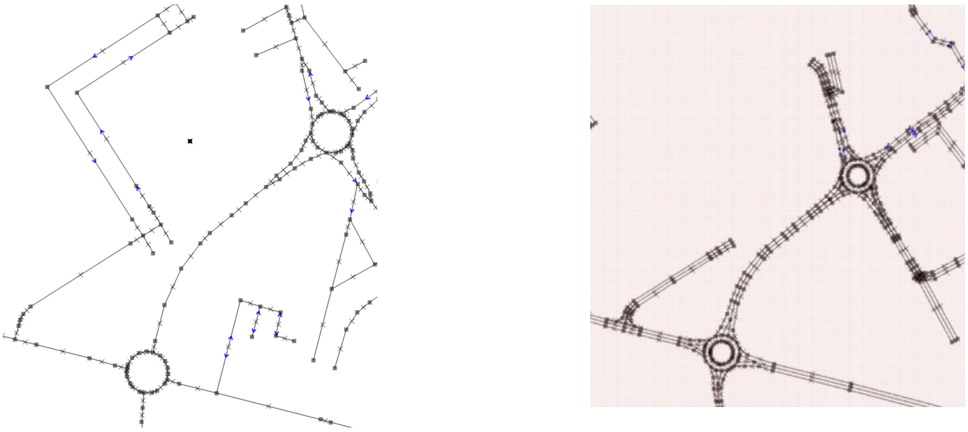


Figure 7. Data conversion. (a) Raw OSM data, a line represents a road and it is not possible to see the lanes. (b) Modified OSM data, with lane markings.

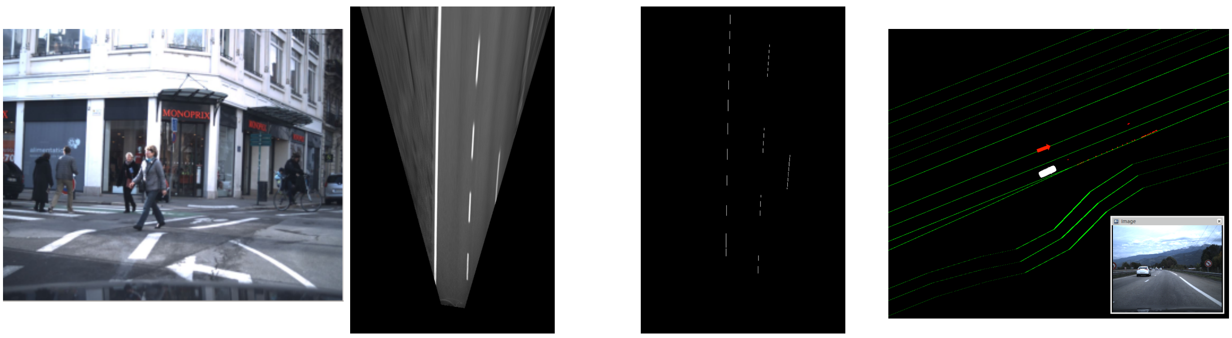


Figure 8. Ridges detection: (a) Input image (b) Projected image (c) Detected ridges. ICP correction on highway (d)

7.2.3.3. ICP-based line matching

The extracted lines are matched and aligned with the map using the ICP algorithm to improve the localization of the vehicle. The ICP algorithm iteratively minimizes the total alignment error between the points detected as ridges and the segments of line extracted from the map. Finally, figure 8 .d shows how the algorithm can correct the vehicle localization. The algorithm is able to accurately track the orientation and position. However, the lateral displacement may be off by a multiple of the lane width, depending on how the algorithm has been initialized. In practice, this effect is often mitigated due to the existence of single-lane roads such as highway entrances.

The results are very promising on highways, but the algorithm has a lower performance on other types of roads, mostly due to irregularities.

7.2.4. Integration of Bayesian Perception System on Embedded Platforms

Participants: Tiana Rakotovao, Christian Laugier, Diego Puschini(cea Dacle), Julien Mottin(cea Dacle).

Safe autonomous vehicles will emerge when comprehensive perception systems will be successfully integrated into vehicles. However, our Bayesian Perception approach requires high computational loads that are not supported by the embedded architectures currently used in standard automotive ECUs.

To address this issue, we first explored new embedded hardware architecture credible for the integration of OGs [0] into autonomous vehicles [19]. We studied in particular recent emerging many-core architectures, which offer higher computing performance while drastically reducing the required power consumption (typically less than 1W). In such architectures, the computation of OGs can be divided into several independent tasks, executed simultaneously on separated processing core of a many-core.

Experiments were conducted on data collected from urban traffic scenario, produced by 8 LIDAR layers mounted on the Inria-Toyota experimental Lexus vehicle. These experiments demonstrate that the many-core produces OGs largely in real-time: 6 time faster than the sensor reading rate.

Besides, we also proposed a mathematical improvement of the OG model, for performing multi-sensor fusion more efficiently than the standard approach presented in [29]. In our approach, the fusion of occupancy probabilities requires fewer operations. This model improvement makes it possible the implementation of OG-based multi-sensor fusion on simple hardware architectures. This perspective applies to microcontroller, ASICs or FPGAs which are more and more present in computing platforms recently present on the automotive market.

7.2.5. Experimental Vehicle Renault ZOE

Participants: Nicolas Turro (sed), Jean-François Cuniberto (sed), Procópio Silveira-Stein, Amaury Nègre, Lukas Rummelhard, Jean-Alix David, Christian Laugier.

7.2.5.1. Experimental Vehicle Renault ZOE

In the scope of the *Perfect* projet of the IRT nanoelec, we have started to develop in 2014, an experimental platform based on an equipped Renault Zoe. The development of this platform has been pursed in 2015.

The vehicle has been enhanced with a tablet to display the new HMI [0], figure 9 (a) illustrates. The HMI displays the detected dynamic objects over the camera image and the graph of collision risk at different time horizon.

New experiments have also been designed to test the perception algorithms and the recent implementation of the collision risk alert. These experiments simulate collisions with people using a fabric mannequin, as shown on figure 9 (b), and an inflatable ball.

[0] Occupancy Grids

[0] Human Machine Interface

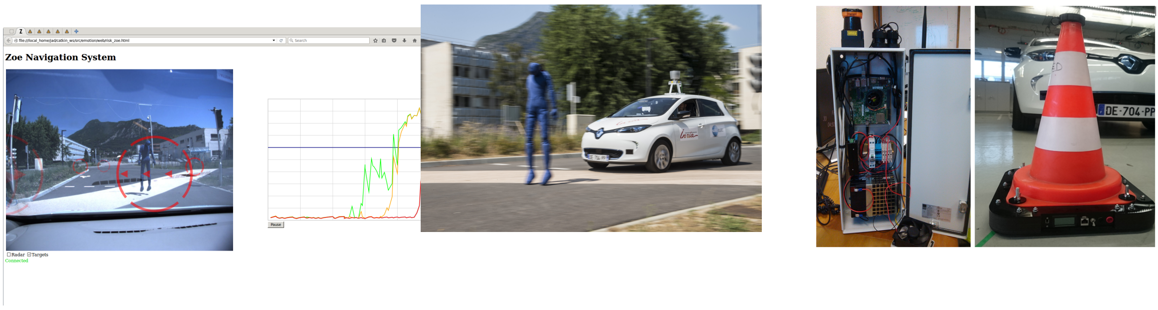


Figure 9. (a) Display of the HMI (b) Collision simulation with a mannequin (c) On left: picture of the smartbox, on right: picture of the cone.

Finally, we have also developed two movable devices in order to enhance V2X [0] communication experiments (see figure 9 (c)):

1. A movable communicating cone equipped with a GPS and a V2X communication box, which broadcast its position to near V2X listeners.
2. A movable smartbox equipped with a GPS, a V2X communication box, a LIDAR sensor and a Nvidia Tegra K1 board. The *CMCDOT* algorithm is implemented on it, and the detected objects are broadcasted to other communicating devices. The smartbox can be mounted on another vehicle or be placed as part of a static infrastructure. Both are alimented by batteries and aim at minimizing their energy consumption.

7.3. Situation Awareness

Participants: Christian Laugier, Alejandro Dizan Vasquez Govea, Procópio Silveira-Stein, David Sierra-Gonzalez, Mathieu Barbier, Stephanie Lefevre(uc Berkeley).

7.3.1. Framework for Motion Prediction and Collision Risk Assessment

Participants: Christian Laugier, Alejandro Dizan Vasquez Govea, Procópio Silveira-Stein, Stephanie Lefevre(uc Berkeley).

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). Throughout 2016, we have continued this line of work by developing several new frameworks for Motion Prediction and Collision Risk Assessment in complex dynamic scenes involving multiple moving agents having various behaviors.

A first contribution has been the extensive experimental validation in real conditions –together with the University of Berkeley– of our *Intention-Expectation* approach: a high-level approach to risk assessment which avoids the complexity of trajectory-level reasoning while being able to take multi-vehicle interactions into account [9]. These results have also been integrated into a Mooc course at the graduate and undergraduate levels [25]. They have also been presented in several invited talks [24] [21] [22] [23].

Another contribution relies in the implementation of some the proposed models on two experimental vehicles (Lexus and Zoé experimental platforms). As mentioned in section 7.2.5 , several experiments on short-term collision risk assessement have been successfuly conducted with these platforms (c.f. [10], [15]). This work will be continued in 2016, in the scope of our ongoing collaborative projects with Toyota, Renault and IRT nanoelec.

[0] Vehicle-to-Vehicle and Vehicle-to-Infrastructure

7.3.2. Planning-based motion prediction for collision risk estimation in autonomous driving scenarios

Participants: David Sierra-Gonzalez, Alejandro Dizan Vasquez Govea, Christian Laugier.

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- Gonzales.

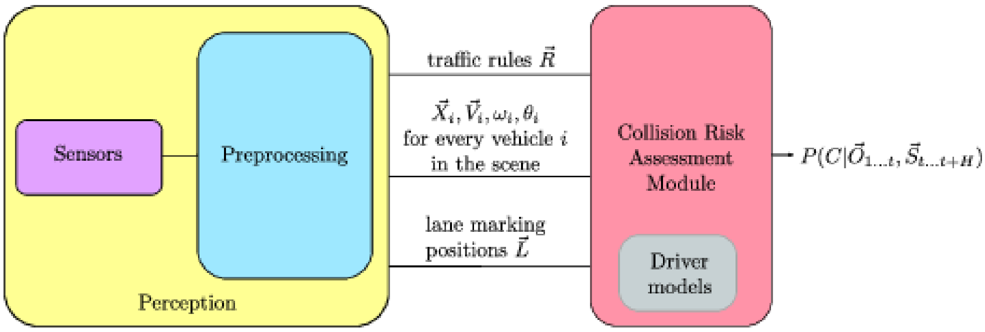


Figure 10. Black box model for the Collision Risk Assessment Module. The number of inputs expected from the perception module is not comprehensive at this point.

Figure 10 shows the black box model of the system. At a given timestep, the system takes the following inputs: the traffic rules in effect; the position, velocity, angular velocity and heading of each vehicle i in the scene; and the position of the lane markings. Thus, at each timestep t we construct an observation vector $1\overrightarrow{O}_t = \left(\overrightarrow{R}_t, \{\overrightarrow{X}_i, \overrightarrow{V}_i, \omega_i, \theta_i\}_t, \overrightarrow{L}_t\right)$ with all the high-level perception inputs, and a state vector $\overrightarrow{S_t}$ with only the minimum variables necessary to describe the scene. The proposed system aims to calculate the probability of collision C of the ego-vehicle for a sequence of future states up until a fixed time horizon H. That can be expressed as $P(C|\overrightarrow{O}_{1\cdots t}, \overrightarrow{S}_{t\cdots t+H})$. This information can then be used by a path-planner to decide upon the safest trajectory.

One key factor for the correct estimation of collision risk is the ability to predict the motion of the dynamic obstacles in the scene, that is, the other drivers. We opt here for a planning-based approach, which assumes that drivers instinctively act as to maximize a reward (or equivalently, minimize a cost). This reward 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. Given such a reward function, Markov Decision Processes (MDP) constitute an adequate framework for the motion prediction problem. Moreover, by using Inverse Reinforcement Learning (IRL) algorithms, we can obtain such reward function directly from expert demonstrations (i.e. simply observing how people drive).

At this point, two well-known IRL algorithms ([26], [58]) have been implemented and used to obtain a generic driver model from human demonstrations performed on a highway simulator. This driver model can now be used to predict the future behavior of the dynamic obstacles in the scene.

7.4. Motion-planning in human-populated environment

7.4.1. Planning-based motion prediction for pedestrians in crowded environments

Participant: Alejandro Dizan Vasquez Govea.

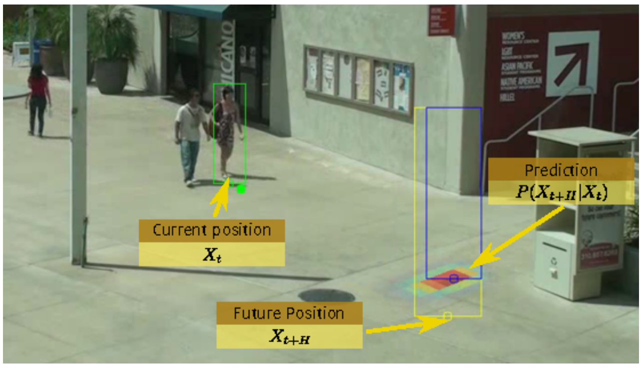


Figure 11. Example of spatiotemporal state prediction for a single agent, indicating its current and future position after 8 sec (green and yellow boxes) as well as the predicted position probability and its maximum value (blue box)

We have also explored the application of motion planning algorithms to the prediction of human motion (Fig. 11). We have proposed a novel planning-based motion prediction approach [12] which addresses the weaknesses of the previous state-of-the-art motion prediction technique [34], namely:

1. *High computational complexity*. This is dealt with by using the Fast Marching Method (FMM) [49] an efficient deterministic planning algorithm which computes the cost-to-go to a given location for every cell of a grid representing the agent's workspace. This grid is then used in a novel goal prediction algorithm and to produce a path-like prediction equivalent to the output of the Markov Decision Processes (MDPs) used by Kitani.
2. *Limited ability to model the temporal evolution along the predicted path*: this is addressed through the use of a velocity-dependent probabilistic motion model which is used to estimate a probability distribution of the future agent's position. This is then fused with a novel cost-based model to produce a full spatiotemporal prediction.
3. *Constant-goal assumption*. We propose a gradient-based goal prediction approach which does not rely on filtering, making it capable of quickly recognizing intended destination changes as they happen.

In our preliminary experiments, the proposed method significantly outperforms the accuracy of Kitani's approach while reducing the computation time by a factor of 30 using a parallel version of our algorithm.

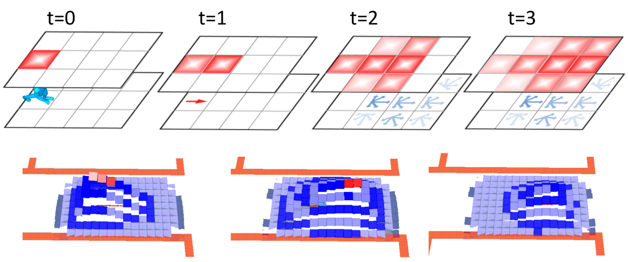


Figure 12. Illustration of (a) the flow-grid model (b) the pheromone model of human motion prediction.

7.4.2. Modeling human-flows from robot(s) perception

Participants: Olivier Simonin, Jacques Saraydaryan, Fabrice Jumel.

To deal with navigation in highly populated environments, eg. flows of humans, we started to investigate the problem of mapping these flows. The challenge is to build such an information from robots perception while they move autonomously to perform their tasks. We also work on predicting humans location from the perceptions and the constructed flow-grid. This led us to define two models : i) a flow-grid mapping computing in each cell the probability to move in each of the k possible directions (illustrated in figure 12 .a), ii) a pheromone-based model allowing to compute the current possible location of humans (flows), see figure 12 .b. We are currently measuring the efficiency of the proposed mapping compared to existing models (which do not model directions). First results will be submitted soon (to IROS 2016).

7.5. Multi-robot Motion-planning in dynamic environments

7.5.1. Benchmarking and extension of multi-robot strategies

7.5.1.1. Exploration of unknown and populated environments

Participants: Olivier Simonin, Nassim Kaldé (phd. Student, Larsen Inria Nancy), François Charpillet (inria Larsen, Nancy), Jan Faigl (ctu, Czech University Of Prague).

Exploration of unknown environment with a group of mobile robots consists mainly to compute a strategy that allows to visit efficiently the area while considering different constraints. These constraints can be trajectory coordination (between robots), presence of humans and limits on time, communication, and computational ressources allowed to robots. The exploration problem is related with mapping, surveillance (eg. patrolling) problems. In this context, O. Simonin and P. Lucidarme (University of Angers) published a general article on multi-robot mapping in the magazine Techniques de l'Ingénieur [5] (2015).

Study of frontier-based strategies

In this context, frontier-based approaches looks for an efficient allocation of the navigational goals which must be situated between the known and unknown areas (the frontiers). Goal candidate locations are repeatedly determined during the exploration. Then, the assignment of the candidates to the robots is solved as the task-allocation problem. A more frequent decision-making may improve performance of the exploration, but in a practical deployment of the exploration strategies, the frequency depends on the computational complexity of the task-allocation algorithm and available computational resources. Therefore, we proposed an evaluation framework to study exploration strategies independently on the available computational resources and we reported a comparison of the selected task-allocation algorithms deployed in multi-robot exploration [30]. This work is supported by the French-Czech PHC "Murotex".

Exploration in populated environments

In the context of the Phd of Nassim Kaldé, co-supervized by F. Chapillet (Inria Nancy, Larsen) and O. Simonin (Chroma), we study exploration in populated environments, in which pedestrian flows can severely impact performances. However, humans have adaptive skills for taking advantage of these flows while moving. Therefore, in order to exploit these human abilities, we propose a novel exploration strategy that explicitly allows for human-robot interactions. Our model for exploration in populated environments combines the classical frontier-based strategy with our interactive approach. For this purpose, we proposed an interaction model where robots can locally choose a human guide to follow and define a parametric heuristic to balance interaction and frontier assignments. This model is introduced in publication [3], where we evaluate to which extent human presence impacts the exploration model in terms of coverage ratio, travelled distance and elapsed time to completion. A simulator, based on V-REP and illustrated in figure 13 .a, has been developped to conduce the experimental measures.

7.5.1.2. Patrolling static and dynamic environments

Participants: Olivier Simonin, Jacques Saraydaryan, Fabrice Jumel, Mihai Popescu, Herve Rivano (inria Urbanet).

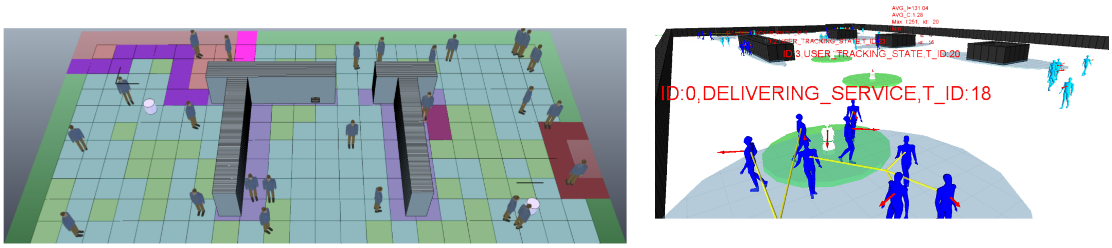


Figure 13. (a) Simulator to study exploration in populated environment based on V-REP (b) Simulator for dynamic patroling of people based on PedSim.

Patrolling moving people

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 re-defined this problem as a dynamic patrolling task, that we called the robot-waiters problem, where robots have to regularly visit all the moving persons. In the publication [11], 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 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 developped a simulator combining a pedestrian model (PedSim) and a robotic model, illustrated in figure 13 .b. Experimental results show the efficiency of the specific new approaches over standard approaches. We also analysed the influence of the number of robots on the performances, for each approach.

We are currently developping new algorithms using the mapping and prediction of human flows based on the work presented in section 7.4.2 .

Patrolling WSN

In the multi-robot patrolling context, we investigated the problem of visiting regularly a set of fixed sensors by computing single-cycles on the graph formed by the WSN (Wireless sensors network). We set this problem as a graph covering with bounded hamiltonian cycles (in the M2R internship of Mihai-Ioan Popescu, now continuing as PhD student in Chroma). After giving insights of NP-hardness, we proposed a generic heuristic algorithm for solving the GCBHC. It works in two steps: the first one partitions the vertices, the second one computes hamiltonian cycles on each partition. We adapted the classic Multilevel Subgraph Partitioning algorithm to the specific requirements yielded by the networking metrics. To avoid the high complexity of this algorithm, we proposed another heuristic which exploits the geometric structure of the graph, the North-Eastern Neighbour heuristic. We implemented two classic hamiltonian cycle heuristics, one is based on Minimum Spanning Trees computations and the other on Christofides algorithm. Comparisons on randmly-generated graphs showed that the Christofides algorithm computes shorter cycles. An article presenting this work has been written and will be submitted soon.

7.5.2. Anytime algorithms for multi-robot cooperation

7.5.2.1. Observation of complex scenes

Participants: Olivier Simonin, Jilles Dibangoye, Laetitia Matignon (liris), Christian Wolf (liris), Jonathan Cohen (internship), Stefan Chitic.

Solving complex tasks with a fleet of robots requires to develop generic strategies that can decide in real time (or time-bounded) efficient and cooperative actions. This is particularly challenging in complex real environments. To this end, we explore anytime algorihms and adaptive/learning techniques.

The INSA BQR project "Crome" [0], led by O. Simonin, motivated the exploration of the joint-observation of complex (dynamic) scenes by a fleet of mobile robots. In our current work, the considered scenes are defined as a sequence of activities, performed by a person in a same place. Then, mobile robots have to cooperate to find a spatial configuration around the scene that maximizes the joint observation of the human pose skeleton. It is assumed that the robots can communicate but have no map of the environment and no external localisation.

To attack the problem, 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. 14 .a). This model is combined with an incremental mapping of the environment in order to limit the complexity of the exploration state space. We have also defined the marginal contribution of each robot observation, to facilitate stability in the search, while the exploration is guided by a meta-heuristics. We developped a simulator (fig. 14 .b) that allows to compare the variants of the approach and to show its features such as adaptation to the dynamic of the scene and robustness to the noise in the observations. Preliminary results have been presented in [8].

We have also developped an experimental framework, using Turtlebot2 robots, presented in figure 14 .c. Experimental measures and validation are in progress.

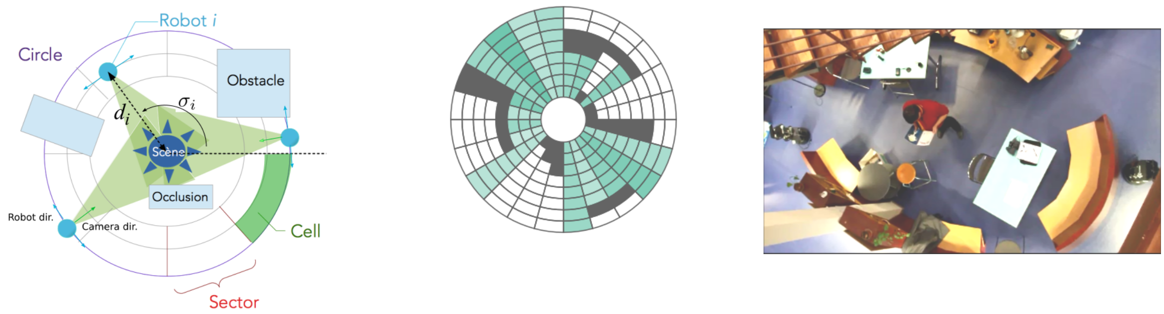


Figure 14. Illustrations (a) Concentric navigation model, (b) Simulator and (c) experimental setup with Turtlebot 2.

7.5.2.2. Middleware for multi-robot systems deployment

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 this context, we addressed 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 presented to the national conference CAR'2015 and will appear in the international conference ICAART 2016 (accepted).

[0] Coordination d'une flottille de robots mobiles pour l'analyse multi-vue de scènes complexes

7.5.3. Sequential decision-making under uncertainty

Sequential decision-making under uncertainty is a core area of artificial intelligence, optimization, operations research, machine learning, and robotics. It involves one or multiple decision makers (or agents or robots) reasoning about the course of actions to achieve collective or self-interested goals while accounting both for the outcomes of current decisions and for future decision-making opportunities. Markov models (e.g., Markov decision processes and Markov games) have emerged as normative frameworks for optimizing decision under uncertainty. These models encompass a wide range of real-world applications: controlling intelligent vehicles; optimizing the production and distribution of energy resources; protecting endangered species; making telecommunication protocols faster and safer; monitoring and assisting elderly patients at home; designing robotic exploration technologies for search and rescue; but also many other applications. Decentralized partially observable Markov decision processes have emerged as the fundamental model to address multiple decision makers' decision-theoretic planning and learning problems. In that direction, we investigate generic, highly scalable and adaptable planning and learning algorithms to apply eventually in multi-robot planning tasks.

7.5.3.1. Structural results for cooperative decentralized control problems

Participants: Jilles S. Dibangoye, Olivier Simonin, Olivier Buffet (inria Nancy, Ex Maia Team), Mamoun Idrissi (internship, Insa de Lyon).

The intractability in cooperative, decentralized control problems is mainly due to prohibitive memory requirements in both optimal policies and value functions. The complexity analysis has emerged as the standard method to estimating the memory needed for solving a given computational problem, but complexity results may be somewhat limited. Our work [13] introduces a general methodology, called the structural analysis, for the design of optimality-preserving concise policies and value functions, which will eventually lead to the development of efficient theory and algorithms. For the first time, we showed that memory requirements for policies and value functions may be asymmetric, resulting in cooperative, decentralized control problems with exponential reductions in memory requirements. To apply this theoretical in robotics, we investigate during M. Idrissi's internship the robotic coverage of unknown areas.

7.5.3.2. State-of-the-art algorithms for optimally solving Dec-POMDPs

Participants: Jilles S. Dibangoye, Christopher Amato (univ. New Hampshire), Olivier Buffet (inria Nancy, Ex Maia Team), François Charpillet (inria Nancy, Larsen Team), Martin Pugnet (master Student, U. Claude Bernard Lyon).

Decentralized partially observable Markov decision processes (Dec-POMDPs) provide a general model for decision-making under uncertainty in cooperative decentralized settings but are difficult to solve optimally (NEXP-Complete). As a new way of solving these problems, we introduced 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, in [14], 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. However, scalability remains limited when the number of agents or problem variables becomes large. To overcome this limitation, we show that, under certain separability conditions of the optimal value function, the scalability of this approach can increase considerably. This separability is present when there is the locality of interaction between agents, which can be exploited to improve performance. 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 confirming that our approach provides significant scalability improvements compared to the state of the art. We push even further the envelope, during Martin's internship, assuming we only have access to an incomplete model of the world. This more realistic assumption that would ease application to robotics leads us directly to learning algorithms inspired from FB-HSVI.

7.5.3.3. Distributed projected gradient-descent algorithm applied to smart grids

Participants: Jilles S. Dibangoye, Arnaud Doniec (uria – Ecole Des Mines de Douai, France), H. Fakham, F. Colas And X. Guillaud (l2ep – Arts Et Métiers Paristech, France).

In a smart grid context, the increasing penetration of embedded generation units leads to a greater complexity in the management of production units. In this work, we focus on the impact of the introduction of decentralized generation for the unit commitment (UC) problem. Unit commitment problems consist in finding the optimal schedules and amounts of power to be generated by a set of generating units in response to an electricity demand forecast. While this problem has received a significant amount of attention, classical approaches assume that these problems are centralized and deterministic. However, these two assumptions are not realistic in a smart grid context. Indeed, finding the optimal schedules and amounts of power to be generated by multiple distributed generator units is not trivial since it requires to deal with distributed computation, privacy, stochastic planning, etc. Our contribution focuses on smart grid scenarios where the main source of complexity comes from the proliferation of distributed generating units. In solving this issue, we consider distributed stochastic unit commitment problems. In [2], we introduce a novel distributed gradient descent algorithm which allows us to circumvent classical assumptions. This algorithm is evaluated through a set of experiments on real-time power grid simulator.

DEFROST Team

7. New Results

7.1. Aggregate Constraints for Virtual Manipulation with Soft Fingers

In this work, we propose a new formulation of contact and friction laws, in the context of virtual graping. The work allows to reduce the number of contact and friction constraints, using volume interpenetration measure, instead of interpenetration distance. The work has been conducted in collaboration with Antony Talvas and Maud Machal (Inria Hybrid Team, Rennes) and Miguel Otaduy (URJC Madrid). It has been presented at the conference IEEE VR and published in the journal TVCG [5].

7.2. Haptic Rendering of Hyperelastic Models with Friction

We have reached an important milestone with this work: we have merge two important research tracks of these last years: On one hand, haptic rendering of friction contact between deformable objects ; on the ohter hand, real-time simulation of hyperelastic objects (particularly to simulate soft-tissues). This work has been conducted in collaboration with Hadrien Courtecuisse (Inria team Mimesis) and Hervé Delingette (Inria team Asclepios) [6]

7.3. Augmentation of Elastic Surfaces with Self-Occlusion Handling

In this work, we propose to recover the 3D shape and to augment elastic objects with self-occlusions handling, using only single view images. Shape recovery from a monocular video sequence is an underconstrained problem and many approaches have been proposed to enforce constraints and resolve the ambiguities. State-of-the art solutions enforce smoothness or geometric constraints, consider specific deformation properties such as inextensibility or resort to shading constraints. We propose a real-time method that uses a mechanical model and that is able to handle highly elastic objects. The problem is formulated as an energy minimization problem accounting for a non-linear elastic model constrained by external image points acquired from a monocular camera. This method prevents us from formulating restrictive assumptions and specific constraint terms in the minimization. In addition, we propose to handle self-occluded regions thanks to the ability of mechanical models to provide appropriate predictions of the shape. This result has been published in the journal TVCG [2] and has been extended to handle cutable objects and has been published as a SIGGRAPH poster [12].

7.4. Real-time control of soft-robots using asynchronous finite element modeling

Finite Element analysis can provide accurate deformable models for soft-robots. However, using such models is very difficult in a real-time system of control. In this study, we introduce a generic solution that enables a high-rate control and that is compatible with strong real-time constraints. From a Finite Element analysis, computed at low rate, an inverse model of the robot outputs the setpoint values for the actuator in order to obtain a desired trajectory. This inverse problem uses a QP (quadratic-programming) algorithm based on the equations set by the Finite Element Method. To improve the update rate performances, we propose an asynchronous simulation framework that provides a better trade-off between the deformation accuracy and the computational burden. Complex computations such as accurate FEM deformations are done at low frequency while the control is performed at high frequency with strong real-time constraints. The two simulation loops (high frequency and low frequency loops) are mechanically coupled in order to guarantee mechanical accuracy of the system over time. Finally, the validity of the multi-rate simulation is discussed based on measurements of the evolution in the QP matrix and an experimental validation is conducted to validate the correctness of the high-rate inverse model on a real robot. [8]

7.5. Domain decomposition approach for FEM quasistatic modeling and control of Continuum Robots with rigid vertebras

This study focuses on a new method dedicated to the modeling and control of Continuum Robots, based on the Finite Element Method (FEM) using quasi-static assumption. The modeling relies on a discretization of the continuum robots using 6 DoFs Frames along the structure of the robot that is compatible with the modeling of a sequence of rigid vertebras. When the robot's structure relies on rods with constant sections, internal forces are computed with beam elements, placed between two adjacent frames, that applies forces and torques. In the opposite, when the robot is composed of a complex shape deformable backbone separated by the rigid vertebras, a domain decomposition strategy is used to obtain an equivalent stiffness between two vertebras using volumetric FEM. In both cases, for solving the whole robot model and inverting it in real-time, the numerical method takes advantage of the serial nature of continuum robots, using a Block-Tri-Diagonal solver. The factor of improvement in the computation time reaches several order of magnitude compared to a classical FEM model, while keeping a good precision. The method has also been implemented and tested on a real pneumatic CBHA trunk designed by Festo Robotics and some complementarity examples have been generated numerically.[10]

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, Adrien Baranes, William Schueller, Sebastien Forestier, Nabil Daddaouda, Nicholas Foley.

This project involves a collaboration between the Flowers team and the Cognitive Neuroscience Lab of J. Gottlieb at Columbia Univ. (NY, US) on the understanding and modeling of mechanisms of curiosity and attention that until now have been little explored in neuroscience, computer science and cognitive robotics. It is organized around the study of the hypothesis that information gain could generate intrinsic reward in the brain (living or artificial), controlling attention and exploration independently from material rewards. The project combines expertise about attention and exploration in the brain and a strong methodological framework for conducting experimentations with monkeys and humans (Gottlieb's lab) together with cognitive modeling of curiosity 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 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 (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 [105], [122], [130]. Therefore we hypothesize that, rather than using a single optimization process as it has been the case in most previous theoretical work [90], 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[16] [74], [112]. 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 [103], [129], and rats show individual variation in learning styles and novelty seeking behaviors [88], 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 [127], [136], [70], [140]. 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 [103], [100] 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

During the first period of the associated team (2013-2015), we layed the operational foundations of the collaboration resulting in several milestone joint journal articles [110], [90], [84][27], new experimental paradigms for the study of curiosity, and organized a major scientific event: the first international interdisciplinary symposium on information seeking, curiosity and attention (web: https://openlab-flowers.inria.fr/t/first-interdisciplinary-symposium-on-information-seeking-curiosity-and-attention/21).

In particular, new results in 2015 include:

7.1.1.4. Eye movements reveal epistemic curiosity in human observers

Saccadic (rapid) eye movements are primary means by which humans and non-human primates sample visual information. However, while saccadic decisions are intensively investigated in instrumental contexts where saccades guide subsequent actions, it is largely unknown how they may be influenced by curiosity – the intrinsic desire to learn. While saccades are sensitive to visual novelty and visual surprise, no study has examined their relation to epistemic curiosity – interest in symbolic, semantic information. To investigate this question, we tracked the eye movements of human observers while they read trivia questions and, after a brief delay, were visually given the answer. We showed that higher curiosity was associated with earlier anticipatory orienting of gaze toward the answer location without changes in other metrics of saccades or fixations, and that these influences were distinct from those produced by variations in confidence and surprise. Across subjects, the enhancement of anticipatory gaze was correlated with measures of trait curiosity from personality questionnaires. Finally, a machine learning algorithm could predict curiosity in a cross-subject manner, relying primarily on statistical features of the gaze position before the answer onset and independently of covariations in confidence or surprise, suggesting potential practical applications for educational technologies, recommender systems and research in cognitive sciences. We published these results in [27], providing full access to the annotated database allowing readers to reproduce the results. Epistemic curiosity produces specific effects on oculomotor anticipation that can be used to read out curiosity states.

7.1.1.5. Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates

Intelligent animals have a high degree of curiosity – the intrinsic desire to know – but the mechanisms of curiosity are poorly understood. A key open question pertains to the internal valuation systems that drive curiosity. What are the cognitive and emotional factors that motivate animals to seek information when this is not reinforced by instrumental rewards? 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. [84]

7.1.2. Computational Models Of Speech Development: the Roles of Active Learning, Curiosity and Self-Organization

Participants: Pierre-Yves Oudeyer [correspondant], Clement Moulin-Frier, Sébastien Forestier.

7.1.2.1. Special issue on the cognitive nature of speech sounds

Together with Jean-Luc Schwartz and Kenneth de Jong, Flowers members Clément Moulin-Frier and Pierre-Yves Oudeyer guest-edited a milestone special issue of the Journal of Phonetics focusing on theories of the cognitive nature of speech sounds, and with a special emphasis on presenting and analyzing a rich series of computational models of speech evolution and acquisition developped in the past years internationally, including models developped by the guest-editors. The editorial of this special issue was published in [35] and the special issue is accessible at: http://www.sciencedirect.com/science/journal/00954470/53.

7.1.2.2. The COSMO model: A Bayesian modeling framework for studying speech communication and the emergence of phonological systems

(Note: this model was developped while C. Moulin-Frier was at GIPSA Lab, and writing was partly achieved while he was at Inria). While the origin of language remains a somewhat mysterious process, understanding how human language takes specific forms appears to be accessible by the experimental method. Languages, despite their wide variety, display obvious regularities. In this paper, we attempt to derive some properties of phonological systems (the sound systems for human languages) from speech communication principles. The article [33] introduces a model of the cognitive architecture of a communicating agent, called COSMO (for "Communicating about Objects using Sensory–Motor Operations') that allows a probabilistic expression of the main theoretical trends found in the speech production and perception literature. This enables a computational comparison of these theoretical trends, which helps us to identify the conditions that favor the emergence of linguistic codes. It presents realistic simulations of phonological system emergence showing that COSMO is able to predict the main regularities in vowel, stop consonant and syllable systems in human languages.

7.1.2.3. The role of self-organization, motivation and curiosity in speech development and evolution

In the article [34], Oudeyer discusses open scientific challenges for understanding development and evolution of speech forms. Based on the analysis of mathematical models of the origins of speech forms, with a focus on their assumptions, the article studies the fundamental question of how speech can be formed out of non-speech, at both developmental and evolutionary scales. In particular, it emphasizes the importance of embodied self-organization, as well as the role of mechanisms of motivation and active curiosity-driven exploration in speech formation. Finally, it discusses an evolutionary-developmental perspective of the origins of speech.

7.1.2.4. Robotic models of the joint development of speech and tool use

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 [90]. 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 [124].

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 [73]). On the other hand, other algorithmic models have shown how several social learning mechanisms could allow robots to acquire elements of speech and language [79], 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 [125].

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 [113]. 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 [120].

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 [79]. 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 [120], [73], [118], [113]. 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.

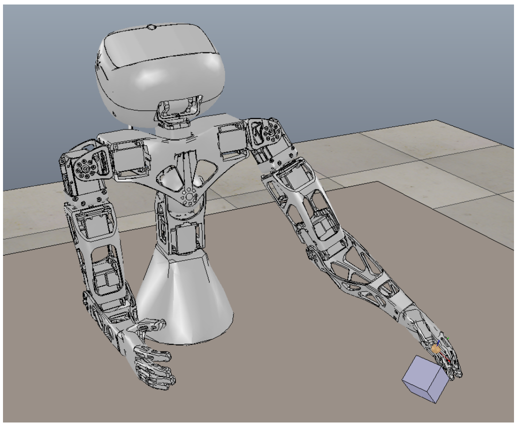
We presented preliminary results on that question in a poster session [89] of the ICDL/Epirob conference in Providence, RI, USA in August 2015. In this work, we rely more specifically on the R-IAC and SAGG-RIAC series of architectures developed in the Flowers team and we develop different ways to extend those architectures to the learning of several task spaces that can be explored in a hierarchical manner. We describe an interactive task to evaluate different hierarchical learning mechanisms, where a robot has to explore its motor space in order to push an object to different locations. The task can be decomposed into two subtasks where the robot can first explore how to make movements with its hand and then integrate this skill to explore the task of pushing an object.

In the Simplest First strategy, the agent explores successively but with a fixed curriculum the different tasks to learn in the good order: from the simplest one (learning hand movements given motor parameters) to the more complex one (pushing a block with hand movements) that need knowledge about the simpler task to be learned.

In the Top-Down Guidance strategy, the module learning the more complex task (pushing a block with hand movements) gives goals (hand movements) to be reached by the lower-level module (learning hand movements given motor parameters) that will explore for a while to reach that goal before switching to a new given goal.

We also compare our architectures to the control ones where the robot learns directly the not decomposed task, with a competence-based intrinsic motivation (goal babbling) or a fully random motor babbling.

The results show a better exploration for the Top-Down Guidance than the Simplest First hierarchical exploration strategy, and that learning intermediate representations is beneficial in this setup.



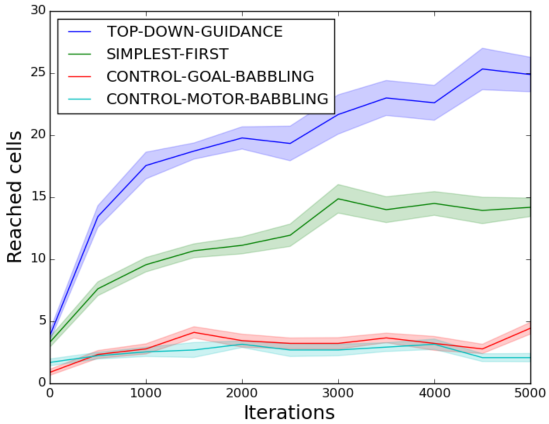


Figure 8. Left: Poppy Torso in the V-REP simulator pushing a block. Right: Exploration results of the different strategies.

7.1.3. Learning in Adult-Child and Human-Robot Interaction

Participants: Anna-Lisa Vollmer [correspondant], Pierre-Yves Oudeyer.

Learning in Adult-Child and Human-Robot Interaction

7.1.3.1. The Change of 'Motionese' Parameters Depending on Children's Age.

Two adult-child interaction studies were analyzed with the focus on the parental teaching behavior, in particular on *motionese* parameters (modifications of child-directed movement). In the first cross-sectional study, parental action demonstrations to three groups of 8–11, 12– 23 and 24–30 month-olds ($N = 84$ parents) were investigated. The youngest group of participants was investigated longitudinally in the second study ($N = 18$ parents). Together the results suggest that some motionese parameters (motion pauses, velocity, acceleration) persist over different ages while other parameters (action length, roundness and pace) occur predominantly in the younger group and seem to be primarily used to attract infants' attention on the basis of movement. In contrast, parameters appearing to be more in charge of structuring the action by organizing it in motion pauses seem to persist. We discuss the results in terms of facilitative vs. pedagogical learning in a paper currently under review for the Journal of Experimental Child Psychology.

7.1.3.2. An Alternative to Mapping a Word onto a Concept in Language Acquisition: Pragmatic Frames

According to the mapping metaphor, for a child to learn a word, s/he has to map a word onto a concept of an object/event. We are not convinced that associations can explain word learning, because even though children's attention is on the objects, they do not necessarily remember the connection of the word with the referent. In this theoretical paper, we propose an alternative to the mapping process that is classically assumed as a mechanism for word learning. Our main point holds that word learning is a task, in which children accomplish a goal in cooperation with a partner. In our approach, we follow Bruner's (1983) idea and further specify pragmatic frames as learning units that drive language acquisition and cognitive development. These units consist of a sequence of language and actions that are co-constructed with a partner to achieve a joint goal. We elaborate on this alternative, offer some initial parametrizations of the concept and embed it in the current language learning approaches in a paper currently under review for Frontiers in Psychology, section Cognitive Science.

7.1.3.3. Meta-Analysis of Pragmatic Frames in Human-Robot Interaction for Learning and Teaching: State-of-the-Art and Perspectives

One of the big challenges in robotics today is to learn from inexperienced human users. Despite tremendous research efforts and advances in human-robot interaction (HRI) and robot learning in the past decades, learning interactions with robots remain brittle and rigidly organized, and often are limited to learning only one single task. In this work, we applied the concept of pragmatic frames known from developmental research in humans in a meta-analysis of current approaches on robot learning. This concept offers a new research perspective in HRI as multiple flexible interaction protocols can be used and learned to teach/learn multiple kinds of skills in long-term recurring social interaction. This perspective, thus, emphasizes teaching as a collaborative achievement of teacher and learner. Our meta-analysis focuses on robot learning from a human teacher with respect to the pragmatic frames they (implicitly) use. We show that while the current approaches offer a variety of different learning and teaching behavior, they all employ highly pre-structured, hard-coded pragmatic frames. Compared to natural human-human interaction, interactions are lacking flexibility and expressiveness, and mostly are hardly viable for being realized with truly naive and uninstructed users. We elaborated an outlook on the future research direction with its relevant key challenges that need to be solved for leveraging pragmatic frames for robot learning. These results have been submitted to the Frontiers in Neurorobotics Journal.

7.1.3.4. Alignment to the Actions of a Robot

Alignment is a phenomenon observed in human conversation: Dialog partners' behavior converges in many respects. Such alignment has been proposed to be automatic and the basis for communicating successfully. Recent research on human-computer dialog promotes a mediated communicative design account of alignment according to which the extent of alignment is influenced by interlocutors' beliefs about each other. Our work aims at adding to these findings in two ways. a) Our work investigates alignment of manual actions, instead of lexical choice. b) Participants interact with the iCub humanoid robot, instead of an artificial computer dialog system. Our results confirm that alignment also takes place in the domain of actions. We were not able to replicate the results of the original study in general in this setting, but in accordance with its findings, participants with a high questionnaire score for emotional stability and participants who are familiar with robots align their actions more to a robot they believe to be basic than to one they believe to be advanced. Regarding alignment over the course of an interaction, the extent of alignment seems to remain constant, when participants believe the robot to be advanced, but it increases over time, when participants believe the robot to be a basic version. These results were published in [38].

7.1.4. Models of Multimodal Concept Acquisition with Non-Negative Matrix Factorization

Participants: Pierre-Yves Oudeyer, Olivier Mangin [correspondant], David Filliat, Louis Ten Bosch.

In the article [32] we introduced MCA-NMF, a computational model of the acquisition of multi-modal concepts by an agent grounded in its environment. More precisely our model finds patterns in multimodal sensor input that characterize associations across modalities (speech utterances, images and motion). We propose this computational model as an answer to the question of how some class of concepts can be learnt. In addition, the model provides a way of defining such a class of plausibly learnable concepts. We detail why the multimodal nature of perception is essential to reduce the ambiguity of learnt concepts as well as to communicate about them through speech. We then present a set of experiments that demonstrate the learning of such concepts from real non-symbolic data consisting of speech sounds, images, and motions. Finally we consider structure in perceptual signals and demonstrate that a detailed knowledge of this structure, named compositional understanding can emerge from, instead of being a prerequisite of, global understanding. An open-source implementation of the MCA-NMF learner as well as scripts and associated experimental data to reproduce the experiments are publicly available.

The python code and datasets allowing to reproduce these experiments and results are available at: https://github.com/omangin/multimodal.

7.1.5. Models of Self-organization of lexical conventions: the role of Active Learning in Naming Games

Participants: William Schueller [correspondant], Pierre-Yves Oudeyer.

Our work focuses on the Naming Games framework [135], meant to simulate lexicon evolution in a population from interactions at the individual level. A quite diverse subset of the possible scenarios and algorithms has already been studied, and those do lead to the self-organization of a shared lexicon (understood as associations between meanings and words). However, high values for some parameters (population size, number of possible words and/or meanings that can be refered to) can lead to really slow dynamics. Following the introductory work done in [119], we introduced a new measure of vocabulary evolution based on information theory, as well as various active learning mechanisms in the Naming Games framework allowing the agents to choose what they talk about according to their past. We showed that it improves convergence dynamics in the studied scenarios and parameter ranges. Active learning mechanisms use the confidence an agent has on its own vocabulary (is it already widely used in the population or not?) to choose between exploring new associations (growing vocabulary) or strengthening already existing ones (spreading its own associations to other agents). This was presented at the ICDL/Epirob conference in Providence, RI, USA in August 2015 [59].

A follow-up to this work consisted of changing slightly the base algorithms, allowing agents to select what they want the others to talk about instead of selecting what they would talk about (hearer's choice scenario, the original one being speaker's choice scenario). In the class of algorithms used, with active learning, it leads to faster convergence, with increased robustness to change in parameter values.

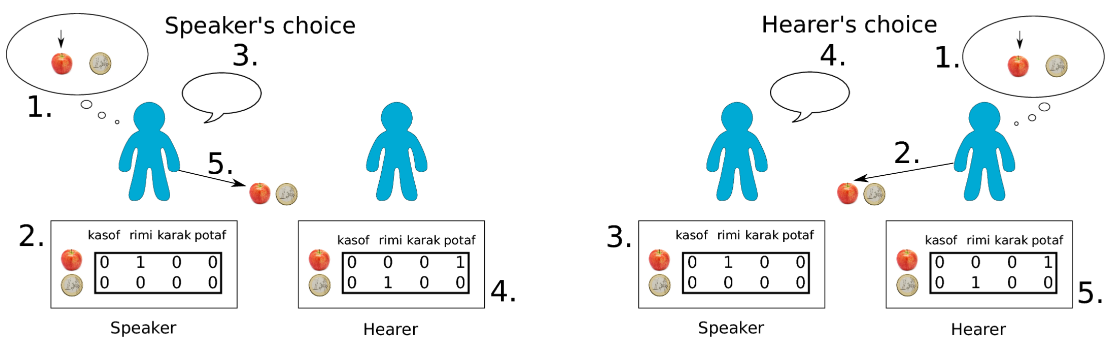


Figure 9. Interaction processes in both active scenarios considered in our work. Beforehand, two individuals have been randomly selected among a population, designated as speaker (S) and hearer (H). Speaker's choice: 1. S chooses a topic, 2. S checks its vocabulary to find or invent an associated word, 3. S utters the word, 4. H guesses the intended meaning, 5. S indicates the intended meaning. Hearer's choice: 1. H chooses a topic, 2. H indicates the intended meaning, 3. S checks its vocabulary to find or invent an associated word, 4. S utters the word, 5. H checks its vocabulary for a meaning associated to the uttered word. In both cases, if all meanings match, the interaction is considered a success, otherwise a failure. After the process, both agents can update their vocabularies to take the interaction into account.

All the simulations can be easily rerun using the provided code and explanatory notebooks on https://github.com/flowersteam/naminggamesal.

7.2. Life-Long Robot Learning And Development Of Motor And Social Skills

7.2.1. Uncalibrated BCI

Participants: Manuel Lopes [correspondant], Pierre-Yves Oudeyer, Jonathan Grizou, Inaki Iturrate, Luis Montesano.

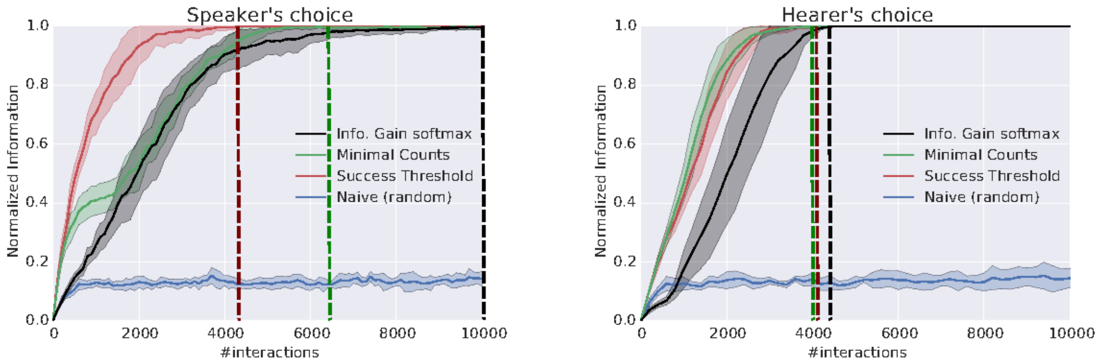


Figure 10. Strategy comparisons in both active scenarios. The measure indicates level of convergence towards a population-wide shared vocabulary (a value of 1 means every agent has the exact same vocabulary). Naive (random, no active learning) strategy converges slowly. Hearer's choice policy is more efficient for all active learning strategies. Last 5% of information are acquired slower when the speaker is choosing. Vertical lines show full convergence time for each strategy. (number of meanings = number of words = population size =20, averaged over 8 trials)

We developed an new approach for self-calibration BCI for reaching tasks using error-related potentials. The proposed method exploits task constraints to simultaneously calibrate the decoder and control the device, by using a robust likelihood function and an ad-hoc planner to cope with the large uncertainty resulting from the unknown task and decoder. The method has been evaluated in closed-loop online experiments with 8 users using a previously proposed BCI protocol for reaching tasks over a grid. The results show that it is possible to have a usable BCI control from the beginning of the experiment without any prior calibration. Furthermore, comparisons with simulations and previous results obtained using standard calibration hint that both the quality of recorded signals and the performance of the system were comparable to those obtained with a standard calibration approach. [30]

7.2.2. Learning from Demonstration

Participants: Manuel Lopes, Thibaut Munzer [correspondant], Marc Toussaint, Li Wang Wu, Yoan Mollard, Andrea Baisero, Bilal Piot, Matthieu Geist, Olivier Pietquin.

Learning from Demonstration

7.2.2.1. Relational Activity Processes for Modeling Concurrent Cooperation

In multi-agent domains, human-robot collaboration 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 multiple 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 Monte-Carlo planning and learning from demonstration, to concurrent cooperation domains. We formally compare the formulation to previous propositional models of concurrent decision making and demonstrate the planning and learning from demonstration methods on a real-world human-robot assembly task.

7.2.2.2. Interactive Learning

In paper [56] we consider that robot programming can be made more efficient, precise and intuitive if we leverage the advantages of complementary approaches such as learning from demonstration, learning from feedback and knowledge transfer. We designed a system that, starting from low-level demonstrations of assembly tasks, is able to extract a high-level relational plan of the task. A graphical user interface (GUI) allows then the user to iteratively correct the acquired knowledge by refining high-level plans, and low-level geometrical knowledge of the task. A final process allows to reuse high-level task knowledge for similar tasks in a transfer learning fashion. We conducted a user study with 14 participants asked to program assembly tasks of small furniture (chair and bench) to validate this approach. The results showed that this combination of approaches leads to a faster programming phase, more precise than just demonstrations, and more intuitive than just through a GUI.

7.2.2.3. Inverse Reinforcement Learning in Relational Domains

We introduced a first approach to the Inverse Reinforcement Learning (IRL) problem in relational domains. IRL has been used to recover a more compact representation of the expert policy leading to better generalize among different contexts. Relational learning allows one to represent problems with a varying number of objects (potentially infinite), thus providing more generalizable representations of problems and skills. We show how these different formalisms can be combined by modifying an IRL algorithm (Cascaded Supervised IRL) such that it handles relational domains. Our results indicate that we can recover rewards from expert data using only partial knowledge about the dynamics of the environment. We evaluate our algorithm in several tasks and study the impact of several experimental conditions such as: the number of demonstrations, knowledge about the dynamics, transfer among varying dimensions of a problem, and changing dynamics. This was published in [49]

7.2.3. A Unified Model for Regression

Regression is the process of learning relationships between inputs and continuous outputs from example data, which enables predictions for novel inputs. Regression lies at the heart of imitation learning, and value function approximation for reinforcement learning. In [37], we provide a novel perspective on regression, by distinguishing rigoroulsy between the models and representations assumed in regression, and the algorithms used to train the parameters of these models. A rather surprising insight is that many regression algorithms [0] use very similar models; in fact, we show that the algorithm-specific models are *all* special cases of a “unified model”. This perspective clearly seperates between representations and algorithms, and allows for a modular exchange between them, for instance in the context of evolutionary optimization.

7.2.4. Multiple Virtual Guides

In co-manipulation, humans and robots solve manipulation tasks together. Virtual guides are important tools for co-manipulation, as they constrain the movement of the robot to avoid undesirable effects, such as collisions with the environment. Defining virtual guides is often a laborious task requiring expert knowledge. This restricts the usefulness of virtual guides in environments where new tasks may need to be solved, or where multiple tasks need to be solved sequentially, but in an unknown order.

To this end, we have proposed a framework for *multiple probabilistic virtual guides*, and demonstrated a concrete implementation of such guides using kinesthetic teaching and Gaussian mixture models [57], [58]. Our approach enables non-expert users to design virtual guides through demonstration. Also, they may demonstrate novel guides, even if already known guides are active. Finally, users are able to intuitively select the appropriate guide from a set of guides through physical interaction with the robot.

7.2.5. Legible Motion

Participants: Manuel Lopes, Baptiste Busch [correspondant], Jonathan Grizou, Freek Stulp.

[0]Locally Weighted Regression, Receptive Field Weighted Regression, Locally Weighted Projection Regression, Gaussian Mixture Regression, Model Trees, Radial Basis Function Networks, Kernel Ridge Regression, Gaussian Process Regression , Support Vector Regression Incr. Random Features Regularized Least Squares, Incr. Sparse Spectrum Gaussian Process Regr., Regression Trees, Extreme Learning Machines.

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. We propose an optimization algorithm, based on policy improvement, that brings out the most legible robot's trajectories during the interaction (cf. Figure 11). The conducted user study highlights that humans become better at predicting sooner the robot's intentions. This leads to faster and more robust overall task completion. This work have been published to IROS 2015[60] and was submitted to the International Journal of Social Robotics under the special issue: Towards a framework for Joint Action.

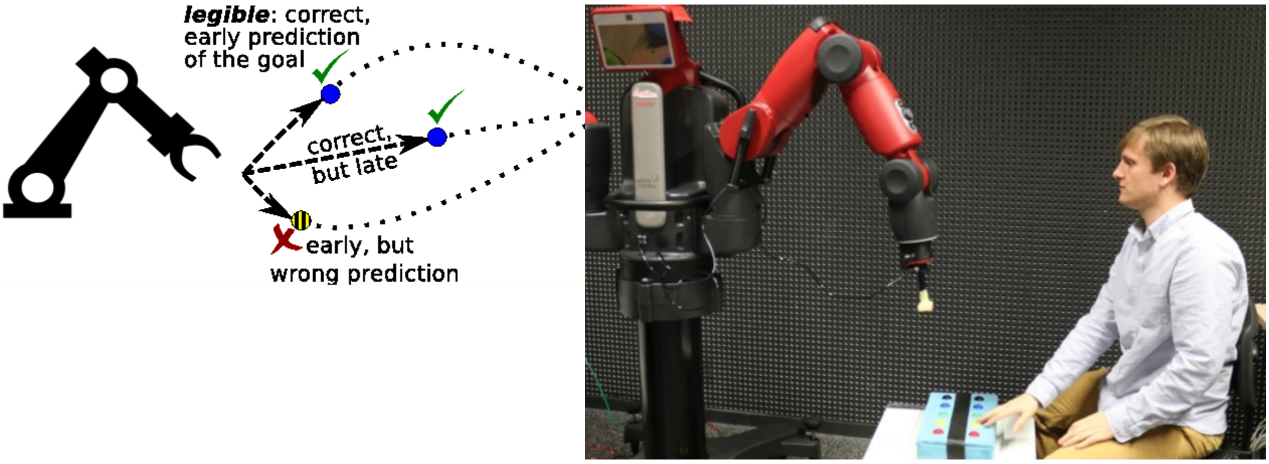


Figure 11. Illustration of the experimental setup. The robots aims for a button and press it. The human predict the target and is instructed to quickly press the button of the same color when sufficiently confident about this prediction.

7.2.6. Demonstrator of human-robot interface for teaching a collaborative task in the context of assistive robotics

Participants: Pierre Rouanet [correspondant], Yoan Mollard, Thibaut Munzer, Baptiste Busch, Manuel Lopes, Pierre-Yves Oudeyer.

In the context of the Roméo 2 project, we have developed a demonstrator of a human-robot interface designed for non-expert users. It allows them to teach a new collaborative task to a robot through simple and intuitive interactions. It is based on the approach of inverse reinforcement learning in relational domains described above.

The context of the demonstrator is assistive robotics where typically an elderly person wants to teach a robot (we use Baxter in this case) how it can help him to prepare a meal. For instance, the user will show the robot that first he wants the robot to hold its bowl and that he stirs it. Then, the robot should put the bowl on a plate. Then, the user will teach the robot that he wants the robot to grab a glass and put it on the right of the bowl...

7.2.7. Diversity-driven curiosity-driven learning and transfer learning

Participants: Fabien Benureau [correspondant], Pierre-Yves Oudeyer.

7.2.7.1. Diversity-driven selection of exploration strategies in multi-armed bandits

In [40], we considered a scenario where an agent has multiple available strategies to explore an unknown environment. For each new interaction with the environment, the agent must select which exploration strategy to use. We provide a new strategy-agnostic method that treat the situation as a Multi- Armed Bandits problem where the reward signal is the diversity of effects that each strategy produces. We test the method empirically on a simulated planar robotic arm, and establish that the method is both able discriminate between strategies of dissimilar quality, even when the differences are tenuous, and that the resulting performance is competitive with the best fixed mixture of strategies.

7.2.7.2. Behavioral Diversity Generation in Autonomous Exploration Through Reuse of Past Experience

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 unknwon 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 the most crucial, i.e. new and unknown ones, is far from trivial. In particular in large and redundant sensorimotor spaces, some effects can typically only be produced by a few number of specific motor commands. We introduced a method to create behavioral diversity by re-enacting past experiences, along with a measure that quantifies this diversity. We showed that our method is robust to morphological and representation changes, that it can learn how to interact with an object by reusing experience from another and how scaffolding behaviors can emerge by simply switching the attention of the robot to different parts of the environment. Finally, we showed that the method can robustly use simulated experiences and crude cognitive models to provide behavioural diversity in the real world. This result are under review.

7.3. Autonomous And Social Perceptual Learning

Participants: David Filliat [correspondant], Freek Stulp, Celine Craye, Yuxin Chen, Clement Masson, Adrien Matricon.

7.3.1. Incremental Learning of Object-Based Visual Saliency

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 [43], [42]

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. This model uses two different strategy: a statistical filtering to remove noise in the speech part and the Non Negative Matrix Factorization algorithm to discover word-meaning in the visual domain. We present experiments on learning object names and color names showing the performance of the model in real interactions with humans, dealing in particular with strong noise in the speech recognition. We published these results in a conference paper [41]

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, 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. To tackle both issue, we are developing a convolutional architectures inspired form Convolutional Neural Networks (CNNs) that provide different modelisations for different parts of the image, which might be usefull to model combinations of transformations. Our Convolutional Gated Auto-Encoder is designed to perform generic feature learning in an unsupervised way (while most CNNs are trained in a supervised fasion) and we are currently testing it on realistic image sequences. We plan to extend this architecture to find relations between modalities as, for instance, proproceptive information and its evolution could be used to predict the next visual features. Similarly, proprioceptive information could be used as a supervising signal to learn visual features.

7.3.4. Learning models by minimizing complexity

In machine learning, it is commonly assumed that simpler models have better chances at generalizing to new, unseen data. Following this principle, we developped an algorithm relying on minimization of a given complexity measure to build a collection of models which jointly explain the observation of the training datapoints. The resulting collection is composed of as few models as possible, each using as few dimensions as possible and each as regular as possible. As of now, each model is a multivariate polynomial, with the complexity of a polynomial of degree N in d variables being N*d+1. The complexity of the collection is the sum of the complexity of all its models. The algorithm starts by associating each datapoint to a local model of complexity 1 (degree 0, no variables), then models are iteratively merged into models of higher complexity, as long as those merges don't increase the complexity of the collection and as long as the resulting models stay within a certain distance of their associated datapoints. We applied this algorithm to the problem of inverse dynamics, which we studied in simulation. For a given robot, torques needed to compensate gravity at equilibrium are entirely determined by the values of its joint angles. As it is common that robots actually perform only low-dimensional tasks, and do not explore their full state space during normal operation, we would like the complexity of our models to mirror the structure of the task. When the task was expressed in the joint space, we got satisfying results on that point, and got good predictions for unseen datapoints. When the task was expressend in end-effector position, it turned out to be impossible to learn the underlying manifolds because a given end-effector position can correspond to various joint configurations, and thus to various torques, making it impossible to predict those torques from the end-effector position alone. We are currently working on applying this model to data generated by an exploration algorithm on a robot arm manipulating objects.

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, Joel Ortiz, Mathilde Couraud, 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 2015 two lines of work were achieved, concerning two important scientific challenges, and in the context of a PEPS CNRS project:

First, an experimental setup was designed to allow fast prototyping of 3D printed robotic prostheses (internship of Joel Ortiz). This work was based on the use of the Poppy open-source modular platform, and resulted in a functional prototype. Several video demonstrations are available at: https://forum.poppy-project.org/t/real-time-control-of-a-prosthetic-robotic-arm-poppy-with-muscle-activities/1656.

Second, first versions of co-adaptation algorithms were designed, implemented and tested with human subjects, based on the combination of advanced models of the arm biomechanics and incremental learning algorithms (internship of Mathilde Couraud). An article is under preparation.

7.5. Applications for Educational Technologies

7.5.1. KidLearn

Participants: Manuel Lopes [correspondant], Pierre-Yves Oudeyer, Didier Roy, Benjamin Clement.

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.

Intelligent Tutoring System (ITS) are computer environments designed to guide students in their learning. Through the proposal of different activities, it provides teaching experience, guidance and feedback to improve learning. The FLOWERS team has developed several computational models of artificial curiosity and intrinsic motivation based on research on psychology that might have a great impact for ITS. Results showed that activities with intermediate levels of complexity, neither too easy nor too difficult but just a little more difficult that the current level, provide better teaching experiences. The system is based on the combination of three approaches. First, it leverages Flowers team's recent models of computational models of artificial curiosity and intrinsic motivation based on research in psychology and neuroscience. 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. In 2014, we have run a second pilot experiment in elementary schools of Région Aquitaine, where 7-8 year old kids could learn elements of mathematics thanks to an educational software that presented the right exercises at the right time to maximize learning progress. [29]

7.5.2. Poppy System

Participants: Matthieu Lapeyre [correspondant], Nicolas Rabault, Pierre Rouanet, Pierre-Yves Oudeyer.

In the Poppy project we are working on the Poppy System which is a new modular and open-source robotic architecture. It is design to help people create and build custom robots. It permits, in a similar approach as Lego, building robots or smart objects using standardized elements.

Poppy System is an unified system where each essential robotic components (actuators, sensors, ...) is an independant module, connected with other through standardized interfaces.

- Unified mechanical interfaces which simplifies the assembly process and the design of 3D printable parts.
- Unified communication between elements using the same connector and bus for each module.
- Unified software makes it easy to program each module independantly.

The current Poppy robots (Humanoid, Torso, Ergo) will be updated using this novel architecture.

Our ambition is to create an ecosystem around this system so communities can develop custom modules, following the Poppy System standards, which can be compatible with all other Poppy robots.

7.5.3. Poppy Education

Participants: Pierre-Yves Oudeyer [correspondant], Didier Roy, Théo Segonds, Stéphanie Noirpoudre, Marie Demangeat, Thibault Desprez, Matthieu Lapeyre, Pierre Rouanet, Nicolas Rabault.

Poppy Education aims to create, evaluate and disseminate pedagogical kits “turnkey solutions” complete, 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, Highschools, engineers schools, ...). For secondary education and higher education, scientific literacy centers, Fablabs.

The Poppy robotic platform used in the project is free hardware and software, printed in 3D, and is intended primarily for:

- learning of computer science and robotics,
- introduction to digital manufacturing (3D printing ...)
- initiation to the integration of IT in physical objects in humanoid robotics, mechatronics.
- artistic activities.

Educational sectors covered by the project are mainly: Enseignement d’exploration ICN en seconde, enseignement ISN en terminale S et bientôt en 1ère , filière STI2D, MPS seconde.

Users and their needs are placed at the center of this project. The pedagogical tools of the project are being created directly with them and evaluated in real life by experiments. Poppy Education is based on the robotic platform poppy, from which it is possible to construct different robots, including:

- Poppy Humanoid is a robust and complete robotics platform designed for genuine experiments in the real world and can be adapted to specific user needs.
- Poppy Torso is a variant of Poppy Humanoid. It is a torso humanoid robot that can be easily installed on a table.
- Poppy Ergo Jr is a robotic arm. Solid and inexpensive, it is perfect to be used in class. Poppy robots are easy to program. Different options are possible based on students level and teaching objectives :
- Pixl is a board who manage power and communication between a raspberry pi and robotis XL320 low cost motors. We use this bord for all our low cost robots.
- 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 IT.
- 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.

Poppy Humanoid, Torso and Ergo robots 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.

Experimentations have began to be setup in 10 high-schools of Region Aquitaine, and 3 university level institutions: Lycée Camille Jullian (Bordeaux), Lycée Victor Louis (Talence), Lycée Saint Genès (Talence), Lycée François Mauriac (Bordeaux), Lycée Jean Moulin (Langon), Lycée des Graves (Gradignan), Lycée Sud Medoc (Le Taillan Medoc), Lycée Alfred Kastler (Talence), Lycée Raoul Follereau (Nevers), Aérocampus Auqitaine, ENSEIRB/IPB, ENSAM Talence.

7.5.4. IniRobot : Education and Thymio II Project (partnership with EPFL)

Participants: Didier Roy [correspondant], Pierre-Yves Oudeyer.

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 https://dm1r.inria.fr/inirobot or http://www.inirobot.fr. 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.

Deployment: After 16 months of activity, IniRobot is used by about 900 adults and 8000 children in 35 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.

HEPHAISTOS Project-Team

7. New Results

7.1. Robotics

7.1.1. Cable-driven parallel robots (CDPR)

7.1.1.1. Analysis of Cable-driven parallel robots

Participants: Alessandro Berti, Laurent Blanchet, Houssein Lamine, Jean-Pierre Merlet [correspondant], Yves Papegay, Rémy Ramadour.

We have continued the analysis of suspended CDPRs for control and design purposes. 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 [21], [22]. The FK problem is different if it is intended to be used in a real-time context as in that case we have the extra information of the platform pose a short time before. After a small change in the cable lengths we may assume a small change in the pose platform but using Newton method with the previous pose cannot guarantee to provide the current pose. We have proposed an algorithm that is guaranteed to get the current pose and is also able to determine if the CDPR may be sufficiently close to a singularity so that multiple solutions are possible [11]. However the assumption of a small change in the platform pose may not always hold, a point that we have shown theoretically and experimentally. We have then proposed an algorithm that uses a model of the coiling process to determine if a drastic change in the pose may occur between two sampling time [11] and also allows one to better estimate the cable tensions on a trajectory. We have for example shown that sudden and important changes in these tensions may occur. Another issue arises for non-deformable cables and CDPR with more than 6 cables in a suspend configuration. In the past we have shown that there always will be at most 6 cables under tension whatever the number of cables. For a given pose there may be several possible set of cables under tension (called *cable configuration*), each of them having different characteristics in terms of maximal tension, sensitivity to disturbances, ... From a control viewpoint it makes sense to impose a given cable configuration at the pose by setting the lengths of slack cables to larger values than the one required for the pose. To determine the best cable configuration we have proposed several ranking index [12].

Even more complex kinematic problems are involved if we assume that the cable are deformable (e.g. are elastic or catenary-like). The cable model is included in the kinematic equations for getting a complete model. We have been interested in the catenary-like model that involves inverse hyperbolic functions and is valid for steel cable of relatively high length. As the IK has never been addressed with such a model we have proposed a solving algorithm [10] that has shown that the IK may have multiple solutions but also may have no solution for poses that are reachable with non-deformable cables. In the same way the DK has several solutions [13]. Finally efficient cables interference detection for sagging cables and the management of modular CDPR, whose geometry may be changed according to the task at hand, have been addressed [9]

7.1.1.2. 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 this area.

In a collaboration with CNAM and Ecole Nationale Supérieure d'Architecture Paris-Malaquais, we worked on additive manufacturing of building based on ultra-high performance concrete and developed a CDPR as a proof of concept to power a large scale 3D-printer.

A real size industrial robot will be developed by the XtreeE start-up company.

7.1.2. Assistance

This is now the core of our activity and our work on CDPR is deeply connected to this field as they are an efficient solution for mobility assistance, a high priority for the elderly, helpers and medical community. We have presented our vision of assistance robotics in several occasions [24], [25], [23].

7.1.2.1. Assessment of elderly frailty

Participants: Karim Bakal, Jean-Pierre Merlet.

The assessment of elderly frailty is a difficult concept because it involves the physical capacities of a person and its environment (health-care services, families, funds...). We consider the assessment of upper limb capabilities by looking at the joint torques τ of the arm and the maximal force F that can be exerted by the hand, which are related by the equation

$$\tau = \mathbf{J}^{\mathbf{T}} F \tag{1}$$

where $\mathbf{J}$ is a matrix which depends only upon the configuration of the arm. These equations constitute an underconstrained linear system. In biomechanics the torque τ is measured together with the configuration of the arm and the force F is evaluated by using the method of Chiacchio, that involves the pseudo-inverse of $\mathbf{J}^{\mathbf{T}}$ to calculate F. But there are several uncertainties that are neglected when using this method: the measurement errors on τ and on the configuration of the arm together with uncertainties on the physical parameters of the arm (such as the length of the bones). The method of Chiacchio provides one of the possible solutions of equation (2) and not necessary the one corresponding to the force at the hand. We use another approach based on interval analysis. We assume that all uncertainties may be bounded (τ is an interval vector τ_m, $\mathbf{J}^{\mathbf{T}}$ is an interval matrix) so that equation (2) become an interval linear system. Interval analysis then allows one to determine an approximation as accurate as wanted of the set F_s of all forces F that satisfy the equation and therefore this set includes the real force at the hand. Now assume that with the same arm configuration we measure the force at the hand, here again with some bounded uncertainties (i.e. F is an interval vector F_m). Here again we may use interval analysis applied on equation (2) in order to determine an interval vector τ_v for the τ that is guaranteed to include the real τ. Furthermore τ must be included in the intersection τ_i of τ_v and τ_m while F must be included in the intersection F_i of F_m and F_s. If τ_i is strictly included in τ_m, then we may compute a better approximation of F_s. Reciprocally if F_i is strictly included in F_m we will get a better τ_v. If one of these situation occurs we repeat the process until no significant improvement of F_s or τ_v is obtained. In a second step we consider that the uncertainties that lead to uncertainties in the matrix $\mathbf{J}^{\mathbf{T}}$ are constrained as we have to satisfy $\tau_v = \mathbf{J}^{\mathbf{T}} F_s$. Here again we use interval analysis to determine if this constraint does not allow to reduce the size of the interval on the physical parameters in which case we may obtain a new $\mathbf{J}^{\mathbf{T}}$ that is included in the initial one. In turn this may allow to obtain better τ_v and F_s. The process stops when no improvement has been obtained for F_s, τ_v and the physical parameters.

To test this approach the right upper limb joint torque of 10 males and the force capacity at the right hand was measured by a dynamometer (Biodex III, Biodex Medical Systems) and respectively by a 6-axis load sensor during an experiment performed at HandiBio laboratory. The configuration of the upper limb was measured with a motion capture system (Qualisys, Sweden). The approach is currently being evaluated.

7.1.2.2. Walking analysis and Rehabilitation

Participants: Claire Maillard, Ting Wang, Jean-Pierre Merlet [correspondant].

The walkers of the ANG family allow one to determine accurately the trajectory of the walker and therefore to analyze the walking of the user. But these walkers may also be used to assess a rehabilitation process or the progress of an end-user involved in rehabilitation. For that purpose after having identified needs and requirements [17] we developed a new walker ANG-med that used infra-red distance sensors to measure the position of the subject during a rehabilitation exercise. Furthermore the software of this walker has been designed to support a message-passing scheme based on the HOP language of the INDES project team so that the walker may exchange information and control order with an external computer, together with allowing the download of new rehabilitation exercise through the robotics RAPP-store [26]. New exercises are designed as a set of such messages, that may include the calculation of exercise assessment indicators. ANG-med supports various modes: stand-alone (no external connection), passive mode (the walker only report indicator and status using a wifi connection) or full external control (an external computer fully control the walker except for emergency and real-time procedures).

ANG-med has been tested for one month in Centre Héliomarin de Vallauris and is now deployed in the rehabilitation center of MATIA in Spain, as part of the RAPP project. A start-up plan was proposed in November 2014 to transfer the walking analysis technology of HEPHAISTOS with the ANG walker in a company called Euthenia 9.2.1.3 .

7.1.2.3. Design and evaluation of assistive devices, ethics

Participants: Marc Beninati, Bernard Senach [correspondant], Jean-Pierre Merlet.

Providing appropriate support, services and information to the elderly, to their caregivers and to the medical profession, through a fleet of communicating devices must rely on a structured processes. A generic design and evaluation framework is being elaborated and will be validated through field experiments [20], [19], [18].

Assistance robotics raises many ethical questions. We started reflection about conducting experiments with frail and old people. A listing of questions to be addressed at each step of an experiment has been written (internal document). We have also hired a joint PhD student with University Bologna about the legal aspects of assistance robotics and we have initiated, together with Nathalie <Nevejans from University of Douai, a meeting with the OPECST at the French National Assembly to discuss legal and ethical aspects of robotics.

7.1.2.4. Smart Environment for Human Behaviour Recognition

Participants: Aurélien Massein, Yves Papegay, Odile Pourtallier.

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 information 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 part of a PhD work, optimal motion planning of a mobile robot with range sensors to locate targets in a room has been studied. Work in progress also include algorithms to deploy infra-red barriers in a large area with several interest places, to be able to locate people. An experimental set-up is in use in the lab and data analysis methods are developed to infer people behaviors.

7.2. Miscellaneous results

7.2.1. 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 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 [14].

Technology demonstrated by our prototype has been transferred, final version of our modeling and simulation environment has been delivered to Airbus in November 2012. Developer level know-how has been transferred in 2013 to a software company in charge of industrialization and maintenance of the modeling and simulation environment.

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.

LAGADIC Project-Team

7. New Results

7.1. Visual tracking

7.1.1. Object detection

Participant: Eric Marchand.

We addressed the challenge of detecting and localizing a poorly textured known object, by initially estimating its complete 3D pose in a video sequence [45]. Our solution relies on the 3D model of the object and synthetic views. The full pose estimation process is then based on foreground/background segmentation and on an efficient probabilistic edge-based matching and alignment procedure with the set of synthetic views, classified through an unsupervised learning phase. Our study focuses on space robotics applications and the method has been tested on both synthetic and real images, showing its efficiency and convenience, with reasonable computational costs.

7.1.2. Registration of multimodal images

Participant: Eric Marchand.

This study has been realized in collaboration with Brahim Tamadazte and Nicolas Andreff from Femto-ST, Besançon. Following our developments in visual tracking and visual servoing from the mutual information [3], it concerned mutual information-based registration of white light images vs. fluorescence images for micro-robotic laser microphonosurgery of the vocal folds. Nelder-Mead Simplex for nonlinear optimization has been used to minimize the cost-function [43].

7.1.3. Pose estimation from RGB-D sensor

Participant: Eric Marchand.

RGB-D sensors have become in recent years a product of easy access to general users. They provide both a color image and a depth image of the scene and, besides being used for object modeling, they can also offer important cues for object detection and tracking in real-time. In this context, the work presented in this paper investigates the use of consumer RGB-D sensors for object detection and pose estimation from natural features. Two methods based on depth-assisted rectification are proposed, which transform features extracted from the color image to a canonical view using depth data in order to obtain a representation invariant to rotation, scale and perspective distortions. While one method is suitable for textured objects, either planar or non-planar, the other method focuses on texture-less planar objects [18]

7.1.4. 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 have started to perform this localization from a set of keyframe images corresponding to the landing trajectory.

7.2. Visual servoing

7.2.1. Histogram-based visual servoing

Participants: Quentin Bateux, Eric Marchand.

Classically visual servoing considers the regulation in the image of a set of visual features (usually geometric features). Direct visual servoing schemes, such as photometric visual servoing, have been introduced in order to consider every pixel of the image as a primary source of information and thus avoid the extraction and the tracking of such geometric features. This year, we proposed a method to extend these works by using a global descriptor, namely intensity histograms, on the whole or multiple sub-sets of the images in order to achieve control of a 6 degrees of freedom (DoF) robot [30][53].

7.2.2. Photometric moment-based visual servoing

Participants: Manikandan Bakthavatchalam, François Chaumette.

This work also belongs to the class of direct visual servoing. Its goal was to use photometric moments as visual features in order to increase the convergence domain of this approach by reducing the non linearity of the control problem. In order to cope with appearance and disappearance of some parts of the environment during the camera motion, a spatial weight has been introduced in the definition of photometric moments. Thanks to a particular design of this weight, the analytical form of the interaction matrix has been obtained, from which it was possible to select a set of moment combinations to control all the six degrees of freedom of the system. Satisfactory experimental results have been obtained [29][8], even if the loss of invariance properties makes the optimal design of visual features still an open problem.

7.2.3. Model predictive visual servoing

Participants: Nicolas Cazy, Paolo Robuffo Giordano, François Chaumette.

The goal of this work is to exploit Model Predictive Control (MPC) techniques for dealing in a robust way with loss of features during a IBVS task. The work [31] provides an experimental validation of different correction schemes able to cope with loss of features due to occlusions of limited camera field of view. The reported results show the effectiveness of the proposed techniques during the servoing of four point features.

7.2.4. Nanomanipulation

Participants: Le Cui, Eric Marchand.

Following our work related to scanning electron micro- scope (SEM) calibration [12] we considered the control of a micro robot using a direct photometric visual servoing that uses only the pure image information as a visual feature, instead of using classic geometric features such as points or lines. However, in micro-scale, using only image intensity as a visual feature performs unsatisfactorily in cases where the photometric variation is low, such as motions along vision sensor's focal axis under a high magnification. In order to improve the performance and accuracy in those cases, an approach using hybrid visual features is proposed in this paper. Image gradient is employed as a visual feature on z axis while image intensity is used on the other 5 DoFs to control the motion. A 6-DoF micro-positioning task is accomplished by this hybrid visual servoing scheme [34].

We also considered a full scale autofocus approach for SEM [35]. The optimal focus (in-focus) position of the microscope is achieved by maximizing the image sharpness using a vision-based closed-loop control scheme. An iterative optimization algorithm has been designed using the sharpness score derived from image gradient information. The proposed method has been implemented and validated using a tungsten gun SEM at various experimental conditions like varying raster scan speed, magnification at real-time.

7.2.5. Audio-based control

Participants: Aly Magassouba, François Chaumette.

This study is not concerned with visual servoing, but to the application of the same principle of sensor-based control to audio sensors. It is made in collaboration with Nancy Bertin from Panama group at Irisa, Inria Rennes-Bretagne Atlantique. In a first step, we have determined the analytical form of the interaction matrix of audio features based on the time difference of arrival on two microphones. From this modeling step, we have determined the different virtual linkages that can be realized in function of the number and configuration of sources [41]. First experimental results using two microphones monunted on the Pioneer mobile robot (see Section 6.9) have been recently obtained.

7.3. Visual navigation of mobile robots

7.3.1. Visual navigation from straight lines

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 localisation as well as for the visual servoing is not based on points of interest, but straight lines that are more common indoors. Satisfactory experimental results have been obtained using the Pioneer mobile robot (see Section 6.9) [23].

7.3.2. Autonomous navigation of a wheelchair and social navigation

Participants: Vishnu Karakkat Narayanan, François Pasteau, Marie Babel.

Navigating within an unknown indoor environment using an electric wheelchair is a challenging task, especially if the user suffers from severe disabilities. We presented in [22] a framework for vision-based autonomous indoor navigation in an electric wheelchair capable of following corridors, and passing through open doorways using a single doorpost. The designed control schemes have been implemented onto a robotized wheelchair and experimental results show the robust behaviour of the designed system.

We then introduced in [40] a task-based control law which can serve as a low-level system for equitably joining interacting groups, while confirming to social conventions. The system uses the position and orientation of the participating humans with respect to a rigid sensor frame in order to control the translational and rotational velocity of a wheelchair so that the robot positions itself aptly at the meeting point

7.3.3. Semi-autonomous control of a wheelchair for navigation assistance

Participants: Vishnu Karakkat Narayanan, François Pasteau, Marie Babel.

To address the wheelchair driving assistance issue, we proposed in [56][28] a unified shared control framework able to smoothly correct the trajectory of the electrical wheelchair. 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.

The robotic solution has been then validated through clinical trials that have been conducted within the Rehabilitation Center of Pôle Saint Hélier (France) with 25 volunteering patients presenting different disabling neuro-pathologies. This assistive tool is shown to be intuitive and robust as it respects the user intention, it does not alter perception while reducing the number of collisions in case of hazardous maneuvers or in crowded environment [27].

7.4. 3D Scene Mapping

7.4.1. Structure from motion

Participants: Riccardo Spica, Paolo Robuffo Giordano, François Chaumette.

Structure from motion (SfM) is a classical and well-studied problem in computer and robot vision, and many solutions have been proposed to treat it as a recursive filtering/estimation task. However, the issue of *actively* optimizing the transient response of the SfM estimation error has not received a comparable attention. In the work [50] we have addressed the active estimation of the 3D structure of an observed planar scene by comparing three different techniques: a homography decomposition (a well-established method taken as a baseline), a least-square fitting of a reconstructed 3D point cloud, and a direct estimation based on the observation of a set of discrete image moments made of a collection of image points belonging to the observed plane. The experimental results confirmed the importance of actively controlling the camera motion in order to obtained a faster convergence for the estimation error, as well as the superiority of the third method based on the machinery of image moments for what concerns robustness against noise and outliers. In [51] the active estimation scheme has been improved by considering a set of features invariant to camera rotations. This way, the dynamics of the structure estimation becomes independent of the camera angular velocity whose measurement is, thus, no longer required for implementing the active SfM scheme. Finally, in [46] the issue of determining online the 'best' combination of image moments for reconstructing the scene structure has been considered. By defining a new set of weighted moments as a weighted sum of traditional image moments, it is indeed possible to optimize for the weights online during the camera motion. The SfM scheme then automatically selects online the best combination of image moments to be used as measurements as a function of the current scene.

7.4.2. Scene Registration based on Planar Patches

Participants: Eduardo Fernandez Moral, Patrick Rives.

Scene registration consists of estimating the relative pose of a camera with respect to a scene previously observed. 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 [14].

7.4.3. Robust RGB-D Image Registration

Participants: Tawsif Gokhool, Renato José Martins, Patrick Rives.

Estimating dense 3D maps from stereo sequences remains a challenging task where building compact and accurate scene models is relevant for a number of tasks, from localization and mapping to scene rendering [20], [10] . In this context, this work deals with generating complete geometric and photometric "minimal" model of indoor/outdoor large-scale scenes, which are stored within a sparse set of spherical images to asset photo-geometric consistence of the scene from multiple points-of-views . To this end, a probabilistic data association framework for outlier rejection is formulated, enhanced with the notion of landmark stability over time. The approach was evaluated within the frameworks of image registration, localization and mapping, demonstrating higher accuracy and larger convergence domains over different datasets [39].

7.4.4. Accurate RGB-D Keyframe Representation of 3D Maps

Participants: Renato José Martins, Eduardo Fernandez Moral, Patrick Rives.

Keyframe-based maps are a standard solution to produce a compact map representation from a continuous sequence of images, with applications in robot localization, 3D reconstruction and place recognition. We have present a approach to improve keyframe-based maps of RGB-D images based on two main filtering stages: a regularization phase in which each depth image is corrected considering both geometric and photometric image constraints (planar and superpixel segmentation); and a fusion stage in which the information of nearby frames (temporal continuity of the sequence) is merged (using a probabilistic framework) to improve the accuracy and reduce the uncertainty of the resulting keyframes. As a result, more compact maps (with less keyframes) are created. We have validated our approach with different kind of RGB-D data including both indoor and outdoor sequences, and spherical and perspective sensors, demonstrating that our approach compares and outperforms the state-of-the-art [42].

7.4.5. Semantic Representation For Navigation In Large-Scale Environments

Participants: Romain Drouilly, Patrick Rives.

Autonomous navigation is one of the most challenging problem to address to allow robots to evolve in our everyday environments. Map-based navigation has been studied for a long time and researches have produced a great variety of approaches to model the world. However, semantic information has only recently been taken into account in those models to improve robot efficiency.

Mimicking human navigation is a challenging goal for autonomous robots. This requires to explicitly take into account not only geometric representation but also high-level interpretation of the environment [9]. We propose a novel approach demonstrating the capability to infer a route in a global map by using semantics. Our approach relies on an object-based representation of the world automatically built by robots from spherical images. In addition, we propose a new approach to specify paths in terms of high-level robot actions. This path description provides robots with the ability to interact with humans in an intuitive way. We perform experiments on simulated and real-world data, demonstrating the ability of our approach to deal with complex large-scale outdoor environments whilst dealing with labelling errors [37].

Mapping evolving environments requires an update mechanism to efficiently deal with dynamic objects. In this context, we propose a new approach to update maps pertaining to large-scale dynamic environments with semantics. While previous works mainly rely on large amount of observations, the proposed framework is able to build a stable representation with only two observations of the environment. To do this, scene understanding is used to detect dynamic objects and to recover the labels of the occluded parts of the scene through an inference process which takes into account both spatial context and a class occlusion model. Our method was evaluated on a database acquired at two different times with an interval of three years in a large dynamic outdoor environment. The results point out the ability to retrieve the hidden classes with a precision score of 0.98. The performances in term of localisation are also improved [36].

7.5. Control of single and multiple Unmanned Aerial Vehicles

7.5.1. Single UAV

Participant: Paolo Robuffo Giordano.

Over the last years the robotics community witnessed an increasing interest in the Unmanned Aerial Vehicle (UAV) field. In particular quadrotor UAVs have become more and more widespread in the community as experimental platform for, e.g., testing novel 3D planning, control and estimation schemes in real-world indoor and outdoor conditions. Indeed, in addition to being able to take-off and land vertically, quadrotors can reach high angular accelerations thanks to the relatively long lever arm between opposing motors. This makes them more agile than most standard helicopters or similar rotorcraft UAVs, and thus very suitable to realize complex tasks such as aerial mapping, air pollution monitoring, traffic management, inspection of damaged buildings and dangerous sites, as well as agricultural applications such as pesticide spraying.

Despite these clear advantages, a clear shortcoming of the quadrotor design lies in its inherent underactuation (only 4 actuated propellers for the 6 dofs of the quadrotor pose). This underactuation limits the quadrotor flying ability in free or cluttered space and, furthermore, it also degrades the possibility of interacting with the environment by exerting desired forces in arbitrary directions. In [24], a novel design for a quadrotor UAV with tilting propellers which is able to overcome these limitations has been presented and experimentally validated. Indeed, the additional set of 4 control inputs actuating the propeller tilting angles can be shown to yield full actuation to the quadrotor position/orientation in space, thus allowing it to behave as a fully-actuated flying vehicle and to overcome the aforementioned underactuation problem.

Furthermore, the issue of estimating online the UAV self-motion from vision has been considered. To this end, a novel nonlinear estimation scheme able to recover the metric UAV linear velocity from the *scaled* one obtained from the decomposition of the optical flow has been proposed in [15]. The observability conditions (in terms of persistency of excitation) needed to ensure a converging estimation have also been studied. The reported experimental results confirmed the effectiveness of the estimation scheme in recovering a reliable and accurate estimation of the UAV self-motion (linear and angular velocities) in realistic conditions.

This work has been realized in collaboration with the Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

7.5.2. Collective control of multiple UAVs

Participants: Fabrizio Schiano, 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 fulfilment 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. However, connectivity alone is not sufficient to perform certain tasks when only *relative sensing* is used. 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 [26], a decentralized gradient-based rigidity maintenance action for a group of quadrotor UAVs has been proposed and tested in real experimental conditions. By starting in a rigid configuration, the group of UAVs is 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. Furthermore, in [52] the novel case of *bearing rigidity* for directed graphs has been considered: here, rather than distances the measurements are the 3D bearing vectors expressed in the local body-frame of each agent. The theory has been developed for the case of planar agents in $SE(2)$ and a 'scale-free' bearing controller has been proposed, able to steer the robot group towards a desired bearing formation.

These works were realized in collaboration with the robotics group at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany and with Technion, Israel.

7.5.3. Cooperative localization using interval analysis

Participants: Vincent Drevelle, Ide Flore Kenmogne Fokam.

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 bounder-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. Encouraging results have already been obtained for multi-robot underwater positioning with acoustical range measurements. Ongoing work focuses on cooperative localization in a multi-UAV fleet with image-based measurements (bearings).

7.6. Medical robotics

7.6.1. Non-rigid target tracking in ultrasound images combining dense information and physically-based model

Participants: Lucas Royer, Alexandre Krupa.

This study concerns the real-time tracking of deformable targets within a sequence of ultrasound (US) images. The proposed approach combines dense information with a physically-based model and has therefore the advantage of not using any fiducial marker. The physical model is represented by a mass-spring damper system driven by external and internal forces. The external forces are obtained by maximizing an image similarity metric between a reference target and the deformed target along the time. The internal forces of the mass-spring damper system constrain the deformation to be physically plausible and therefore efficiently reduce the sensitivity to the speckle noise. This approach was first validated from simulated and real sequences of 2D US images [49]. It was then extended for deformable target tracking in a sequence of 3D ultrasound volumes and tested on a robotic setup used to apply deformation on an organic phantom [48]. The performance of this deformable 3D target tracking approach was evaluated with visual assessment combined with robotic odometry ground truth. This method was also tested and compared with respect to state-of-the-art techniques by using 3D image databases provided by MICCAI CLUST'14 and CLUST'15 challenges [47] (MICCAI Challenge on Liver Ultrasound Tracking). It was awarded by the organizers of the CLUST challenges as being the best method for accurate target tracking in 3D ultrasound sequences. We recently improved our approach in order to increase its robustness to the presence of ultrasound shadows, local illumination changes and image occlusions.

7.6.2. 3D steering of flexible needle by ultrasound visual servoing

Participants: Pierre Chatelain, 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 previous work, we designed a control approach based on a duty cycling technique for steering a beveled-tip flexible needle actuated by a robotic arm in such a way to control the needle curvature in 3D space and reach a desired target by visual servoing. In this preliminary work, the control approach was validated by using visual features extracted from 2 images provided by 2 orthogonal cameras observing a translucent gelatin phantom where the needle was inserted. This year, we have pursued our work towards this needle steering robotic assistance by developing a new algorithm able to track in real-time a flexible needle in a sequence of 3D ultrasound images (volumes). The flexible needle modeled as a polynomial curve is tracked during the automatic insertion using particle filtering. This new tracking algorithm enables real-time closed-loop needle control with 3D ultrasound feedback. The target to reach was manually defined by the user in the US image and can be on-line tracked thanks to the template tracking algorithm proposed in [21] based on ultrasound dense visual servoing [7]. Experimental results of an automatic needle tip positioning in a home-made gelatine phantom demonstrate the feasibility of 3D ultrasound-guided needle steering for reaching a desired target by ultrasound visual servoing [33]. Recently a new control law for needle steering that uses both direct manipulation of the needle base and the duty cycling method has been studied. It is based on a 3D model of a beveled tip needle using virtual springs that characterize the needle mechanical interaction with soft tissue. 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. Preliminary simulations show that this hybrid control allows better targeting capabilities in terms of larger needle workspace and reduced needle bending.

7.6.3. Optimization of ultrasound image quality by visual servoing

Participants: Pierre Chatelain, Alexandre Krupa.

This study focuses on a new ultrasound-based visual servoing approach that optimizes the positioning of an ultrasound probe manipulated by a robotic arm in order to improve the quality of the acquired ultrasound images. To this end, we use the recent framework of ultrasound confidence map, developed in the Chair for Computer Aided Medical Procedures and Augmented Reality of Prof. Nassir Navab, which aims at estimating the per-pixel quality of the ultrasound signal based on a model of sound propagation in soft tissues. More specifically, we treat the ultrasound confidence maps as a new modality and designed a visual servoing control law for image quality optimization. We illustrated our approach with the application of robotic tele-echography where the in-plane rotation of a 2D probe is visually servoed by the confidence map and the other degrees of freedom are teleoperated by the user. Experiments performed on both an ultrasound examination training phantom and ex vivo tissue samples validated this new concept [32]. Currently, we consider the confidence-driven servoing of other degrees of freedom, in particular out-of-plane motions that were controlled in our previous works from image moments [6], which could provide finer control of the image quality.

7.6.4. Visual servoing based on ultrasound elastography

Participants: Pedro Alfonso Patlan Rosales, Alexandre Krupa.

This study concerns the use of the ultrasound elastography as a new image modality for the control of the motion of an ultrasound probe actuated by a robotic manipulator. 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 a user-dependent quality of the elastography image. To improve the US elastography imaging, we recently developed an assistant robotic palpation system that automatically moves an ultrasound probe in such a way to optimize ultrasound elastography. The main originality of this preliminary work concerns the use of the elastography modality directly as input of the robot controller thanks to an innovative ultrasound elastography-based visual servoing approach.

7.6.5. Visual servoing using shearlet transform

Participants: Lesley-Ann Duflot, Alexandre Krupa.

Similar to wavelet transform, shearlet transform is usually used in the field of signal or image compression. At the best of our knowledge these image representations were never used directly as feedback of a closed-loop control scheme. The objective of this work is to study the feasibility of using the coefficients of shearlet transform of the observed ultrasound image directly as the visual features of an image-based visual servoing. In this study we estimated numerically the interaction matrix that links the time variation of the coarsest coefficients of the shearlet to the motion of the ultrasound probe. This shearlet-based visual servoing was experimentally tested for automatically positioning a 2D US probe, held by a robot, on a desired section of an abdominal phantom. The first results demonstrated promising performances.

LARSEN Team

7. New Results

7.1. Lifelong Autonomy

7.1.1. Adaptation / Learning

Participant: Jean-Baptiste Mouret.

We collaborate on this subject with Jeff Clune (University of Wyoming, USA).

7.1.1.1. Adaptation to Unforeseen Damage Conditions

Whereas animals can quickly adapt to injuries, current robots cannot “think outside the box” to find a compensatory behaviour when they are damaged: they are limited to their pre-specified self-sensing abilities and can diagnose only anticipated failure modes, an impracticality for complex robots. A promising approach to reducing robot fragility involves having robots learn appropriate behaviours in response to damage, but current techniques are slow even with small, constrained search spaces. We introduced an intelligent trial-and-error algorithm that allows robots to adapt to damage in less than two minutes in large search spaces without requiring self-diagnosis or pre-specified contingency plans [11]. Before the robot is deployed, it uses a novel technique (based on evolutionary algorithms) to create a detailed map of the space of high-performing behaviours. This map represents the robot’s prior knowledge about what behaviours it can perform and their value. When the robot is damaged, it uses this prior knowledge to guide a trial-and-error learning algorithm (based on Bayesian optimization) that conducts intelligent experiments to rapidly discover a behaviour that compensates for the damage. Experiments reveal successful adaptations for a legged robot injured in five different ways, including damaged, broken, and missing legs, and for a robotic arm with joints broken in 14 different ways. This new algorithm will enable more robust, effective, autonomous robots, and may shed light on the principles that animals use to adapt to injury.

This work was the cover of Nature on the 28th of May, 2015 (see the “highlights” section).

7.1.2. Robotics Perception

Participants: François Charpillet, Francis Colas, Abdallah Dib, Van Quan Nguyen.

We collaborate on this subject with Emmanuel Vincent from the Multispeech team (Inria Nancy - Grand Est).

7.1.2.1. Audio Source Localization

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. However, the robot movements are typically fixed or they obey heuristic strategies, such as turning the head and moving towards the source, which may be suboptimal. This work proposes an approach to control the robot movements so as to locate the source as quickly as possible [17]. We represent the belief about the source position by a discrete grid and we introduce a dynamic programming algorithm to find the optimal robot motion minimizing the entropy of the grid. We report initial results in a real environment.

This work is carried on through the PhD Thesis of Van Quan Nguyen under the supervision of Emmanuel Vincent and Francis Colas.

7.1.2.2. State Estimation for Autonomous Surface Vessels

Autonomous Surface Vessels (ASVs) are increasingly proposed as tools to automatize environmental data collection, bathymetric mapping and shoreline monitoring. For many applications it can be assumed that the boat operates on a 2D plane. However, with the involvement of exteroceptive sensors like cameras or laser rangefinders, knowing the 3D pose of the boat becomes critical. We formulated three different algorithms based on 3D extended Kalman filter (EKF) state estimation for ASVs localization [12]. We compared them using field testing results with ground truth measurements, and demonstrated that the best performance is achieved with a model-based solution in combination with a complementary filter for attitude estimation. Furthermore, we presented a parameter identification methodology and showed that it also yielded accurate results when used with inexpensive sensors. Finally, we presented a long-term series (i.e., over a full year) of shoreline monitoring data sets and discussed the need for map maintenance routines based on a variant of the Iterative Closest Point (ICP) algorithm.

7.1.2.3. Geometric Registration

We proposed a review of geometric registration in robotics [16]. Registration algorithms associate sets of data into a common coordinate system. They have been used extensively in object reconstruction, inspection, medical application, and localization of mobile robotics. We focus on mobile robotics applications in which point clouds are to be registered. While the underlying principle of those algorithms is simple, many variations have been proposed for many different applications. In this work, we gave a historical perspective of the registration problem and showed that the plethora of solutions can be organized and differentiated according to a few elements. Accordingly, we presented a formalization of geometric registration and cast algorithms proposed in the literature into this framework. Finally, we reviewed a few applications of this framework in mobile robotics that cover different kinds of platforms, environments, and tasks. These examples allowed us to study the specific requirements of each use case and the necessary configuration choices leading to the registration implementation. Ultimately, the objective of this work is to provide guidelines for the choice of geometric registration configuration.

7.1.2.4. Robust Dense Visual Odometry for RGB-D Cameras in a Dynamic Environment

Visual odometry is a fundamental challenge in robotics and computer vision. The aim of our work is to estimate RGB-D camera motion (onboard a mobile robot) from RGB-D images in a dynamic scene with people moving in the scene. Most of the existing methods have a poor localization performance in such case, which makes them inapplicable in real world conditions. This year, we have proposed a new dense visual odometry method [27] that uses random sampling consensus (RANSAC) to cope with dynamic scenes. We show the efficiency and robustness of the proposed method on a large set of experiments in challenging situations and from publicly available benchmark datasets. Additionally, we compare our approach to another state-of-art method based on M-estimator that is used to deal with dynamic scenes. Our method gives similar results on benchmark sequences and better results on our own dataset.

7.1.3. Distributed Sensing and Acting

Participants: Mihai Andries, Amine Boumaza, François Charpillet, Iñaki Fernández Pérez, Nassim Kaldé.

We collaborate on this subject with Olivier Simonin from the Chroma team (Inria Grenoble - Rhône Alpes).

7.1.3.1. Localisation of Humans, Objects and Robots Interacting on Load-Sensing Floors

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 Hikob company (http://www.hikob.com/) and Inria SED (*service d'expérimentation et de développement*). 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 intelligent apartment experimental platform. This load-sensing floor includes as well a LED lighting system which sits flush with the floor surface. This provides people with a new way to interact with their environment at home. This year, we have focused on localisation, tracking and recognition of humans, objects and robots interacting on load-sensing floors [9]. Inspired by computer vision, the proposed technique processes the floor pressure-image by segmenting the blobs containing objects, tracking them, and recognizing their contents through a mix of inference and combinatorial search. The result lists the probabilities of assignments of known objects to observed blobs. The concept was successfully evaluated in daily life activity scenarii, involving multi-object tracking and recognition on low resolution sensors, crossing of user trajectories, and weight ambiguity.

7.1.3.2. Online Distributed Learning for a Swarm of Robots

We propose a novel innovation marking method [22] for neuro-evolution of augmenting topologies in embodied evolutionary robotics. This method does not rely on a centralized clock, which makes it well suited for the decentralized nature of embodied evolution where no central evolutionary process governs the adaptation of a team of robots exchanging messages locally. This method is inspired from event dating algorithms, based on logical clocks, that are used in distributed systems, where clock synchronization is not possible. We compare our method to odNEAT, an algorithm in which agents use local time clocks as innovation numbers, on two multi-robot learning tasks: navigation and item collection. Our experiments showed that the proposed method performs as well as odNEAT, with the added benefit that it does not rely on synchronization of clocks and is not affected by time drifts.

The effect of selection pressure on evolution in centralized evolutionary algorithms (EA's) is relatively well understood. Selection pressure pushes evolution toward better performing individuals. However, distributed EA's in an Evolutionary Robotics (ER) context differ in that the population is distributed across the agents, and a global vision of all the individuals is not available. In this work, we analyze the influence of selection pressure in such a distributed context. We propose a version of mEDEA [22] that adds a selection pressure, and evaluate its effect on two multi-robot tasks: navigation and obstacle avoidance, and collective foraging. Experiments show that even small intensities of selection pressure lead to good performances, and that performance increases with selection pressure. This is opposed to the lower selection pressure that is usually preferred in centralized approaches to avoid stagnating in local optima.

7.1.3.3. Online Distributed Exploration of an Unknown Environment by a Swarm of Robots

This year, we have proposed a new taboo-list approach [18] for multi-robot exploration of unknown structured environments, in which robots are implicitly guided in their navigation on a globally shared map. Robots have a local view of their environment, inside which they navigate in a asynchronous manner. When the exploration is complete, robots gather at a rendezvous point. The novelty consists in using a distributed exploration algorithm which is not guided by frontiers to perform this task. Using the Brick and Mortar Improved ant-algorithm as a base, we add robot-perspective vision, variable vision range, and an optimization which prevents agents from going to the rendezvous point before exploration is complete. The algorithm was evaluated in simulation on a set of standard maps.

Another work [14] carried out within the PhD of Nassim Kaldé concerns exploration in populated environments. The difficulty here is that pedestrian flows can severely impact performances. However, humans have adaptive skills for taking advantage of these flows while moving. Therefore, in order to exploit these human abilities, we propose a novel exploration strategy that explicitly allows for human-robot interactions. Our model for exploration in populated environments combines the classical frontier-based strategy with our interactive approach. We implement interactions where robots can locally choose a human guide to follow and define a parametric heuristic to balance interaction and frontier assignments. Finally, we evaluate to which extent human presence impacts our exploration model in terms of coverage ratio, travelled distance and elapsed time to completion.

7.2. Natural Interaction with Robotics Systems

7.2.1. Human Characterization

Participants: François Charpillet, Abdallah Dib, Xuan Son Nguyen, Vincent Thomas.

We collaborate on this subject with Olivier Buffet and Alain Dutech from Inria Nancy - Grand Est, Arsène Fansi Tchango and Fabien Flacher from Thales ThereSIS, and Alain Filbois from SED Inria Nancy - Grand Est.

7.2.1.1. Multi-Camera Tracking in Partially Observable Environment

In collaboration with Thales ThereSIS - SE&SIM Team (Synthetic Environment & Simulation), we focus on the problem of following the trajectories of several persons with the help of several controllable cameras. This is a difficult problem since the set of cameras cannot simultaneously cover the whole environment, since some persons can be hidden by obstacles or by other persons, and since 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 (1) the HMM (Hidden Markov Models) formalism to represent the state of the system (the persons and their internal states), (2) a simulator provided and developed by Thales ThereSIS, and (3) particle filtering approaches based on this simulator. Since activity and location depend on each other, we adopt a Simultaneous Tracking and Activity Recognition approach.

After having shown that it was possible to use a complex behavioral simulator to infer the behavior of complex individuals (motivation, possession, ...) even in case of long periods of occlusions [40], we investigated how to propose a factored particle filter (with one distribution per target) for efficiently tracking multiple targets simultaneously. To that end, we use a Joint Probabilistic Data Association Filter with a particular model of dynamics that largely decouples the evolution of several targets, and turns out to be very natural to apply. We proposed to use a small number of “representatives” of each target to determine and consider only effective interactions among targets.

This work has been published in Arsène Fansi Tchango’s PhD thesis which has been defended in December [7].

7.2.1.2. Human Posture Recognition

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/).

7.2.2. Social Robotics

Participants: Amine Boumaza, Serena Ivaldi.

We collaborate on this subject with Yann Boniface from Loria, Alain Dutech from Inria Nancy - Grand Est and Nicolas Rougier from the Mnemosyne team (Inria Bordeaux - Sud-Ouest).

7.2.2.1. PsyPhINe: Cogito Ergo Es

PsyPhINe is an interdisciplinary and exploratory project (see 9.1.2) 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?

Preliminary experiments with human subjects conducted this past year on a simple device helped to design an experimental protocol and test simple hypotheses which set the ground for the full fledged non verbal Turing Test. This project was funded under a PEPS Mirabelle grant (see 9.1.2) which helped build a robotic device with many degrees of freedom to perform further experiments. We also organized an inter-disciplinary workshop gathering top researchers from philosophy, anthropology, psychology and computer science to discuss and exchange on our methodology (see 10.1.1.1).

7.2.2.2. Multimodal Object Learning During Human-Robot Interaction

Robots working in evolving human environments need the ability to continuously learn to recognize new objects. Ideally, they should act as humans do, by observing their environment and interacting with objects, without specific supervision. However, if object recognition simply relies on visual input, then it may fail during human-robot interaction, because of the superposition of human and body parts. A multimodal approach was then proposed in [15], where visual input from cameras was combined with the robot proprioceptive information, in order to classify objects, robot, and human body parts. We proposed a developmental learning approach that enables a robot to progressively learn appearances of objects in a social environment: first only through observation, then through active object manipulation. We focused on incremental, continuous, and unsupervised learning that does not require prior knowledge about the environment or the robot. In the first phase of the proposed method, we analyse the visual space and detect proto-objects as units of attention that are learned and recognized as possible physical entities. The appearance of each entity is represented as a multi-view model based on complementary visual features. In the second phase, entities are classified into three categories: parts of the body of the robot, parts of a human partner, and manipulable objects. The categorization approach is based on mutual information between the visual and proprioceptive data, and on motion behaviour of entities. The ability to categorize entities is then used during interactive object exploration to improve the previously acquired objects models. The proposed system was implemented and evaluated with an iCub and a Meka robot learning 20 objects. The system was able to recognize objects with 88.5% success rate and create coherent representation models that are further improved by learning during human-robot interaction.

7.2.2.3. Robot Functional and Social Acceptance

To investigate the functional and social acceptance of a humanoid robot, we carried out an experimental study with 56 adult participants and the iCub robot. Trust in the robot has been considered 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), and measured by the participants' conformation to the iCub's answers to specific questions. In particular, we were interested in understanding whether specific (i) user-related features (i.e., desire for control), (ii) robot-related features (i.e., attitude towards social influence of robots), and (iii) context-related features (i.e., collaborative vs. competitive scenario), may influence their trust towards the iCub robot. 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 savvy does not thus seem to be a pre-requisite for trust in its social savvy. Finally, desire for control, attitude towards social influence of robots and type of interaction scenario did not influence the trust in iCub. Results are also discussed with relation to methodology of HRI research in a currently submitted paper (http://arxiv.org/abs/1510.03678 [cs.RO]). This work follows the research on engagement with social robots that was previously published [10].

7.2.2.4. Relation Between Extroversion and Negative Attitude Towards Robot

Estimating the engagement is critical for human - robot interaction. Engagement measures typically rely on the dynamics of the social signals exchanged by the partners, especially speech and gaze. However, the dynamics of these signals is likely to be influenced by individual and social factors, such as personality traits, as it is well documented that they critically influence how two humans interact with each other. We assess the influence of two factors, namely extroversion and negative attitude toward robots, on speech and gaze during a cooperative task, where a human must physically manipulate a robot to assemble an object [23]. We evaluate if the score of extroversion and negative attitude towards robots co-variate with the duration and frequency of gaze and speech cues. The experiments were carried out with the humanoid robot iCub and 56 adult participants. 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.

RITS Project-Team

7. New Results

7.1. 2D Laser Based Road Obstacle Classification for Road Safety Improvement

Participants: Pierre Merdrignac, Evangeline Pollard, Fawzi Nashashibi.

Vehicle and pedestrian collisions often result in fatality to the vulnerable road users (VRU), indicating a strong need to protect such persons. Laser sensors have been extensively used for moving obstacles detection and tracking. Laser impacts are produced by reflection on these obstacles which suggests an information is available to recognize multiple road obstacles classes (pedestrian, cyclists, vehicles,...). We introduce a new system to address this problem that is divided in three parts: definition of geometric features describing road obstacles, multi-class object classification from an adaboost trained classifier and Bayesian estimation of the obstacle class. This approach benefits from consecutive observations of a single obstacle to estimate its class more precisely. We tested our system on some laser sequences and showed that it can estimate the class of some road obstacles around the vehicle with an accuracy of 87.4%. The vehicle class is determined with more than 97% of success. However, the main source of confusion is for static obstacles (posts and trees) for which 15% are classified as pedestrians. More detail can be fund in [36], [16].

7.2. On line Mapping and Global Positioning technique based on evidential SLAM

Participants: Guillaume Trehard, Evangeline Pollard, Fawzi Nashashibi.

Locate a vehicle in an urban environment remains a challenge for the autonomous driving community. By fusing information from a LIDAR, a Global Navigation by Satellite System (GNSS) and the vehicle odometry, we introduced and developed an original solution based on evidential grids and a particle filter to map the static environment and simultaneously estimate the position in a global reference at a high rate and without any prior knowledge (see [39]).

7.3. PML-SLAM

Participants: Zayed Alsayed, Fawzi Nashashibi, Anne Verroust-Blondet.

Our goal is to improve localization systems performances in order to be able to navigate in large-scale urban environments. In this context, we first optimized CPU and memory consumption of a SLAM laser-based technique [52] by introducing a map manager system. This strategy allows a smooth navigation while saving and loading probabilities-grid submaps into/from a hard-disc in a transparent way (cf. [27]). This work was validated and extended in the context of ITS Bordeaux demonstrations (VEDECOM demonstrator), where GPS information was integrated into SLAM environment Maps.

7.4. Motion planning techniques

Participants: David Gonzalez Bautista, Fernando Garrido Carpio, Joshué Pérez Rastelli, Vicente Milanés Montero, Fawzi Nashashibi.

Intelligent vehicles have increased their capabilities for highly, and even fully, automated driving under controlled environments. Scene information is received using on-board sensors and communication network systems–i.e. infrastructure and other vehicles. Considering the available information, different motion planning techniques have been implemented to autonomously driving on complex environments. The main goal is focused on executing strategies to improve safety, comfort and energy optimization. However, 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. We have recently carried out a deep state-of-the-art review to find the gaps in this hot topic into the autonomous vehicle field, paying special attention to overtaking and obstacle avoidance maneuvers.

Based on this review, we have mainly identified two main gaps: trajectory and speed planning with dynamics obstacle avoidance capabilities and real-time performance of the algorithms in the sense of significantly reducing the computational time, moving the system closer to what a vehicle should be able to provide in the real world.

According to this review, a speed planner has been designed with specific considerations on computing time efficiency, with an optimal comfort and avoiding to exceed speed and acceleration limits [31]. The comfort is evaluated as the minimization and smoothness of acceleration and jerk profiles, while maintaining a coherent speed profile with respect to traffic rules, the geometry of the path and the lateral accelerations associated to it. Specifically, this speed planner uses fifth order polynomial curves. These curves are C2 continuous and smooth, meaning that the jerk profile is also continuous and smooth. The method proposed computes the velocity in terms of the length of the path, instead of time, greatly reducing the errors. Specific targets for the speed planner are:

- Compute a smooth and continuous speed profile accounting for acceleration limits (longitudinal and lateral) according to ISO 2631-1 standard.
- Minimize distance error problems by associating the speed profile in the path speed planner instead of the time.

This speed planner was tested against other techniques providing better results in terms of computational time and smoothness (cf. [32]).

Additionally, a novel trajectory planning with a significant reduction on the computational time with respect to prior implementations from the team has been implemented. Our approach is mainly affected by vehicle's kinematics and physical road constraints. Based on these assumptions, computational time for path planning can be significantly reduced by creating a database containing already optimized versions of all the potential trajectories in each curve the vehicle can carry out. Therefore, this algorithm generates a database of smooth and continuous curves considering a big set of different intersection scenarios, taking into account the constraints of the infrastructure and the physical limitations of the vehicle. According to the real scenario, the local planner selects from the database the appropriate curves, searching for the ones that fit with the intersections defined on it. The path planning algorithm has been tested in simulation against the previous control architecture. The results obtained show path generation improvements in terms of smoothness and to continuity. Next steps on this algorithm is to test its performance in real platform and add the dynamics obstacle avoidance capabilities, establishing the link with the perception algorithms research line currently open in the team.

7.5. Control techniques

Participants: Francisco Navas Matos, Carlos Eduardo Flores Pino, David Gonzalez Bautista, Joshué Pérez Rastelli, Vicente Milanés Montero.

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 the fully automated capabilities. Having this in mind, there are two research lines currently open in the time: the first one is mainly related to what we call "naturalistic driving" in the sense of adding the human reasoning to the vehicle. We are mainly focusing our effort on artificial intelligent algorithms as neuro-fuzzy techniques. The main reason is the growing interest of the car makers in adding sharing control capabilities (between the vehicle and the driver) to the automated car. Our initial results show a big potential of using this approach and we already achieved some simulations results that were well-accepted by the scientific community and will be shown in mid-December at the final event of the EU project DESERVE.

On the other hand, we are also further investigating robust control algorithms for providing stability not only to an automated vehicle but also to a chain of automated vehicles that should be able to cooperate intelligently. This work is mainly divided in two main research lines:
1) Controllability and stability of dynamic complex systems are the key aspects when it comes to design intelligent control algorithms for vehicles. Current advances in the field are mainly oriented to advanced multi-sensor fusion toward multi-target decision-making systems. These artificial intelligence-based algorithms are able to provide reasonable responses under controlled environments (i.e. highly-detailed maps). However, new trends are proposing intelligent algorithms able to handle any unexpected circumstances as unpredicted uncertainties or even fully outages from sensors. The goal of this new research line at RITS is to further investigate control algorithms able to provide stability responses for autonomous vehicles under uncontrolled circumstances, including modifications on the input/output sensors. Dynamic plant models where different inputs/outputs can be added or subtracted in real-time during its operation is one of the hot topics in the control research arena. This system has to provide stable enough response when these operations occur. This is especially true on high-risk environments as autonomous driving; and
2) Data-driven control techniques based on model-free algorithms. Vehicles exhibit a highly non-linear behavior, especially at low speeds (as occur in urban environments). The research on novel data-driven techniques that are independent of the plant model provides huge benefits when applying them to automated vehicles. This novel research line in the team tries to further investigate on stable algorithm that doesn't need an accurate model of the vehicle dynamic, leading to compensate the effects of nonlinear dynamics, disturbances, or uncertainties in the parameters. [35]

7.6. Study on Perception and Communication Sytems for Safety

Participants: Pierre Merdrignac, Oyunchimeg Shagdar, Ines Ben Jemaa, Fawzi Nashashibi.

The existing R&D efforts for protecting vulnerable road users (VRU) are mainly based on perception techniques, which aim to detect VRUs utilizing vehicle embedded sensors. The efficiency of such a technique is largely affected by the sensor's visibility condition. Vehicle-to-Pedestrian (V2P) communication can also 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 combination of such approaches can be a solution to the VRU safety. This is the motivation of this work. We develop theoretical models to present the characteristics of perception and communications systems. Experimental studies are conducted to compare the performances of these techniques in real-world environments. Our results show that the perception system reliably detects pedestrians and other objects within 50 m of range in the line-of-sight (LOS) condition. In contrast, the V2P communication coverage is approximately 340 and 200 meters in LOS and non-LOS (NLOS) conditions, respectively. However, the communication-based system fails to correctly position the VRU w.r.t the vehicle, preventing the system from meeting the safety requirement. Finally, we propose a cooperative system that combines the outputs of the communication and perception systems. More detail can be fund in [37], [16].

7.7. Asynchronous Reactive Distributed Congestion Control Algorithms for the ITS G5 Vehicular Communications

Participant: Oyunchimeg Shagdar.

The IEEE 802.11p is the technology dedicated to vehicular communications to support road safety, efficiency, and comfort applications. A large number of research activities have been carried out to study the characteristics of the IEEE 802.11p. The key weakness of the IEEE 802.11p is the channel congestion issue, where the wireless channel gets saturated when the road density increases. The European Telecommunications Standardization Institute (ETSI) is in the progress of studying the channel congestion problem and proposed so-called Reactive Distributed Congestion Control (DCC) algorithm as a solution to the congestion issue. In this work we investigate the impacts of the Reactive DCC mechanism in comparison to the conventional IEEE 802.11p with no congestion control. Our study shows that the Reactive DCC scheme creates oscillation on channel load that consequently degrades communication performance. The results reveal that the channel load oscillation is due to the fact that in the Reactive DCC, the individual CAM (Cooperative Awareness Message) controllers react to the channel congestion in a synchronized manner. To reduce the oscillation, we propose a simple extension to Reactive DCC, Asynchronous Reactive DCC, in which the individual CAM controllers adopt randomized rate setting, which can significantly reduce the oscillation and improve the network performance. See [45] for more detail.

7.8. Vehicle to vehicle visible light communication

Participants: Mohammad Abu Alhoul, Oyunchimeg Shagdar, Fawzi Nashashibi.

Visible Light Communication (VLC) technology utilizes the light spectral range between 380 nm and 750 nm, which enables the dual functionality of lightning and information delivery. A use of VLC for the ITS domain has many benefits including that it can be a complementary technology to the IEEE 802.11p, which is the radio communications technology dedicated to the V2X communication but suffers from its channel congestion problem.

This year, we conducted theoretical and experimental studies on the optical channel characteristics. Based on our studies and the previous contributions, we developed a transmitter and receiver VLC prototype to be integrated to the vehicle lightning systems dedicated to platooning applications. Using the low-cost Arduino micro-controller, a transmitter broadcasts the vehicle status information including the vehicle identity, velocity, orientation, acceleration through the vehicle rear Light Emitting Diodes (LED). The receiver is based on a simple Photo Diode (PD) with an accurate 635 nm optical filtering stage to overcome the saturation and the unwanted ambient noise issues. Experimental studies show that the system can provide 8.5 Kbps of information delivery between vehicles with up to 30 meters of bumper to bumper distance.

7.9. Analysis of broadcast strategies in IEEE 802.11p VANETs

Participants: Younes Bouchaala, Oyunchimeg Shagdar, Paul Muhlethaler.

We analyze different broadcast strategies in IEEE 802.11p Vehicular Ad-hoc NETworks (VANETs). The first strategy is the default IEEE 802.11p strategy. Using a model derived from the Bianchi model, we provide the network performance in terms of throughput and success rate. The second strategy consists in using an acknowledgment technique similar to the acknowledgment with point-to-point traffic. A node will send its broadcast packet as in the default case, but it requires an acknowledgment from a neighbor node. This node may be a random neighbor or may be selected according to precise rules. We analyze this second strategy in terms of throughput and success rate. Somewhat surprisingly, we show that this second strategy improves the delivery ratio of the transmitted packets but reduces the overall throughput. This means that if the CAM messages (Cooperative Awareness Messages) are broadcasted, the total number of packets actually delivered will be greater with the default strategy than with the improved strategy. We propose a third strategy which consists in using the default strategy for normal packets, but we add random redundant transmissions to ensure greater reliability for very important packets. We show that with this simple technique, not only do we obtain suitable reliability, but we also achieve larger global throughput than with the acknowledgment-oriented technique. This is described in [26]. Another contribution of this paper is to compute network performance in terms of throughput and success rate with respect to the network parameters and to analyze their impact on performances.

7.10. Multicast communications for cooperative vehicular systems

Participants: Ines Ben Jemaa, Oyunchimeg Shagdar, Paul Muhlethaler, Arnaud de La Fortelle.

With the advancement of wireless communications technologies, users can now have multicast services while they are driving. Majority of the multicast services require Internet-to-vehicle multicast message dissemination. Conventional group management approaches in Internet is relatively simple because it is performed on the local networks of the multicast members which are usually a priori configured to receive the service. In addition to this, multicast packets flows follow a fixed routing structure that is built between the source and the destinations. These approaches could not be applied to vehicular networks (VANET) due to their dynamic and distributed nature. In order to enable such multicasting, our work deals with two aspects. First, reachability of the moving vehicles to the multicast service and second, multicast message dissemination in the VANET. Regarding the first issue, we find that neither current multicast addressing nor existing mobility management mechanisms are suitable for VANET. We introduce first a self-configuring multicast addressing scheme that allows the vehicles to auto-configure a dynamic multicast address without a need to exchange signalling messages with the Internet. Second, we propose a simplified approach that extends Mobile IP and Proxy Mobile IP. About message dissemination, we first propose to revisit traditional multicast routing techniques that rely on a tree structure. In particular, as vehicular networks are known to have changing topology, we present a theoretical study of the link lifetime between vehicles in urban environments. Then, we propose then Motion-MAODV, an improved version of a tree-based routing mechanism (MAODV) that aims at guaranteeing longer route lifetime. Finally, we also propose a geographic routing protocol Melody that provides a geocast dissemination in urban environments. Through simulations, we show that Melody ensures more reliable and efficient packet delivery to a given geographic area compared to traditional geo-brodcasting schemes in highly dense scenarios. More detail can be found in [28], [41], [47].

7.11. Context Awareness and Priority Control for ITS based on Automatic Speech Recognition

Participants: Oyunchimeg Shagdar, Sakriani Watiasri Sakti.

Bringing rapid assistance to motorists involved in a traffic accident is an important service to be provided by Intelligent Transportation System (ITS). Existing proposals to automatic accident detection are based on the vehicle's perception point of view. In [38] we introduce situational awareness based on the "understanding" of conversational speech of drivers/passengers using an automatic speech recognition (ASR) system. Context-aware priority control and congestion control schemes are presented to ensure coexistence of ASR-triggered applications and cooperative awareness messages (CAM) in the IEEE 802.11p system. Finally, application risk analysis and performance evaluations of ASR and V2X communications are carried out.

7.12. Emergent Behaviors and Traffic Density among Heuristically-Driven Intelligent Vehicles using V2V Communication

Participants: Oyunchimeg Shagdar, Fawzi Nashashibi.

We study the global traffic density and emergent traffic behavior of several hundreds of intelligent vehicles, as a function of V2V communication (for the ego vehicle to perceive traffic) and path-finding heuristics (for the ego vehicle to reach its destination), in urban environments. Ideal/realistic/no V2V communication modes are crossed with straight-line/towards-most-crowded/towards-least-crowded pathfinding heuristics to measure the average trip speed of each vehicle. The behaviors of intelligent vehicles are modeled by a finite state automaton. The V2V communication model is also built based on signal propagation models in an intersection scenario and a Markov-chain based MAC model. Our experiments in simulation over up to 400 vehicles exhibit attractive insights: 1) communication's impact is positive for the performance of the emergent vehicles' behavior, however, 2) the path-finding heuristics may not obtain their expected collective behavior due to the communications errors in realistic road environment (cf. [43]).

7.13. Time-bounded message dissemination in strings

Participant: Gérard Le Lann.

In 2015, besides reviewing prominent open issues regarding safety in IVNs (see [42]), we have investigated coordination problems that arise in string formations. Since the inception of the platoon concept (1977), a number of solutions have been proposed for achieving string control (platoons are a particular case of ad hoc/open string). String control must be exercised in order to avoid rear-end collisions, string instability, and for coping with emergency situations. The cyber components essential for string control have not been fully identified yet. For example, considering the cooperative adaptive cruise control paradigm, data collected in recent platooning experiments show that it is inappropriate to rely on V2V broadcast from a lead vehicle, thus the quest for other approaches. In strings, one can take advantage of short-range directional antennas which enable fast messaging among consecutive string neighbors, leading to the concept of neighbor-to-neighbor (N2N) communications and the cohort construct (a cohort is a string with a specification). String control problems translate into communication protocol issues and distributed algorithmic problems, notably:

- Time-bounded string-wide acknowledged message delivery and dissemination (TBMD),
- Bounded channel access delay (BCAD), a MAC-level problem,
- Time-bounded message acknowledgment (TBMA).

Acceptable solutions shall achieve small non-stochastic worst-case channel access time bounds (BCAD) and bounded delays for successful message delivery (TBMA and TBMD), under worst-case conditions regarding channel contention and message/acknowledgment losses. Non-stochastic worst-case bounds can only be established analytically (obviously, simulations cannot be considered). The importance of the TBMD problem can be exposed simply as follows: would TBMD be solved, then the string instability problem vanishes. Rather than resting solely on stepwise detection-and-reaction strategies based on radars/lasers, every string member adjusts its acceleration/deceleration rate according to observed motions of its predecessor, TBMD delivers a N2N message carrying the newly string-wide targeted velocity, in less than 100 milliseconds in strings comprising in the order of 20 members, in the presence of message/acknowledgment losses. The TBMD problem has been solved (see [34]). The solution rests on assuming that TBMA and TBMD have solutions. Both problems have been solved (solutions are under review). Contrary to strings, groups are ad hoc/open multilane formations. It turns out that solutions aimed at the 3 problems referenced above are instrumental in solving problems arising with multilane SC scenarios. For example, the 3-way handshakes at the core of safe lane changes published previously now achieve significantly better performance figures. Work in progress also includes:

- conflicting concurrent lane changes at high velocities,

- fully automated zipper merging at high velocities, in non-line-of-sight conditions (radio communications), in line-of-sight conditions (optical communications).

7.14. Broadcast Transmission Networks with Buffering

Participants: Guy Fayolle, Paul 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 [44].

7.15. Belief propagation inference for traffic prediction

Participants: Cyril Furtlehner, Jean-Marc Lasgouttes.

This work [51] 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.

These studies have been done in particular in the framework of the projects Travesti and Pumas.

This year, the work about the theoretical aspects of encoding real valued variables into a binary Ising model has been accepted for publication in Annals of Mathematics and Artificial Intelligence [23]. Moreover, an informal collaboration has been started with the company SISTeMA ITS, in order to assess the performance of our techniques in real-world city networks.

7.16. Random Walks in Orthants

Participant: Guy Fayolle.

7.16.1. Explicit criterion for the finiteness of the group in the quarter plane

In the book [3], original methods were proposed to determine the invariant measure of random walks in the quarter plane with small jumps, the general solution being obtained via reduction to boundary value problems. Among other things, an important quantity, the so-called *group of the walk*, allows to deduce theoretical features about the nature of the solutions. In particular, when the *order* of the group is finite, necessary and sufficient conditions have been given in [3] for the solution to be rational or algebraic. When the underlying algebraic curve is of genus 1, we propose, in collaboration with R. Iasnogorodski (St-Petersburg, Russia), a concrete criterion ensuring the finiteness of the group. It turns out that this criterion is always tantamount to the cancellation of a single constant, which can be expressed as the determinant of a matrix of order 3 or 4, and depends in a polynomial way on the coefficients of the walk [20].

7.16.2. Second Edition of the Book Random walks in the Quarter Plane

In collaboration with R. Iasnogorodski (St-Petersburg, Russia) and V. Malyshev, we prepared the second edition of the book [3], which will be published by Springer, in the collection *Probability Theory and Stochastic Processes*. Part II of this second edition borrows specific case-studies from queueing theory, and enumerative combinatorics. Five chapters will be added, including examples and applications of the general theory to enumerative combinatorics. Among them:

- Explicit criterions for the finiteness of the group, both in the genus 0 and genus 1 cases.
- Chapter *Coupled-Queues* shows the first example of a queueing 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.

7.17. Global optimization for online resource allocation

Participant: Jean-Marc Lasgouttes.

As part of the Mobility 2.0 FP7 project, we have considered the possibility to allocate charging stations to Full Electric Vehicle (FEV) users in a way that, instead of merely minimizing their travel time, tries to improve the travel time for the whole community.

Our setting can be seen as a resource allocation problem, known as the Transportation Problem in Operations Research literature. It is solvable using several algorithms, among which the simplex algorithm or the Hungarian algorithm. Unfortunately, these algorithms are not well-adapted here for two reasons:

- The allocation of slots to users is done on-line, when the user does a request. It is not possible to wait until all the users are known before doing the allocation;
- The complexity of these algorithms is very high, especially since, due to the effect of range limitations, each request has different characteristics, which is equivalent to increasing the types of customers.

We therefore present a simple heuristic approach, which is fast enough for systems with thousands of stations. Its principle is to penalize the cost for the user with an approximation of the extra cost incurred to future users who compete for the same resource (a charging or parking slot).

This work has been presented at the ITSC'2015 conference [33].

AYIN Team

7. New Results

7.1. Markov Random Fields

7.1.1. New hierarchical joint classification method of SAR and optical multiresolution remote sensing datas

Participants: Ihsen Hedhli, Josiane Zerubia [contact].

This work was carried out in collaboration with Prof. Gabriele Moser and Prof. Sebastiano Serpico from DITEN departement [www.diten.unige.it/], University of Genoa, Italy.

Nowadays, a wide variety of remote sensing images is available. Therefore, it becomes more and more important to be able to analyze compound data sets consisting of different types of images acquired by different sensors, as they allow a spatially distributed and temporally repetitive view of the monitored area at the desired spatial scales. In particular, the opportunity of joint availability of synthetic aperture radar (SAR) and optical images offers high resolution (HR), all-weather, day/night, short revisit time data, as well as polarimetric and multifrequency acquisition capabilities. Similarly, the strong differences in terms of wavelength range (microwave vs. visible and near infrared), sensitivity to cloud cover and sun illumination (strong for optical imagery vs. almost negligible for SAR), and noise-like properties (speckle in SAR vs. generally low noise variance in current HR optical sensors) make the joint use of HR optical and SAR imagery especially interesting for many applications to environmental monitoring and natural risk management. Within this framework, there is a definite need for classification methods that automatically correlate different sets of images taken on the same area from different sensors and at different resolutions. This year we developed a novel classification approach for multiresolution, multisensor (optical and synthetic aperture radar), and/or multiband images. Accurate and time-efficient classification methods are particularly important tools to support rapid and reliable assessment of the ground changes. Given the huge amount and variety of data available currently from last-generation satellite missions, the main difficulty is to develop a classifier that can benefit from multiband, multiresolution, and multisensor input imagery. As shown in Figure 1 , the proposed method addresses the problem of multisensor fusion of SAR with optical data for classification purposes, and allows input data collected at multiple resolutions and additional multiscale features derived through wavelets to be fused. The proposed approach formalizes a supervised Bayesian classifier within a multiple quadtree topology that combines a class-conditional statistical model for pixel-wise information and a hierarchical Markov random field (MRF) for multisensor and multiresolution contextual information.

7.2. Marked point processes

7.2.1. Integrating RJMCMC and Kalman filters for multiple object tracking

Participants: Paula Craciun, Josiane Zerubia [contact].

This work has been done in collaboration with Dr. Mathias Ortner from Airbus D&S [http://www.space-airbusds.com/fr/]

Recently, we have proposed a new spatio-temporal marked point process model for tracking small, rigid objects in high resolution images. We have shown very good detection and tracking results for synthetic biological data as well as remotely sensed sequences. The model is based on defining a dedicated energy function that is highly non-convex. The solution is found by minimizing this energy function using a suitable batch-optimization scheme based on Reversible Jump Markov Chain Monte Carlo (RJMCMC) sampler. This approach is motivated by the low temporal frequency of the sequences (1Hz).

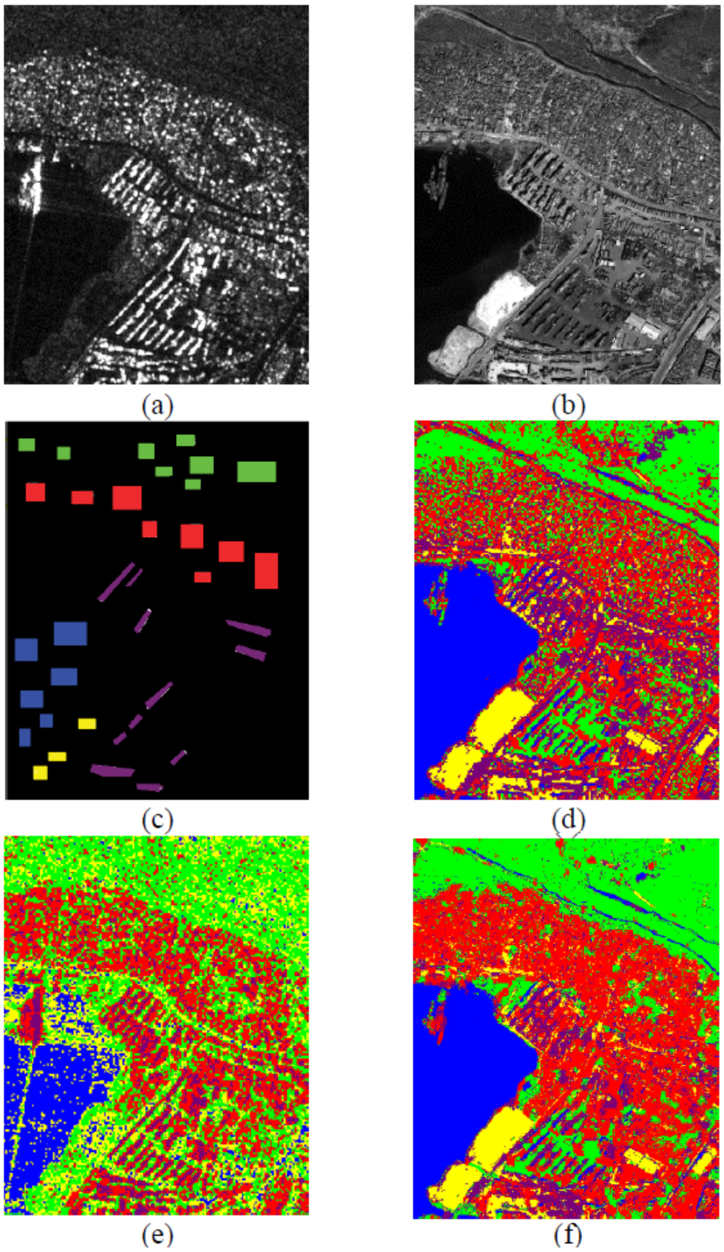


Figure 1. (a) SAR image (© ASI), (b) one channel from the optical image (© GeoEye), (c) the available ground truth, (d) hierarchical MRF-based classification obtained from the optical image, using Laferté method, (e) hierarchical MRF-based classification obtained for the SAR image, using Laferté method, (f) hierarchical MRF-based classification obtained by the proposed method.

Sequential filters have proven to provide relatively fast and reliable tracking performances in particular for single target tracking. We have efficiently exploited the properties of sequential filters within the RJMCMC sampling scheme. The filter is used to generate more meaningful perturbation proposals which are then evaluated using an appropriate Green acceptance ratio. Better perturbation proposals increase the acceptance probability of the overall RJMCMC sampling scheme which in turn leads to a faster convergence.
Figure 2 shows the detection and tracking results on two synthetic biological sequences as well as on two sequences of simulated satellite images of Toulon by courtesy of Airbus Defence & Space, France. The evolution of the energy with the number of iterations for the standard RJMCMC sampler and the proposed sampler is also shown. The proposed sampler is depicted blue.

7.2.2. Initialization and estimation of parameters for marked point processes applied to automatic object detection on satellite images

Participants: Aurélie Boisbunon, Josiane Zerubia [contact].

This work has been done in collaboration with Dr. Rémi Flamary (Université de Nice Sophia Antipolis), Prof. Alain Rakotomamonjy (Université de Rouen) et Alain Giros (CNES). It was partially funded by the French Spatial Agency CNES [http://www.cnes.fr].

Sparse representations, large scale, stochastic algorithms, machine learning, image processing Marked point processes (MPP) strongly rely on parameters, whose estimation affects both computation time and performances. In this work, we proposed two approaches: the first one consists in initializing MPPs with a first coarse solution obtained very quickly from sparse regularization methods, while the second one estimates the parameters by the Stochastic Approximation Expectation-Maximization (SAEM) algorithm. We give details on both approaches below. The first coarse solution is obtained from a deterministic sparse regularization method. This method is based on the representation of an image with objects as a sum of convolutions between atoms of a dictionary and matrices of positions of the objects in the image. The atoms of the dictionary are fixed in advance and correspond to different instances of the objects (scales, angles, shapes, etc). This way, we transform the problem of object detection into the problem of estimating extremely sparse matrices. The algorithm we derived for solving the associated optimization problem is both parallelized and very efficient. This work started last year, and continued this year by conducting more tests.

7.2.3. Generic curvilinear structure modeling via marked point process theory

Participants: Seong-Gyun Jeong, Yuliya Tarabalka, Josiane Zerubia [contact].

This work has been done in collaboration with Dr. Nicolas Nisse (COATI team [https://team.inria.fr/coati/], Inria-SAM) and Dr. Yuliya Tarabalka (Titane team [https://team.inria.fr/titane/team/], Inria-SAM)

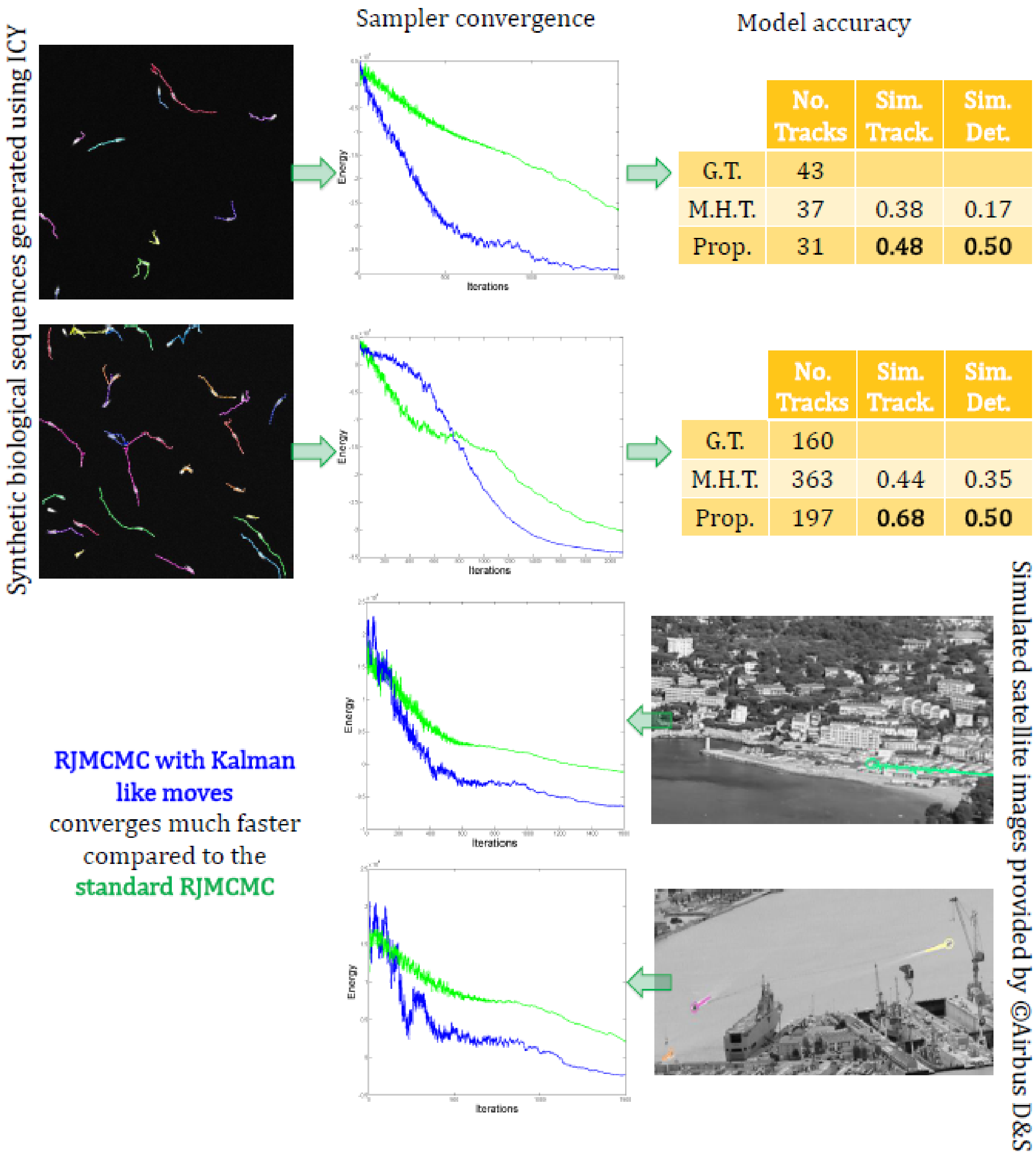
We propose a novel curvilinear structure reconstruction algorithm based on ranking learning and graph theory. In this work we reconstruct the curvilinear structure as a set of small line segments (via MPP). Specifically, we infer the structured output ranking of the line segments via Structured Support Vector Machine(SSVM). To predict the existence of the curvilinear structure, we measure oriented image gradient maps and morphological profiles. We propose an orientation-aware curvilinear feature descriptor and a feature grouping operator to improve the structural consistency for learning system. In order to provide topological information, we develop a graph-based curvilinear structure reconstruction algorithm. The proposed algorithm builds a graph based on the output ranking scores and searches the longest geodesic paths which are associated with the latent curvilinear structure. Experimental results (see Figure 3) show that the proposed algorithm faithfully detects the curvilinear structures and preserves topological information compared with the competing algorithms.

7.3. Other approaches

7.3.1. Acne detection on polarized or non-polarized images

Participants: Zhao Liu, Josiane Zerubia [contact].

This work is in collaboration with Dr. Queille-Roussel and Prof. Bahadoran in CHU Nice, France. Now Dr. Zhao Liu is a post-doc at Manchester University [www.manchester.ac.uk/], Manchester, UK.



	No. Tracks	Sim. Track.	Sim. Det.
G.T.	43		
M.H.T.	37	0.38	0.17
Prop.	31	**0.48**	**0.50**

	No. Tracks	Sim. Track.	Sim. Det.
G.T.	160		
M.H.T.	363	0.44	0.35
Prop.	197	**0.68**	**0.50**

Figure 2. Tracking results and sampler convergence on two synthetic biological sequences (generated using ICY [http://icy.bioimageanalysis.org/], a free software offered by the Quantitative Analysis Unit from the Pasteur Institute, France) as well as two sequences of simulated satellite images of Toulon (by courtesy of Airbus Defence & Space, France). The RJMCMC sampler with Kalman like moves (shown in blue) requires a significantly lower number of iterations until convergence as compared to the standard RJMCMC.

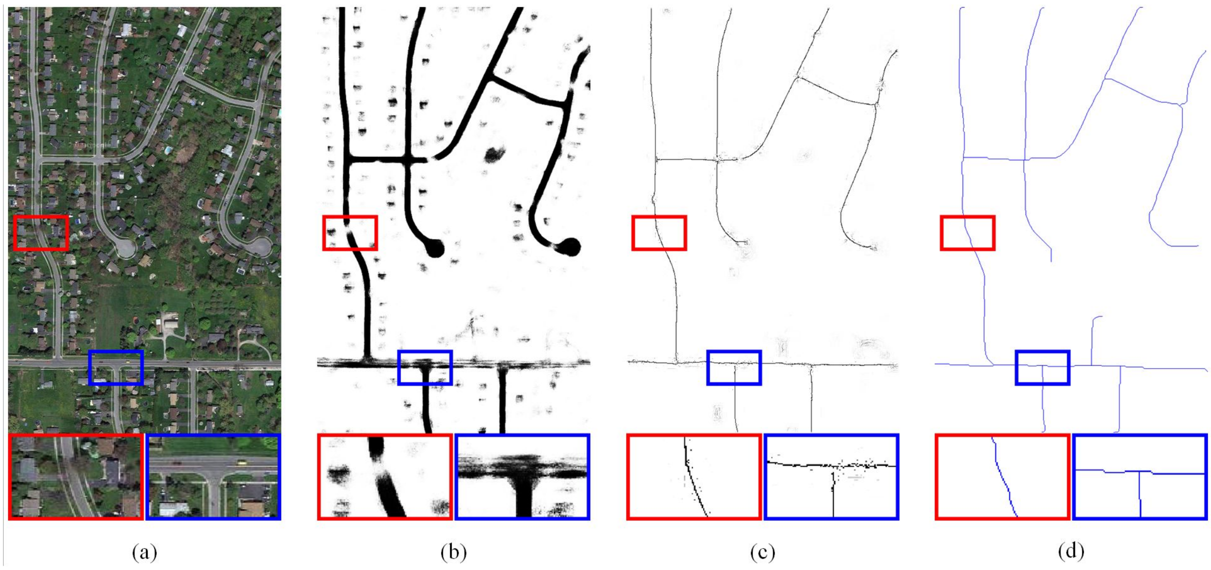


Figure 3. Compared with (b) the segmentation and (c) the centerline detection methods, (d) the proposed algorithm well represents topological features of the curvilinear structure. Setting a threshold value yields to lose correlated information of the pixels on the reconstructed curvilinear structure. In this example, road network is partially occluded by trees or cars, so that the local measure often fails to detect the underlying curvilinear structures. Although the centerline is able to quantify scale (width) of curvilinear structure, it is inaccurate to classify pixels around junctions. In this work we learn spatial patterns of the curvilinear structures with structured output ranking scores. We also propose a graph-based representation algorithm to obtain the topological information.

This work is in collaboration with Dr. Queille-Roussel and Prof. Bahadoran in CHU Nice, France.

Acne vulgaris, a highly prevalent skin disease, has a significant life quality impact on patients. It is generally believed that this type of skin disorder results from proliferation of propionibacterium acnes in pilosebaceous units, which can lead to inflammatory lesions due to increase of oxyhemoglobin level. So far there is no golden standard for acne diagnosis in clinics. It entirely depends on dermatologists' experience for acne assessment. But significant variability among individual diagnosis may lead to less trustworthy results, and less reproducibility of human evaluation makes the comparison of acne follow-up difficult. This work, incorporating the knowledge of optical characteristics of human skin, identifies cutaneous chromophore distribution using bilateral decomposition. Then the inflammatory acne lesions are detected by a Markov random field (MRF) model associating the chromophore descriptors. Experimental results (see Figure 4) show that the proposed method is robust to large dynamic range intensity, and the derived automatic segmentation of inflammatory acne appears to be highly consistent to human visual assessment. This research work was started in 2013. This year, more tests have been conducted on polarized and non-polarized images.

Figure 4. Acne detection using proposed method. (Left) Original image provided by CHU Nice. (Right) Acne detection result.

7.3.2. Finer registration of facial wrinkles in time series images

Participants: Nazre Batool, Josiane Zerubia [contact].

Dr. Batool was funded by the Inria-DPEI fellowship for the period Feb. 2014 – May 2015. Currently she is a postdoc researcher at CMIV, Linköping University [www.liu.se/cmiv], Linköping, Sweden.

The goal of this work is to evaluate quantitatively the subtle variations in facial wrinkles for the same subject in response to treatment using image-based analysis. Any image analysis technique for the analysis of such subtle image variations would require high accuracy and precision for good performance. As in other imaging problems geared towards detection of temporal changes, accurate registration of key image features (wrinkles) is mandatory as a first step. We propose to compare image features in key wrinkle sites only while excluding the noise introduced by changes in surrounding skin texture. Therefore, previously we proposed a 2-step registration algorithm where the initial registration was based on the alignment of facial landmarks such as corners of eyes, nose, and mouth. Then a method based on Large Deformation Diffeomorphic Metric Mapping (LDDMM) was used to achieve finer local registration for wrinkles. However, the LDDMM algorithm had the shortcoming of the unavailability of time invariant finer facial landmarks and that the deformations were guided by image intensities which were varying among images as well due to subtle changes in skin texture. The deformation of skin due to underlying movement can be categorized loosely as locally rigid because the local skin texture remains constant but globally non-rigid because of the movement of skin areas due to slight expression and misalignment. Due to this dual nature of deformation, registration schemes such as thin plate spline or affine transformations are not applicable. Our improved approach is to guide the LDDMM registration on skin features with higher intensity gradients only (such as due to moles, wrinkles, rough surface) which have the higher probability of being constant and detected across temporal changes. First we detect key landmarks and landmark correspondences using the Gabor feature images where the phase correlation is used to find estimates of landmark correspondences. The phase correlation is based on the well-known Fourier shift property i.e. a shift in the spatial domain of two images results in a linear phase difference in the frequency domain of their respective Fourier Transforms. Figure 5 shows Gabor features of two images captured 4 weeks apart in (a) and (b). Figure 5 (c) shows key landmarks placed at high Gabor amplitude sites and (d) shows their corresponding landmarks detected using Fourier phase correlation.

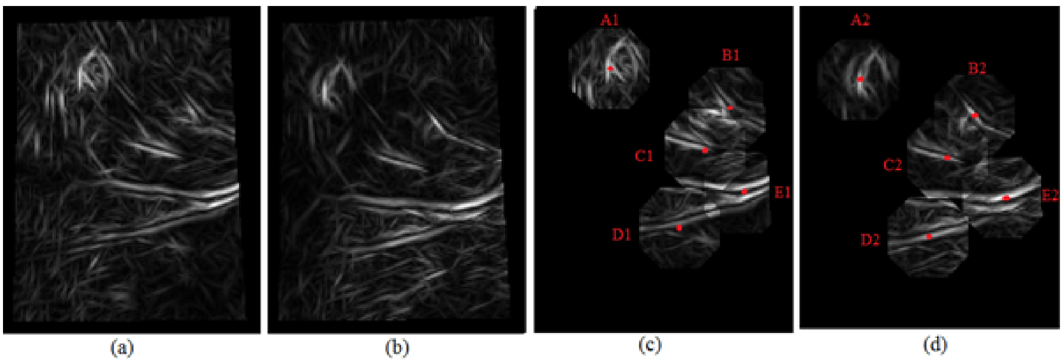


Figure 5. Detection of landmarks and correspondences. (a) Gabor response for source image. (b) Gabor response for target image. (c) 5 landmarks with circular templates in source image. (d) Corresponding detected landmarks with circular templates in target image.

Then, as a next step, the detected key landmarks and their corresponding positions are used in the landmark based LDDMM algorithm to find locally non-rigid deformations between two images. Figure 6 shows an example where the corresponding landmarks are shown as black dots in (a) and (b). Fig. 5 (c) shows the image in (a) wrapped to (b) using LDDMM based on landmark correspondences. In (d) the drifts of landmarks are shown during the LDDMM algorithm and (e) shows the non-rigid deformation of underlying image grid. In the future, the proposed wrinkle registration algorithm will be used to compare wrinkle intensities in time series of images to quantify very minute changes in wrinkles in response to dermatological treatments.

7.3.3. Hyperspectral Image Processing for Detection and Grading of Skin Erythema

Participants: Ali Madooei, Josiane Zerubia [contact].

Ali Madooei worked at Inria Sophia Antipolis on an internship funded by the Canadian Mitacs Globalink Research Award & Inria. He is currently in his last year of PhD at Simon Fraser [www.sfu.ca/] University, Canada. This work has been conducted in collaboration with Ramy M. Abdlaty, Lilian Doerwald-Munoz, Dr. Joseph Hayward and Prof. Qiyin Fang from Mc Master university [http://future.mcmaster.ca/]/Juravinsky cancer center [www.jcc.hhsc.ca/], Canada, and Prof. Joseph Hayward from Simon Fraser University, Canada.

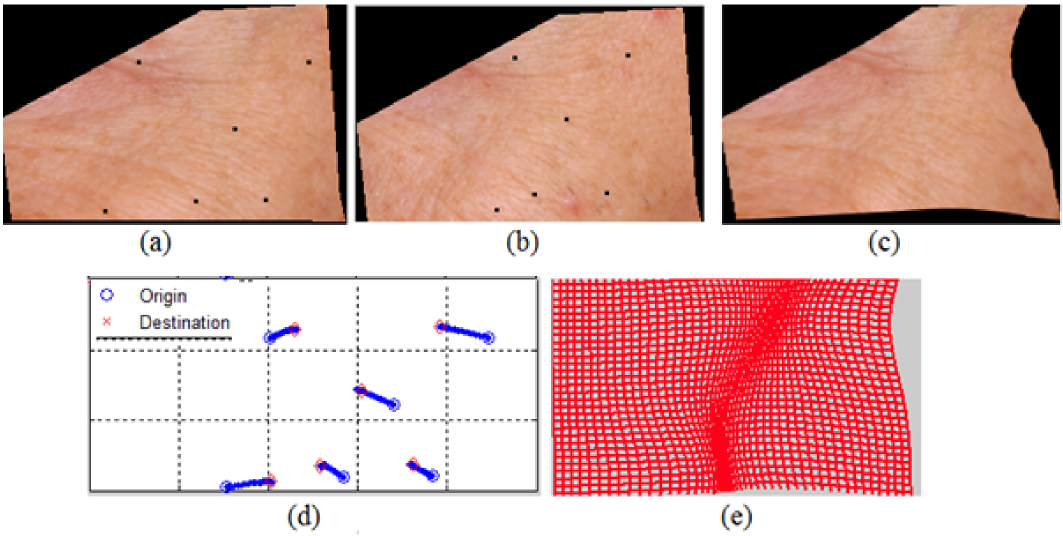


Figure 6. Finer registration of micro features in skin images. (a) Source image with landmarks shown as black dots. (b) Target image with landmarks. (c) Source image wrapped to target image. (d) Path of landmarks during LDDMM registration. (e) Deformation of underlying grid during LDDMM registration.

Acute skin erythema is a common side effect in patients undergoing radiotherapy treatment. It displays itself as an increase in skin redness and irritation. Erythema has been reported to correlate to individual patient response to radiation and therefore may be useful to guide and modify courses of treatment in a timely manner. Currently, upon visual examination, a qualitative score can be assigned to characterize the severity of erythema, which then may be used for assessing radiation response. Due to the subjective nature of this method, additional non-invasive techniques are needed for more accurate evaluation. Previous studies have mainly focused on tissue reflectance spectroscopy or imaging photography. The former retrieves spectral information from point measurements while the latter is obtained with conventional Red, Green, Blue(RGB) colour cameras. Photography has the advantage of offering spatial information but this comes at the cost of losing much of spectral information. We use hyperspectral imaging (HSI) which provides both spatial and spectral representation of the affected area. A hyperspectral camera effectively divides the spectrum into very many thin image slices (the actual number depending on the camera and application see Fig.7). This fine-grained slicing reveals spectral structure that may not be evident to the eye or to an RGB camera but can provide a rich set of information for image processing. As an emerging imaging modality for medical applications, the combination of HSI devices with adequate image processing techniques offers the perfect landscape for developing new methods for noninvasive disease monitoring and diagnosis.

The purpose of our study was to investigate the possibility of monitoring the degree of erythema using HSI data. 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. 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 evaluated the system against the clinical assessment of an experienced clinician. We also compared the performance to that of using digital photography (instead of HSI). The results from this preliminary study are encouraging and indicate that hyperspectral image data do contain relevant information, and indeed outperform imaging photography. In the future, we want to

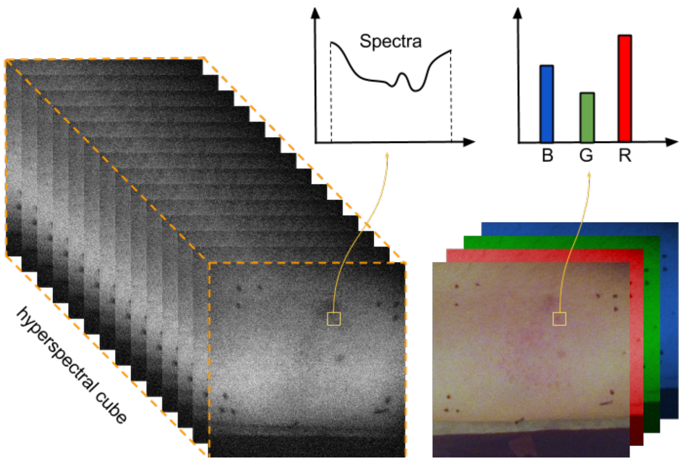


Figure 7. A schematic representation of hyperspectral vs. RGB image data.

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. Our ultimate objective is to build a system for monitoring radiation response in individuals using HSI technology and image processing.

7.3.4. SAR data classification using generalized Gamma mixture model

Participants: Vladimir Krylov, Josiane Zerubia [contact].

Vladimir Krylov is a former AYIN post-doc, now post-doc at DITEN department, University of Genoa [www.diten.unige.it/], Italy. This work has been performed in collaboration with Prof. Heng-Chao Li, Prof. Ping-Zhi Fan (Southwest Jiaotong University, Chengdu [english.swjtu.edu.cn/], China) and Prof. William Emery (University of Colorado [www.colorado.edu/], Boulder, USA).

The accurate statistical modeling of synthetic aperture radar (SAR) images is a crucial problem in the context of effective SAR image processing, interpretation and application. In this work a semi-parametric approach is designed within the framework of finite mixture models based on the generalized Gamma distribution (GΓD) in view of its flexibility and compact analytical form. Specifically, we have developed a generalized Gamma mixture model (GΓMM) to implement an effective statistical analysis of high-resolution SAR images and proved the identifiability of such mixtures. A low-complexity unsupervised estimation method has been derived by combining the proposed histogram-based expectation-conditional maximization algorithm and the Figueiredo-Jain mixture estimation algorithm. This resulted in a numerical maximum likelihood (ML) estimator that can simultaneously determine the ML estimates of component parameters and the optimal number of mixture components. The state-of-the-art performance of the proposed method has been validated experimentally on a wide range of high-resolution SAR amplitude and intensity images.

In Fig. 8 we demonstrate a typical result of the developed statistical modeling technique on a portion of a 2 meter resolution L-band image acquired by an airborne EMISAR system. The unsupervised GΓMM estimate contains five components and reports a very accurate result that outperforms the considered benchmark statistical modeling methods. In order to visualize the estimated five statistical components we also report a maximum likelihood classification map.

7.3.5. Multitemporal image change detection with a False Discovery Rate approach

Participants: Vladimir Krylov, Josiane Zerubia [contact].

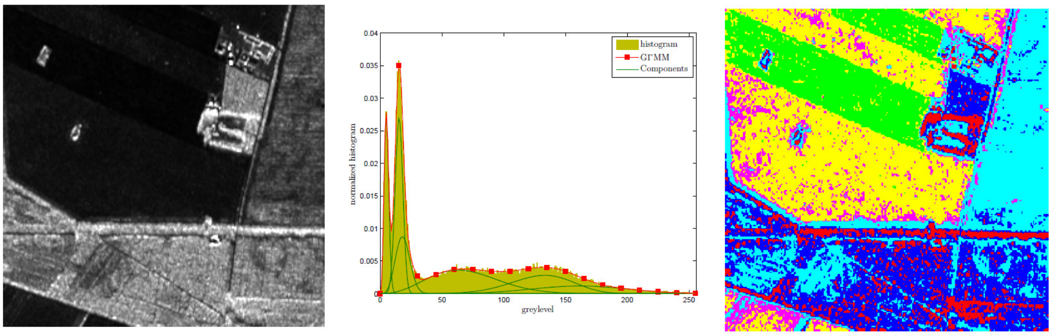


Figure 8. Statistical modeling of a EMISAR (©ESA) image (left) by generalized Gamma mixture model (middle) and its visualization by maximum likelihood classification (right).

This work has been performed in collaboration with Prof. Sebastiano Serpico and Prof. Gabriele Moser, DITEN department, University of Genoa [www.diten.unige.it/], Italy.

Multitemporal change detection is one of the fundamental image processing problems and multiple detection, monitoring and tracking applications rely on its accurate and timely performance. In this work we address the problem of unsupervised change detection on two or more coregistered images of the same object or scene at several time instants. The designed method is appropriate for short image sequences with a relatively small amount of changes. Such analysis is instrumental in various applications where acquisitions are relatively sparse and report limited meaningful changes, in particular, in remote sensing and medical image processing. We develop a novel patch-based hypothesis testing approach which is based on a false discovery rate formulation for statistical significance testing. This alternative error metric allows to adjust the family-wise error rate by imposing control over the proportion of the false positives in the detection. The designed change detector allows the use of various statistical features. The appropriate choice of the latter enables the detector to address application-specific detection problems with a particular set of disturbance factors, like noise, illumination variation, etc. In particular, we demonstrate the use of two rank-based statistics for change detection on image pairs and one multisample statistic for the analysis of image sequences. The experiments with remotely sensed radar, dermatological, and still camera surveillance imagery demonstrate competitive performance and flexibility of the proposed method.

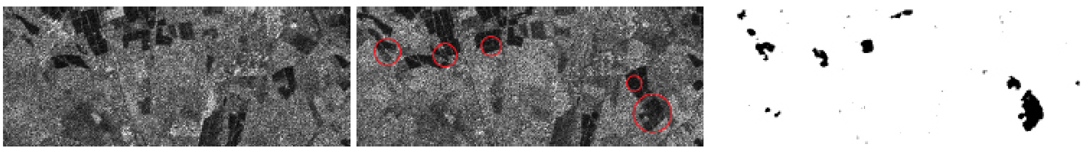


Figure 9. Change detection on a pair of 15 meter resolution XSAR images (first and second) obtained with a false discovery rate error metric based on the Cramer-von Mises statistic. The changes are highlighted with red circles (second), and the unsupervised detection result is reported in black (third).

A typical result obtained with the proposed change detection technique is reported in Fig. 9 . The proposed approach gives a unified statistical thresholding procedure to perform change detection based on statistical features that have a known distribution under the no-change hypothesis. This approach is essentially non-parametric and is highly parallelizable.

LEAR Project-Team

7. New Results

7.1. Visual recognition in images

7.1.1. *Weakly Supervised Object Localization with Multi-fold Multiple Instance Learning*

Participants: Ramazan Cinbis, Cordelia Schmid, Jakob Verbeek.

Object category localization is a challenging problem in computer vision. Standard supervised training requires bounding box annotations of object instances. This time-consuming annotation process is sidestepped in weakly supervised learning. In this case, the supervised information is restricted to binary labels that indicate the absence/presence of object instances in the image, without their locations. In [26], we propose to follow a multiple-instance learning approach that iteratively trains the detector and infers the object locations in the positive training images. Our main contribution is a multi-fold multiple instance learning procedure, which prevents training from prematurely locking onto erroneous object locations. Compared to state-of-the-art weakly supervised detectors, our approach better localizes objects in the training images, which translates into improved detection performance. Figure 1 illustrates the iterative object localization process on several example images. The technical report [26] is a journal paper under review after minor revision which extends a previous conference publication by adding experiments with CNN features, and a refinement procedure for the object location inference. These additions improve over related work that has appeared since the publication of the original paper.

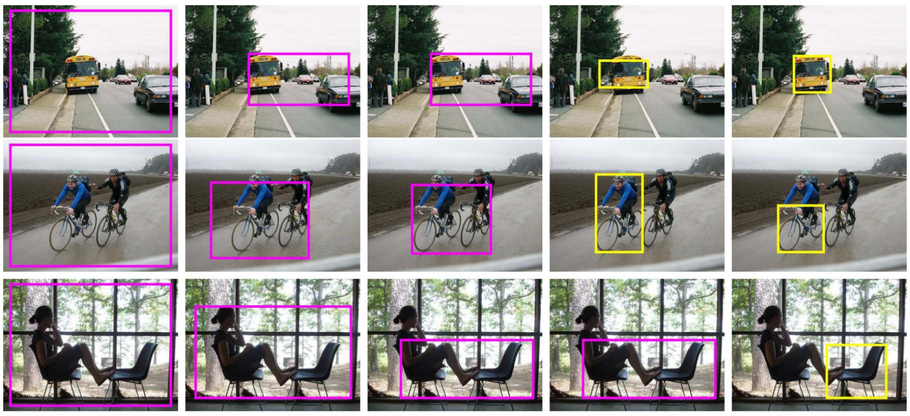


Figure 1. Illustration of our iterative object localization process on several example images, from initialization (left) to final localization (right). Yellow bounding boxes indicate that the object location hypothesis is in agreement with the ground-truth, for pink boxes the hypothesis is incorrect.

7.1.2. *Patch-level spatial layout for classification and weakly supervised localization*

Participants: Valentina Zadrija [University of Zagreb], Josip Krapac [University of Zagreb], Jakob Verbeek, Sinisa Segvic [University of Zagreb].

In [24] we propose a discriminative patch-level spatial layout model suitable for learning object localization models with weak supervision. We start from a block-sparse model of patch appearance based on the normalized Fisher vector representation. The appearance model is responsible for i) selecting a discriminative subset of visual words, and ii) identifying distinctive patches assigned to the selected subset. These patches are further filtered by a sparse spatial model operating on a novel representation of pairwise patch layout. We have evaluated the proposed pipeline in image classification and weakly supervised localization experiments on a public traffic sign dataset. The results show significant advantage of the proposed spatial model over state of the art appearance models.

7.1.3. Approximate Fisher Kernels of non-iid Image Models for Image Categorization

Participants: Ramazan Cinbis, Cordelia Schmid, Jakob Verbeek.

The bag-of-words (BoW) model treats images as sets of local descriptors and represents them by visual word histograms. The Fisher vector (FV) representation extends BoW, by considering the first and second order statistics of local descriptors. In both representations local descriptors are assumed to be identically and independently distributed (iid), which is a poor assumption from a modeling perspective. It has been experimentally observed that the performance of BoW and FV representations can be improved by employing discounting transformations such as power normalization. In [5], an expanded version of a previous conference publication, we introduce non-iid models by treating the model parameters as latent variables which are integrated out, rendering all local regions dependent. Using the Fisher kernel principle we encode an image by the gradient of the data log-likelihood w.r.t. the model hyper-parameters. Our models naturally generate discounting effects in the representations; suggesting that such transformations have proven successful because they closely correspond to the representations obtained for non-iid models. To enable tractable computation, we rely on variational free-energy bounds to learn the hyper-parameters and to compute approximate Fisher kernels. Our experimental evaluation results validate that our models lead to performance improvements comparable to using power normalization, as employed in state-of-the-art feature aggregation methods.

7.1.4. Local Convolutional Features with Unsupervised Training for Image Retrieval

Participants: Mattis Paulin, Matthijs Douze, Zaid Harchaoui, Julien Mairal, Florent Perronnin [Facebook], 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[17], 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. 3 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. Learning and statistical models

7.2.1. A Universal Catalyst for First-order Optimization

Participants: Hongzhou Lin, Julien Mairal, Zaid Harchaoui.

In this paper [16], we introduce a generic scheme for accelerating first-order optimization methods in the sense of Nesterov, which builds upon a new analysis of the accelerated proximal point algorithm. Our approach consists of minimizing a convex objective by approximately solving a sequence of well-chosen auxiliary problems, leading to faster convergence. This strategy applies to a large class of algorithms, including gradient descent, block coordinate descent, SAG, SAGA, SDCA, SVRG, Finito/MISO, and their proximal variants. For all of these methods, we provide acceleration and explicit support for non-strongly convex objectives. In addition to theoretical speed-up, we also show that acceleration is useful in practice, as illustrated in Figure 4 , especially for ill-conditioned problems where we measure significant improvements.

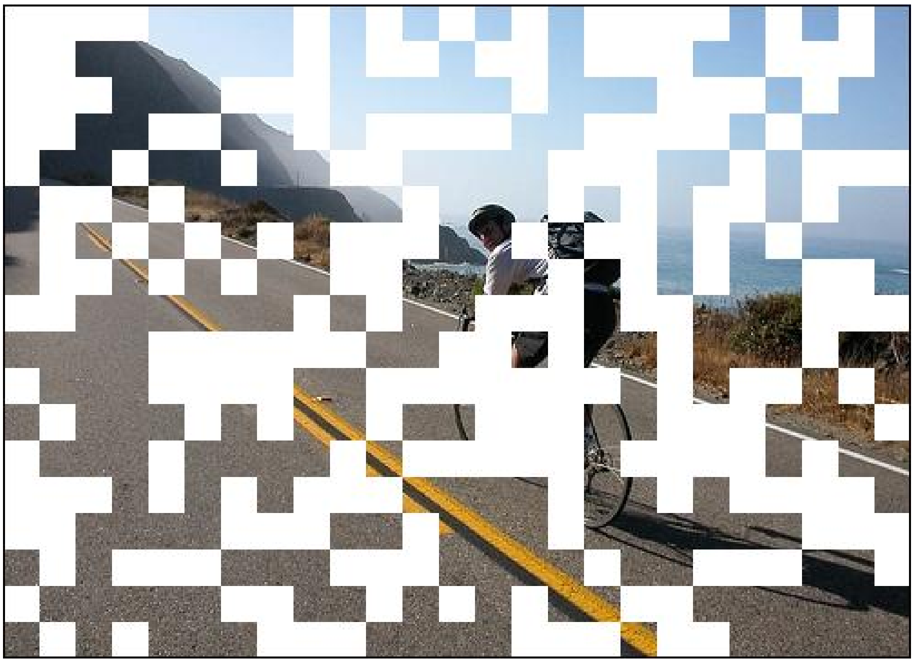


Figure 2. Illustration of why local image patches are not independent: we can easily guess the image content in the masked areas.

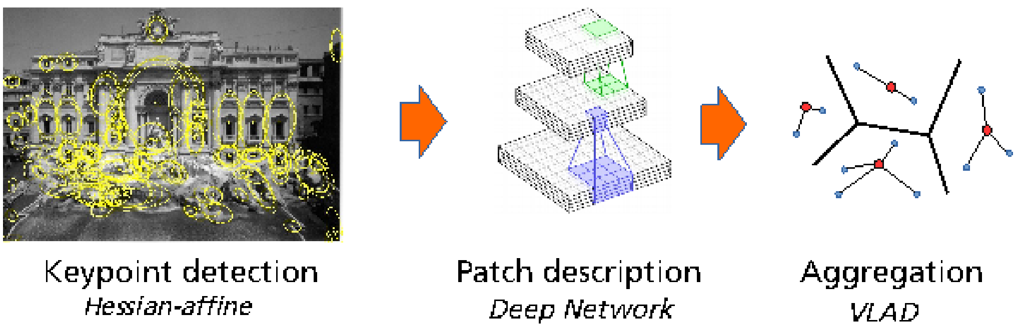


Figure 3. Image retrieval pipeline. Interest points are extracted with the Hessian-affine detector (left), encoded in descriptor space using convolutional features (middle), and aggregated into a compact representation using VLAD-pooling (right).

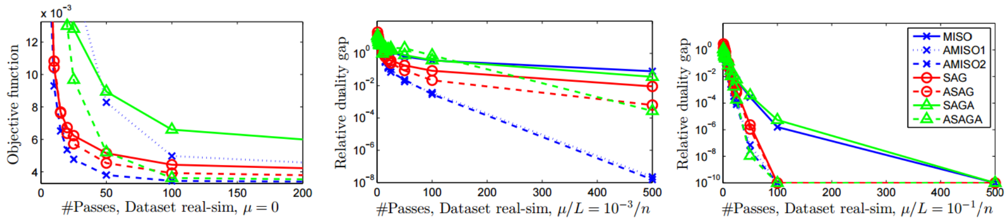


Figure 4. Objective function value (or duality gap) for different number of passes performed over each dataset. The legend for all curves is on the top right. AMISO, ASAGA, ASAG refer to the accelerated variants of MISO, SAGA, and SAG, respectively.

7.2.2. Incremental Majorization-Minimization Optimization with Application to Large-Scale Machine Learning

Participant: Julien Mairal.

In this paper [7], we study optimization methods consisting of iteratively minimizing surrogates of an objective function, as illustrated in Figure 5 . We introduce a new incremental scheme that experimentally matches or outperforms state-of-the-art solvers for large-scale optimization problems typically arising in machine learning.

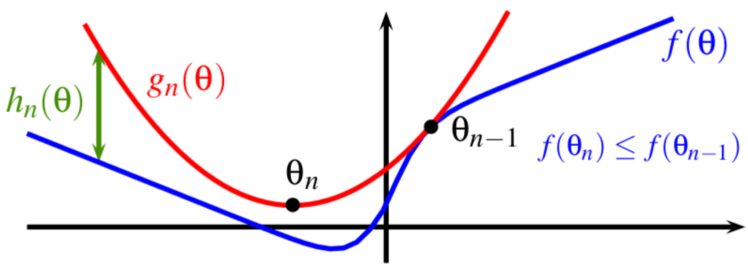


Figure 5. Illustration of the basic majorization-minimization principle. We compute a surrogate g_n of the objective function f around a current estimate θ_{n-1}. The new estimate θ_n is a minimizer of g_n. The approximation error h_n is smooth.

7.2.3. Coordinated Local Metric Learning

Participants: Shreyas Saxena, Jakob Verbeek.

Mahalanobis metric learning amounts to learning a linear data projection, after which the ℓ_2 metric is used to compute distances. In [20], we develop local metric learning techniques which allow more flexible metrics, not restricted to linear projections, see 6 . Most of these methods partition the data space using clustering, and for each cluster a separate metric is learned. Using local metrics, however, it is not clear how to measure distances between data points assigned to different clusters. In this paper we propose to embed the local metrics in a global low-dimensional representation, in which the ℓ_2 metric can be used. With each cluster we associate a linear mapping that projects the data to the global representation. This global representation directly allows computing distances between points regardless to which local cluster they belong. Moreover, it also enables data visualization in a single view, and the use of ℓ_2-based efficient retrieval methods. Experiments on the Labeled Faces in the Wild dataset show that our approach improves over previous global and local metric learning approaches.

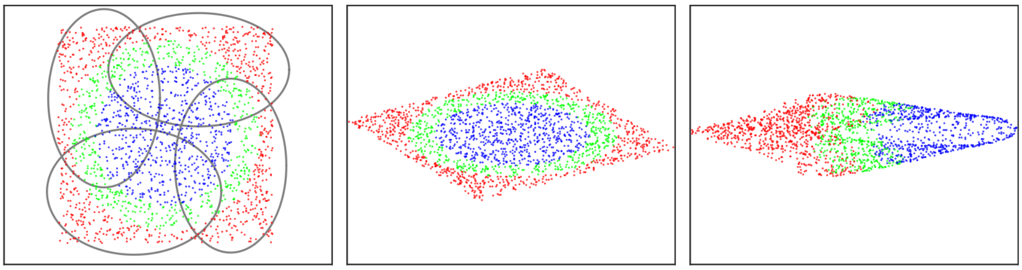


Figure 6. Synthetic dataset with color coded class labels, and the GMM used by our CLML local metric (left). Data projection given by a global Mahalanobis metric (middle) and our local CLML metric (right). The pairwise training constraints are better respected by CLML.

7.2.4. A convex formulation for joint RNA isoform detection and quantification from multiple RNA-seq samples

Participants: Elsa Bernard [Institut Curie, Ecoles des Mines-ParisTech], Laurent Jacob [CNRS, LBBE Laboratory], Julien Mairal, Jean-Philippe Vert [Institut Curie, Ecoles des Mines-ParisTech].

Detecting and quantifying isoforms from RNA-seq data is an important but challenging task. The problem is often ill-posed, particularly at low coverage. One promising direction is to exploit several samples simultaneously. In this paper [4], we propose a new method for solving the isoform deconvolution problem jointly across several samples. We formulate a convex optimization problem that allows to share information between samples and that we solve efficiently, as illustrated in Figure 7 . We demonstrate the benefits of combining several samples on simulated and real data, and show that our approach outperforms pooling strategies and methods based on integer programming. Our convex formulation to jointly detect and quantify isoforms from RNA-seq data of multiple related samples is a computationally efficient approach to leverage the hypotheses that some isoforms are likely to be present in several samples. The software and source code are available at http://cbio.ensmp.fr/flipflop.

7.2.5. Adaptive Recovery of Signals by Convex Optimization

Participants: Zaid Harchaoui, Anatoli Juditsky [Univ. Grenoble], Arkadi Nemirovski [Georgia Tech], Dimitry Ostrovsky [Univ. Grenoble].

In [13], we present a theoretical framework for adaptive estimation and prediction of signals of unknown structure in the presence of noise. The framework allows to address two intertwined challenges: (i) designing optimal statistical estimators; (ii) designing efficient numerical algorithms. In particular, we establish oracle inequalities for the performance of adaptive procedures, which rely upon convex optimization and thus can be efficiently implemented. As an application of the proposed approach, we consider denoising of harmonic oscillations

7.2.6. Semi-proximal Mirror-Prox for Nonsmooth Composite Minimization

Participants: Niao He [Georgia Tech], Zaid Harchaoui.

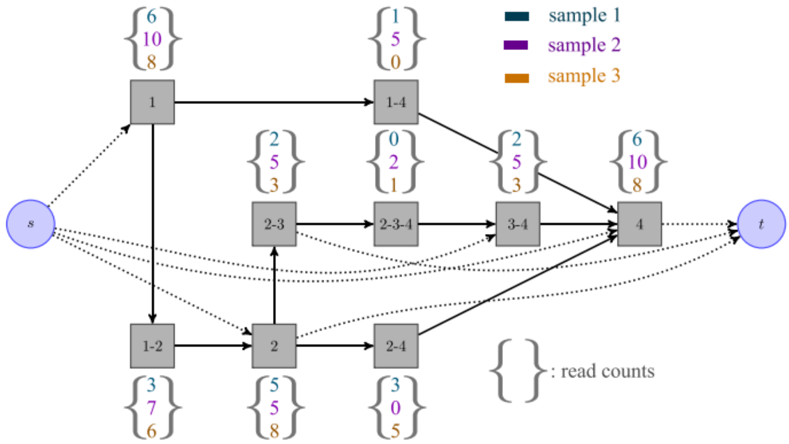


Fig. 1 Multi-dimensional splicing graph with three samples. Each candidate isoform is a path from source node *s* to sink node *t*. Nodes denoted as grey squares correspond to ordered set of exons. Each read is assigned to a unique node, corresponding to the exact set of exons that it overlaps. Note that more than 2 exons can constitute a node, properly modeling reads spanning more than 2 exons. A vector of read counts (one component per sample) is then associated to each node of the graph. Note also that some components of a vector can be equal to zero

Figure 7. Graph on which we perform network flow optimization. Nodes represent observed reads, and paths on the graph correspond to isoforms.

In [28], we propose a new first-order optimisation algorithm to solve high-dimensional non-smooth composite minimisation problems. Typical examples of such problems have an objective that decomposes into a non-smooth empirical risk part and a non-smooth regularisation penalty. The proposed algorithm, called Semi-Proximal Mirror-Prox, leverages the Fenchel-type representation of one part of the objective while handling the other part of the objective via linear minimization over the domain. The algorithm stands in contrast with more classical proximal gradient algorithms with smoothing, which require the computation of proximal operators at each iteration and can therefore be impractical for high-dimensional problems. We establish the theoretical convergence rate of Semi-Proximal Mirror-Prox, which exhibits the optimal complexity bounds, for the number of calls to linear minimization oracle. We present promising experimental results showing the interest of the approach in comparison to competing methods.

7.3. Recognition in video

7.3.1. Beat-Event Detection in Action Movie Franchises

Participants: Danila Potapov, Matthijs Douze, Jerome Revaud, Zaid Harchaoui, Cordelia Schmid.

While important advances were recently made towards temporally localizing and recognizing specific human actions or activities in videos, efficient detection and classification of long video chunks belonging to semantically-defined categories such as “pursuit” or “romance” remains challenging.

In our work [30], we introduce a new dataset, Action Movie Franchises, consisting of a collection of Hollywood action movie franchises. We define 11 non-exclusive semantic categories — called beat-categories — that are broad enough to cover most of the movie footage. The corresponding beat-events are annotated as groups of video shots, possibly overlapping. We propose an approach for localizing beat-events based on classifying shots into beat-categories and learning the temporal constraints between shots, as shown in Figure 8 . We show that temporal constraints significantly improve the classification performance. We set up an evaluation protocol for beat-event localization as well as for shot classification, depending on whether movies from the same franchise are present or not in the training data.

Figure 8. A 5-minute extract from the proposed Action Movie Franchises dataset, ground truth annotation and output of different methods. Each color stands for a different event category: green —pursuit, blue —battle, yellow —victory-good, green —despair-good, pink —romance, gray —victory-bad, cadet blue —good-argue-good. Hashes mark difficult examples. The color code for the classifier evaluation is: white = true positive, gray = ignored, black = false positive.

7.3.2. EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow

Participants: Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid.

In this paper [18], we propose a novel approach for optical flow estimation, targeted at large displacements with significant occlusions. It consists of two steps: i) dense matching by edge-preserving interpolation from a sparse set of matches; ii) variational energy minimization initialized with the dense matches. The sparse-to-dense interpolation relies on an appropriate choice of the distance, namely an edge-aware geodesic distance. This distance is tailored to handle occlusions and motion boundaries – two common and difficult issues for optical flow computation. We also propose an approximation scheme for the geodesic distance to allow fast computation without loss of performance. Subsequent to the dense interpolation step, standard one-level variational energy minimization is carried out on the dense matches to obtain the final flow estimation. The proposed approach, called Edge-Preserving Interpolation of Correspondences *(EpicFlow)* is fast and robust to large displacements. An overview is given in Figure 9 . EpicFlow significantly outperforms the state of the art on MPI-Sintel and performs on par on Kitti and Middlebury.

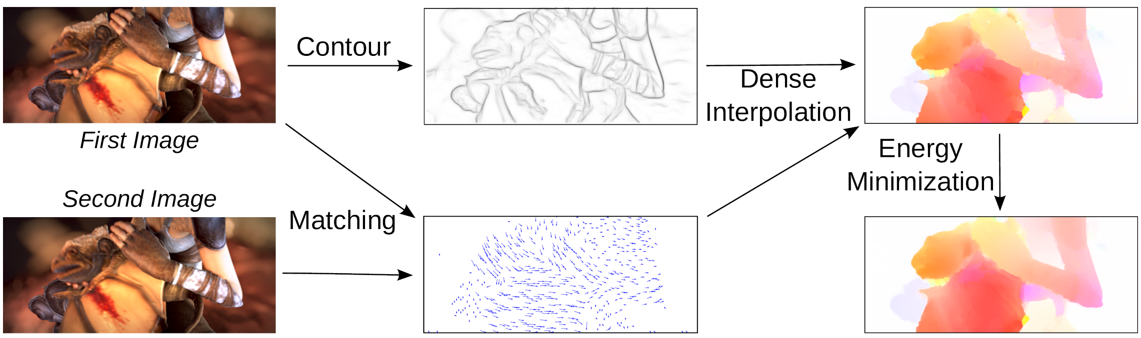


Figure 9. Overview of EpicFlow. Given two images, we compute matches using DeepMatching and the edges of the first image using SED. We combine these two cues to densely interpolate matches and obtain a dense correspondence field. This is used as initialization of a one-level energy minimization framework.

7.3.3. DeepMatching: Hierarchical Deformable Dense Matching

Participants: Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid.

In this paper [31], we introduce a novel matching algorithm, called DeepMatching, to compute dense correspondences between images. DeepMatching relies on a hierarchical, multi-layer, correlational architecture designed for matching images and was inspired by deep convolutional approaches, see Figure 10 . The proposed matching algorithm can handle non-rigid deformations and repetitive textures and efficiently determines dense correspondences in the presence of significant changes between images. We evaluate the performance of DeepMatching, in comparison with state-of-the-art matching algorithms, on the Mikolajczyk, the MPI-Sintel and the Kitti datasets. DeepMatching outperforms the state-of-the-art algorithms and shows excellent results in particular for repetitive textures. We also propose a method for estimating optical flow, called DeepFlow, by integrating DeepMatching in the large displacement optical flow (LDOF) approach of Brox et al. Compared to existing matching algorithms, additional robustness to large displacements and complex motion is obtained thanks to our matching approach. DeepFlow obtains competitive performance on public benchmarks for optical flow estimation.

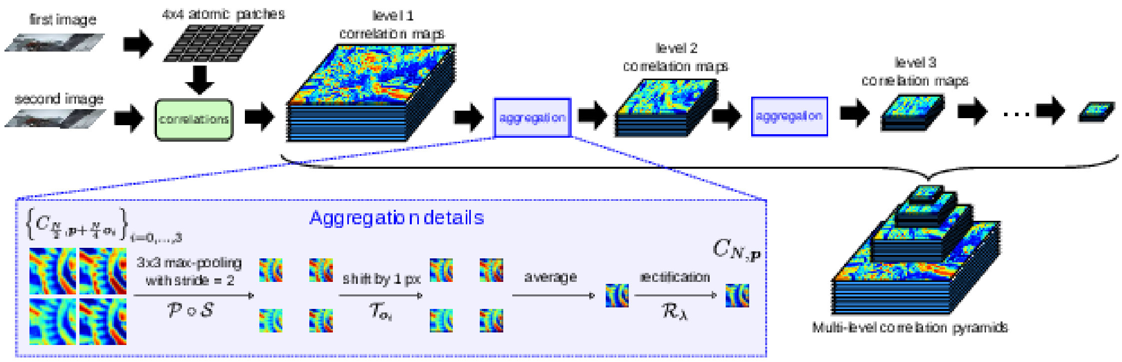


Figure 10. Overview of the bottom-up part of DeepMatching, which builds the multi-level correlation pyramid, from which matches are then extracted.

7.3.4. Learning to Detect Motion Boundaries

Participants: Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, Cordelia Schmid.

In this paper [23], we propose a learning-based approach for motion boundary detection. Precise localization of motion boundaries is essential for the success of optical flow estimation, as motion boundaries correspond to discontinuities of the optical flow field. The proposed approach allows to predict motion boundaries, using a structured random forest trained on the ground-truth of the MPI-Sintel dataset, see Figure 11 . The random forest leverages several cues at the patch level, namely appearance (RGB color) and motion cues (optical flow estimated by state-of-the-art algorithms). Experimental results show that the proposed approach is both robust and computationally efficient. It significantly outperforms state-of-the-art motion-difference approaches on the MPI-Sintel and Middlebury datasets. We compare the results obtained with several state-of-the-art optical flow approaches and study the impact of the different cues used in the random forest. Furthermore, we introduce a new dataset, the YouTube Motion Boundaries dataset (**YMB**), that comprises 60 sequences taken from real-world videos with manually annotated motion boundaries. On this dataset, our approach, although trained on MPI-Sintel, also outperforms by a large margin state-of-the-art optical flow algorithms.

7.3.5. Learning to track for spatio-temporal action localization

Participants: Philippe Weinzaepfel, Zaid Harchaoui, Cordelia Schmid.

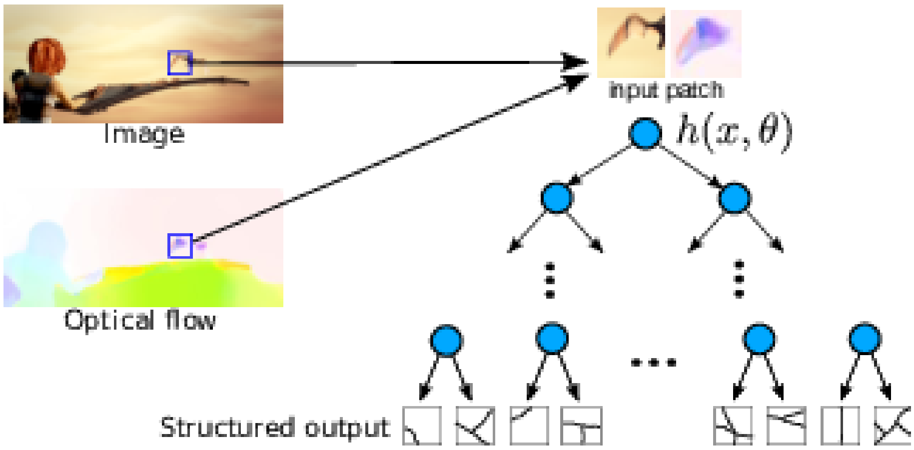


Figure 11. A structured random forest, taking as input a patch of an image and corresponding optical flow, outputs a motion boundaries patch. These motion boundaries patches are then aggregated to build the motion boundaries map for the whole image.

In this paper [22], we propose an effective approach for spatio-temporal action localization in realistic videos. The approach first detects proposals at the frame-level and scores them with a combination of static and motion CNN features. It then tracks high-scoring proposals throughout the video using a tracking-by-detection approach. Our tracker relies simultaneously on instance-level and class-level detectors. The tracks are scored using a spatio-temporal motion histogram, a descriptor at the track level, in combination with the CNN features. Finally, we perform temporal localization of the action using a sliding-window approach at the track level. An overview of our approach is given in Figure 12 . We present experimental results for spatio-temporal localization on the UCF-Sports, J-HMDB and UCF-101 action localization datasets, where our approach outperforms the state of the art with a margin of 15%, 7% and 12% respectively in mAP.

7.3.6. A robust and efficient video representation for action recognition

Participants: Heng Wang, Dan Oneata, Cordelia Schmid, Jakob Verbeek.

In [9] we present a state-of-the-art video representation and apply it to efficient action recognition and detection. We first propose to improve the popular dense trajectory features by explicit camera motion estimation. Local feature trajectories consistent with the homography are considered as due to camera motion, and thus removed. This results in significant improvement on motion-based HOF and MBH descriptors. We further explore the recent Fisher vector as an alternative feature encoding approach to the standard bag-of-words histogram, and consider different ways to include spatial layout information in these encodings. We present a large and varied set of evaluations, considering (i) classification of short basic actions on six datasets, (ii) localization of such actions in featurelength movies, and (iii) large-scale recognition of complex events. We find that our improved trajectory features significantly outperform previous dense trajectories, and that Fisher vectors are superior to bag-of-words encodings for video recognition tasks. In all three tasks, we show substantial improvements over the state-of-the-art results. This journal paper combines and extends earlier conference papers.

7.3.7. Circulant temporal encoding for video retrieval and temporal alignment

Participants: Jerome Revaud, Matthijs Douze, Hervé Jégou [Inria Rennes, Facebook AI Research], Cordelia Schmid, Jakob Verbeek.

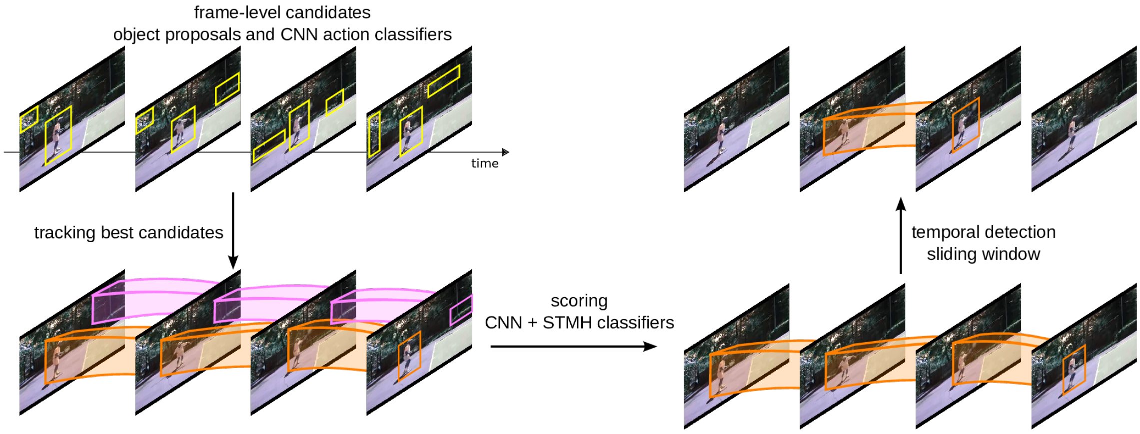


Figure 12. Overview of our spatio-temporal action localization approach. We detect frame-level object proposals and score them with CNN action classifiers. The best candidates, in term of scores, are tracked throughout the video. We then score the tracks with CNN and spatio-temporal motion histogram (STMH) classifiers. Finally, we perform a temporal sliding window for detecting the temporal extent of the action.

In [6] we address the problem of specific video event retrieval. Given a query video of a specific event, e.g., a concert of Madonna, the goal is to retrieve other videos of the same event that temporally overlap with the query. Our approach encodes the frame descriptors of a video to jointly represent their appearance and temporal order. It exploits the properties of circulant matrices to efficiently compare the videos in the frequency domain. This offers a significant gain in complexity and accurately localizes the matching parts of videos. The descriptors can be compressed in the frequency domain with a product quantizer adapted to complex numbers. In this case, video retrieval is performed without decompressing the descriptors. The second problem we consider is the temporal alignment of a set of videos. We exploit the matching confidence and an estimate of the temporal offset computed for all pairs of videos by our retrieval approach. Our robust algorithm aligns the videos on a global timeline by maximizing the set of temporally consistent matches. The global temporal alignment enables synchronous playback of the videos of a given scene. This journal paper extends an earlier conference paper.

7.3.8. Pose Estimation and Segmentation of Multiple People in Stereoscopic Movies

Participants: Guillaume Seguin [Willow], Karteek Alahari, Josef Sivic [Willow], Ivan Laptev [Willow].

The work in [8] presents a method to obtain a pixel-wise segmentation and pose estimation of multiple people in stereoscopic videos, as shown in Figure 13 . This task involves challenges such as dealing with unconstrained stereoscopic video, non-stationary cameras, and complex indoor and outdoor dynamic scenes with multiple people. We cast the problem as a discrete labelling task involving multiple person labels, devise a suitable cost function, and optimize it efficiently. The contributions of our work are two-fold: First, we develop a segmentation model incorporating person detections and learnt articulated pose segmentation masks, as well as colour, motion, and stereo disparity cues. The model also explicitly represents depth ordering and occlusion. Second, we introduce a stereoscopic dataset with frames extracted from feature-length movies “StreetDance 3D" and “Pina". The dataset contains 587 annotated human poses, 1158 bounding box annotations and 686 pixel-wise segmentations of people. The dataset is composed of indoor and outdoor scenes depicting multiple people with frequent occlusions. We demonstrate results on our new challenging dataset, as well as on the H2view dataset from (Sheasby et al. ACCV 2012).

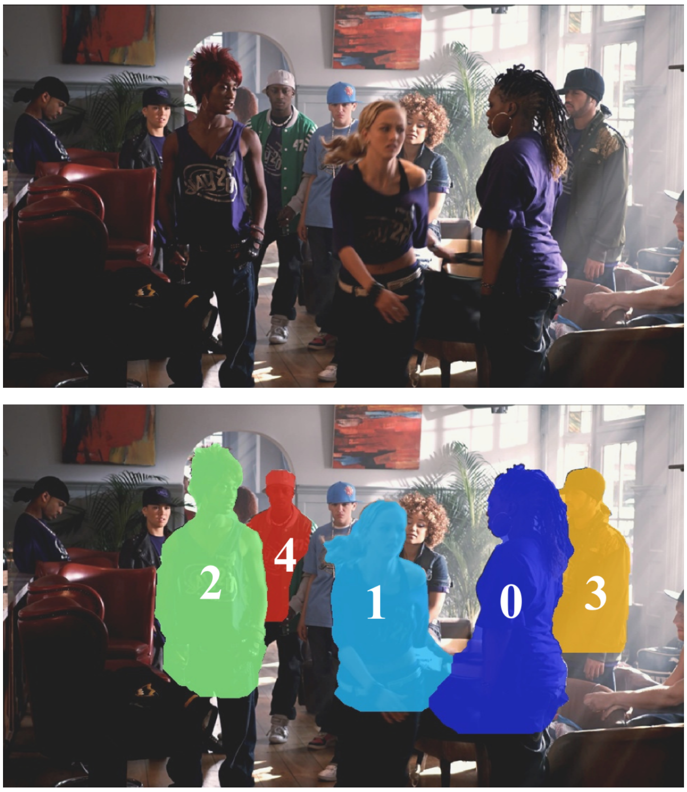


Figure 13. We segment multiple people in the scene, estimate their poses and relative front-to-back order, denoted by the numbers in the image below, in every frame of a video sequence.

7.3.9. Encoding Feature Maps of CNNs for Action Recognition

Participants: Xiaojiang Peng, Cordelia Schmid.

In [29] We describe our approach for action classification in the THUMOS Challenge 2015. Our approach is based on two types of features, improved dense trajectories and CNN features, as illustrated in Figure 14 . For trajectory features, we extract HOG, HOF, MBHx, and MBHy descriptors and apply Fisher vector encoding. For CNN features, we utilize a recent deep CNN model, VGG19, to capture appearance features and use VLAD encoding to encode/pool convolutional feature maps which shows better performance than average pooling of feature maps and full-connected activation features.

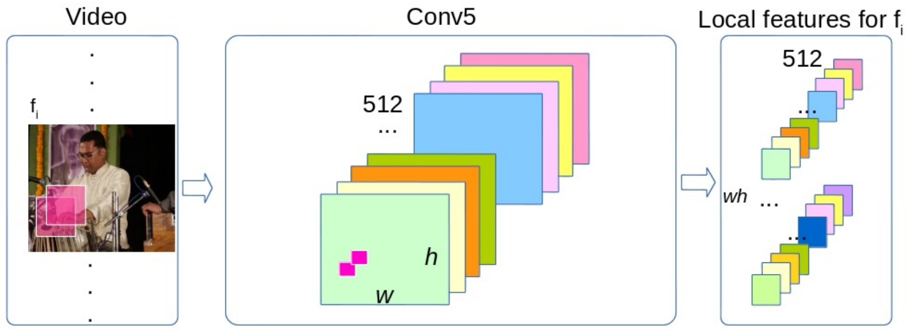


*Figure 14. Local features from convolutional feature maps. Each pixel (pink square in the middle image) in the Conv5 feature map is actually a feature for the corresponding patch in original frame. We obtain w*h 512-D features for frame f_i.*

7.3.10. Online Object Tracking with Proposal Selection

Participants: Yang Hua, Karteek Alahari, Cordelia Schmid.

Tracking-by-detection approaches are some of the most successful object trackers in recent years. Their success is largely determined by the detector model they learn initially and then update over time. However, under challenging conditions where an object can undergo transformations, e.g., severe rotation, these methods are found to be lacking. In [14], we address this problem by formulating it as a proposal selection task and making two contributions. The first one is introducing novel proposals estimated from the geometric transformations undergone by the object, and building a rich candidate set for predicting the object location. The second one is devising a novel selection strategy using multiple cues, i.e., detection score and edgeness score computed from state-of-the-art object edges and motion boundaries. We extensively evaluate our approach on the visual object tracking 2014 challenge and online tracking benchmark datasets, and show the best performance. Sample results are shown in Figure 15 . Our tracker based on this method has recently won the visual object tracking challenge (VOT-TIR) organized as part of ICCV 2015 in Santiago, Chile.

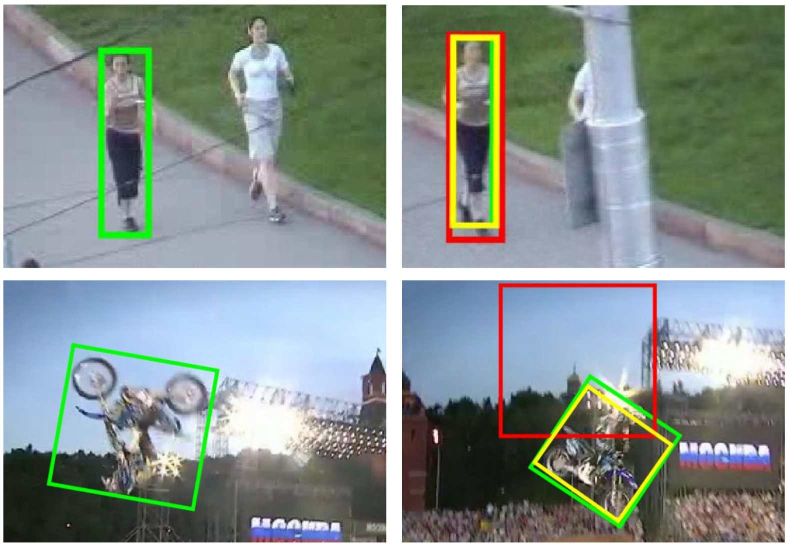


Figure 15. Sample frames (cropped) from the jogging (top row) and motocross (bottom row) sequences. The ground truth annotation (green) in the first frame (left) is used to train our tracker and the winner of VOT2014 challenge. We show these two tracking results (right) on another frame in the sequence. Our method (yellow) successfully tracks objects undergoing deformations unlike winner of VOT2014 challenge (red).

LINKMEDIA Project-Team

6. New Results

6.1. Unsupervised motif and knowledge discovery

6.1.1. Estimation of continuous intrinsic dimension

Participants: Laurent Amsaleg, Teddy Furon.

In collaboration with Michael Houle, National Institute for Informatics (Japan).

Some of our research work was concerned with the estimation of continuous intrinsic dimension (ID), a measure of intrinsic dimensionality recently proposed by Houle. Continuous ID can be regarded as an extension of Karger and Ruhl's expansion dimension to a statistical setting in which the distribution of distances to a query point is modeled in terms of a continuous random variable. This form of intrinsic dimensionality can be particularly useful in search, classification, outlier detection, and other contexts in machine learning, databases, and data mining, as it has been shown to be equivalent to a measure of the discriminative power of similarity functions. In [11], we proposed several estimators of continuous ID that we analyzed based on extreme value theory, using maximum likelihood estimation, the method of moments, probability weighted moments, and regularly varying functions. Experimental evaluation was performed using both real and artificial data.

6.1.2. Supervised multi-scale locality sensitive hashing

Participants: Laurent Amsaleg, Li Weng.

LSH is a popular framework to generate compact representations of multimedia data, which can be used for content based search. However, the performance of LSH is limited by its unsupervised nature and the underlying feature scale. In [42], we proposed to improve LSH by incorporating two elements: supervised hash bit selection and multi-scale feature representation. First, a feature vector is represented by multiple scales. At each scale, the feature vector is divided into segments. The size of a segment is decreased gradually to make the representation correspond to a coarse-to-fine view of the feature. Then each segment is hashed to generate more bits than the target hash length. Finally the best ones are selected from the hash bit pool according to the notion of bit reliability, which is estimated by bit-level hypothesis testing. Extensive experiments have been performed to validate the proposal in two applications: near-duplicate image detection and approximate feature distance estimation. We first demonstrate that the feature scale can influence performance, which is often a neglected factor. Then we show that the proposed supervision method is effective. In particular, the performance increases with the size of the hash bit pool. Finally, the two elements are put together. The integrated scheme exhibits further improved performance.

6.1.3. Rotation and translation covariant match kernels for image retrieval

Participants: Andrei Bursuc, Teddy Furon, Hervé Jégou, Giorgos Tolias.

Most image encodings achieve orientation invariance by aligning the patches to their dominant orientations and translation invariance by completely ignoring patch position or by max-pooling. Albeit successful, such choices introduce too much invariance because they do not guarantee that the patches are rotated or translated consistently. In this work, we propose a geometric-aware aggregation strategy, which jointly encodes the local descriptors together with their patch dominant angle [38] and/or location [10]. The geometric attributes are encoded in a continuous manner by leveraging explicit feature maps. Our technique is compatible with generic match kernel formulation and can be employed along with several popular encoding methods, in particular bag of words, VLAD and the Fisher vector. The method is further combined with an efficient monomial embedding to provide a codebook-free method aggregating local descriptors into a single vector representation. Invariance is achieved by efficient similarity estimation of multiple rotations or translations, offered by a simple trigonometric polynomial. This strategy is effective for image search, as shown by experiments performed on standard benchmarks for image and particular object retrieval, namely Holidays and Oxford buildings.

6.1.4. Sequential pattern mining on audio data

Participants: Laurent Amsaleg, Guillaume Gravier, Simon Malinowski.

M. Sc. Internship of Corentin Hardy, in collaboration with René Quiniou, Inria Rennes, DREAM research team, within the framework of the STIC AmSud Maximum project and of the MOTIF Inria Associate Team.

Analyzing multimedia data is a challenging problem due to the quantity and complexity of such data. Mining for frequently recurring patterns is a task often ran to help discovering the underlying structure hidden in the data. This year, we have explored how data symbolization and sequential pattern mining techniques could help for mining recurring patterns in multimedia data. In [20], we have shown that even if sequential pattern mining techniques are very helpful in terms of computational efficiency, the data symbolization step is a crucial step to find for extracting relevant audio patterns.

6.1.5. Clustering by diverting supervised machine learning

Participants: Vincent Claveau, Teddy Furon, Guillaume Gravier.

M. Sc. Internship of Amélie Royer, ENS Rennes.

Clustering algorithms exploit an input similarity measure on the samples, which should be fine-tuned with the data format and the application at hand. However, manually defining a suitable similarity measure is a difficult task in case of limited prior knowledge or complex data structures for example. While supervised classification systems require a set of samples annotated with their ground-truth classes, recent studies have shown it is possible to exploit classifiers trained on an artificial annotation of the data in order to induce a similarity measure. In this work, we have proposed a unified framework, named similarity by iterative classifications (SIC), which explores the idea of diverting supervised learning for automatic similarity inference. We studied several of its theoretical and practical aspects. We also have implemented and evaluate SIC on three tasks of knowledge discovery on multimedia content. Results show that in most situations the proposed approach indeed benefits from the underlying classifier's properties and outperforms usual similarity measures for clustering applications.

6.1.6. Multimodal person discovery in TV broadcasts

Participant: Guillaume Gravier.

Work in collaboration with Cassio Elias dos Santos Jr. and William Robson Schwartz, in the framework of the Inria Associate Team MOTIF and of the STIC AmSud project Maximum.

Taking advantage of recent results on large-scale face comparison with partial least square, we developed various approaches for multimodal person discovery in TV broadcasts in the framework of the MediaEval 2015 international benchmark [30]. The task consists in naming the persons on screen that are speaking with no prior information, leveraging text overlays, speech transcripts as well as face and voice comparison. We investigated two distinct aspects of multimodal person discovery. One refers to face clusters, which are considered to propagate names associated with faces in one shot to other faces that probably belong to the same person. The face clustering approach consists in calculating face similarities using partial least squares and a simple hierarchical approach. The other aspect refers to tag propagation in a graph-based approach where nodes are speaking faces and edges link similar faces/speakers. The advantage of the graph-based tag propagation is to not rely on face/speaker clustering, which we believe can be errorprone. The face clustering approach ranked among the top results in the international benchmark.

6.1.7. Unsupervised video structure mining with grammatical inference

Participants: Guillaume Gravier, Bingqing Qu.

In collaboration with Jean Carrive and Félicien Vallet, Institut National de l'Audiovisuel.

In [25], we addressed the problem of unsupervised program structuring with minimal prior knowledge about the program. We extended previous work to propose an approach able to identify multiple structures and infer structural grammars for recurrent TV programs of different types. The approach taken involves three sub-problems: i) we determine the structural elements contained in programs with minimal knowledge about which type of elements may be present; ii) we identify multiple structure for the programs if any and model the structures of programs; iii) we generate the structural grammar for each corresponding structure. Finally, we conducted use-case based evaluations on real recurrent programs of three different types to demonstrate the effectiveness of the proposed approach.

6.1.8. Information retrieval for distributional semantics, and vice-versa

Participants: Vincent Claveau, Ewa Kijak.

Distributional thesauri are useful in many tasks of natural language processing. In [33], [3], we address the problem of building and evaluating such thesauri with the help of information retrieval (IR) concepts. Two main contributions are proposed. First, in the continuation of previous work, we have shown how IR tools and concepts can be used with success to build thesauri. Through several experiments and by evaluating directly the results with reference lexicons, we show that some IR models outperform state-of-the-art systems. Secondly, we use IR as an application framework to indirectly evaluate the generated thesaurus. Here again, this task-based evaluation validate the IR approach used to build the thesaurus. Moreover, it allows us to compare these results with those from the direct evaluation framework used in the literature. The observed differences question these evaluation habits.

6.2. Multimedia content description and structuring

6.2.1. Image description using component trees

Participants: Petra Bosilj, Ewa Kijak.

In collaboration with Sébastien Lefèvre from Obelix Team (IRISA).

In this work, we explored the application of a tree-based feature extraction algorithm for the widely-used MSER features, and proposed a tree-of-shapes based detector of maximally stable regions. Changing an underlying component tree in the algorithm allows considering alternative properties and pixel orderings for extracting maximally stable regions. Performance evaluation was carried out on a standard benchmark in terms of repeatability and matching score under different image transformations, as well as in a large scale image retrieval setup, measuring mean average precision. The detector outperformed the baseline MSER in the retrieval experiments [37].

We also proposed a local region descriptor based on 2D shape-size pattern spectra, calculated on arbitrary connected regions, and combined with normalized central moments. The challenges when transitioning from global pattern spectra to the local ones were faced, and an exhaustive study on the parameters and the properties of the newly constructed descriptor was conducted. The descriptors were calculated on MSER regions, and evaluated in a simple retrieval system. Competitive performance with SIFT descriptors was achieved. An additional advantage of the proposed descriptors is their size which is less than half the size of SIFT [14], [15].

6.2.2. Improved motion description for action classification

Participant: Hervé Jégou.

In collaboration with Mihir Jain (University of Amsterdam, The Netherlands) and Patrick Bouthemy (Team-project SERPICO, Inria Rennes, France)

Even though the importance of explicitly integrating motion characteristics in video descriptions has been demonstrated by several recent papers on action classification, our current work concludes that adequately decomposing visual motion into dominant and residual motions, i.e., camera and scene motion, significantly improves action recognition algorithms. This holds true both for the extraction of the space-time trajectories and for computation of descriptors. We designed in [7] a new motion descriptor—the DCS descriptor—that captures additional information on local motion patterns enhancing results based on differential motion scalar quantities, divergence, curl and shear features. Finally, applying the recent VLAD coding technique proposed in image retrieval provides a substantial improvement for action recognition. These findings are complementary to each other and they outperformed all previously reported results by a significant margin on three challenging datasets: Hollywood 2, HMDB51 and Olympic Sports as reported in (Jain et al. (2013)).

6.2.3. Word embeddings and recurrent neural networks for spoken language understanding

Participants: Guillaume Gravier, Christian Raymond, Vedran Vukotić.

Recently, word embedding representations have been investigated for slot filling in spoken language understanding (SLU), along with the use of neural networks as classifiers. Neural networks, especially recurrent neural networks, which are adapted to sequence labeling problems, have been applied successfully on the popular ATIS database. In [29], we make a comparison of this kind of models with the previously state-of-the-art conditional random fields (CRF) classifier on a more challenging SLU database. We show that, despite efficient word representations used within these neural networks, their ability to process sequences is still significantly lower than for CRF, while also having a drawback of higher computational costs, and that the ability of CRF to model output label dependencies is crucial for SLU.

6.2.4. Hierarchical topic structuring

Participants: Guillaume Gravier, Pascale Sébillot, Anca-Roxana Şimon.

Topic segmentation traditionally relies on lexical cohesion measured through word re-occurrences to output a dense segmentation, either linear or hierarchical. We have proposed a novel organization of the topical structure of textual content [28]. Rather than searching for topic shifts to yield dense segmentation, our algorithm extracts topically focused fragments organized in a hierarchical manner. This is achieved by leveraging the temporal distribution of word re-occurrences, searching for bursts, to skirt the limits imposed by a global counting of lexical re-occurrences within segments. Comparison to a reference dense segmentation on varied datasets indicates that we can achieve a better topic focus while retrieving all of the important aspects of a text.

6.2.5. Partial least square hashing for large-scale face identification

Participants: Guillaume Gravier, Ewa Kijak.

Work performed with Cassio Elias dos Santos Jr. during his 3 months visit, in collaboration with William Robson Schwartz (UFMG, Brasil), in the framework of the Inria Associate Team MOTIF.

Face recognition has been largely studied in past years. However, most of the related work focus on increasing accuracy and/or speed to test a single pair probe-subject. In [31], we introduced a novel method inspired by the success of locality sensing hashing applied to large general purpose datasets and by the robustness provided by partial least squares analysis when applied to large sets of feature vectors for face recognition. The result is a robust hashing method compatible with feature combination for fast computation of a short list of candidates in a large gallery of subjects. We provided theoretical support and practical principles for the proposed hashing method that may be reused in further development of hash functions applied to face galleries. Comparative evaluations on the FERET and FRGCv1 datasets demonstrate a speedup of a factor 16 compared to scanning all subjects in the face gallery.

6.2.6. Selection strategies for active learning in NLP

Participants: Vincent Claveau, Ewa Kijak.

Nowadays, many NLP problems are modelized as supervised machine learning tasks, especially when it comes to information extraction. Consequently, the cost of the expertise needed to annotate the examples is a widespread issue. Active learning offers a framework to that issue, allowing to control the annotation cost while maximizing the classifier performance, but it relies on the key step of choosing which example will be proposed to the expert. In [3], we have examined and proposed such selection strategies in the specific case of conditional random fields which are largely used in NLP. On the one hand, we have proposed a simple method to correct a bias of certain state-of-the-art selection techniques. On the other hand, we have detailed an original approach to select the examples, based on the respect of proportions in the datasets. These contributions are validated over a large range of experiments implying several tasks and datasets, including named entity recognition, chunking, phonetization, word sense disambiguation.

6.2.7. Tree-structured named entities extraction from competing speech transcripts

Participant: Christian Raymond.

When real applications are working with automatic speech transcription, the first source of error does not originate from the incoherence in the analysis of the application but from the noise in the automatic transcriptions. In [41], we present a simple but effective method to generate a new transcription of better quality by combining utterances from competing transcriptions. We have extended a structured named entity (NE) recognizer submitted during the ETAPE challenge. Working on French TV and radio programs, our system revises the transcriptions provided by making use of the NEs it has detected. Our results suggest that combining the transcribed utterances which optimize the F-measure, rather than minimizing the WER scores, allows the generation of a better transcription for NE extraction. The results show a small but significant improvement of 0.9 % SER against the baseline system on the ROVER transcription. These are the best performances reported to date on this corpus.

6.3. Content-based information retrieval

6.3.1. A comparison of dense region detectors for image search and fine-grained classification

Participants: Hervé Jégou, Ahmet Iscen, Giorgos Tolias.

In collaboration with Philippe-Henri Gosselin (ETIS team, ENSEA, Cergy, France)

We consider a pipeline for image classification or search based on coding approaches like bag of words or Fisher vectors. In this context, the most common approach is to extract the image patches regularly in a dense manner on several scales. In [6], we propose and evaluate alternative choices to extract patches densely. Beyond simple strategies derived from regular interest region detectors, we propose approaches based on super-pixels, edges, and a bank of Zernike filters used as detectors. The different approaches are evaluated on recent image retrieval and fine-grain classification benchmarks. Our results show that the regular dense detector is outperformed by other methods in most situations, leading us to improve the state of the art in comparable setups on standard retrieval and fined-grain benchmarks. As a byproduct of our study, we show that existing methods for blob and super-pixel extraction achieve high accuracy if the patches are extracted along the edges and not around the detected regions.

6.3.2. Efficient large-scale similarity search using matrix factorization

Participants: Teddy Furon, Ahmet Iscen.

In collaboration with Michael Rabbat (McGill University, Montréal, Canada)

We considered 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. 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 scenario. We evaluate our approach both for global descriptors obtained from SIFT and CNN features. Experiments with standard image search benchmarks, including the Yahoo100M dataset comprising 100 million images, show that our method gives comparable (and sometimes superior) 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 [43].

6.3.3. Explicit embeddings for nearest neighbor search with Mercer kernels

Participant: Hervé Jégou.

In collaboration with Anthony Bourrier and Patrick Pérez (Technicolor, Rennes, France), Florent Perronnin (Xerox, Grenoble, France) Rémi Gribonval (Team-project PANAMA, Inria Rennes, France).

Many approximate nearest neighbor search algorithms operate under memory constraints, by computing short signatures for database vectors while roughly keeping the neighborhoods for the distance of interest. Encoding procedures designed for the Euclidean distance have attracted much attention in the last decade. In the case where the distance of interest is based on a Mercer kernel, we propose a simple, yet effective two-step encoding scheme: first, compute an explicit embedding to map the initial space into a Euclidean space; second, apply an encoding step designed to work with the Euclidean distance. Comparing this simple baseline with existing methods relying on implicit encoding, we demonstrate better search recall for similar code sizes with the chi-square kernel in databases comprised of visual descriptors, outperforming concurrent state-of-the-art techniques by a large margin [2].

6.3.4. Image search with selective match kernels: aggregation across single and multiple images

Participants: Hervé Jégou, Giorgos Tolias.

In collaboration with Yannis Avrithis (National Technical University of Athens, Greece)

Our work [9] considers a family of metrics to compare images based on their local descriptors. It encompasses the VLAD descriptor and matching techniques such as Hamming Embedding. Making the bridge between these approaches leads us to propose a match kernel that takes the best of existing techniques by combining an aggregation procedure with a selective match kernel. The representation underpinning this kernel is approximated, providing a large scale image search both precise and scalable, as shown by our experiments on several benchmarks. We show that the same aggregation procedure, originally applied per image, can effectively operate on groups of similar features found across multiple images. This method implicitly performs feature set augmentation, while enjoying savings in memory requirements at the same time. Finally, the proposed method is shown effective for place recognition, outperforming state of the art methods on a large scale landmark recognition benchmark.

6.3.5. Early burst detection for memory-efficient image retrieval

Participant: Hervé Jégou.

In collaboration with Miajing Shi, visiting Ph. D. student from Pekin University, and Yannis Avrithis (National Technical University of Athens, Greece)

Recent works show that image comparison based on local descriptors is corrupted by visual bursts, which tend to dominate the image similarity. The existing strategies, like power-law normalization, improve the results by discounting the contribution of visual bursts to the image similarity. We proposed to explicitly detect the visual bursts in an image at an early stage. We compare several detection strategies jointly taking into account feature similarity and geometrical quantities. The bursty groups are merged into meta-features, which are used as input to state-of-the-art image search systems such as VLAD or the selective match kernel. Then, we show the interest of using this strategy in an asymmetrical manner, with only the database features being aggregated but not those of the query. Extensive experiments performed on public benchmarks for visual retrieval show the benefits of our method, which achieves performance on par with the state of the art but with a significantly reduced complexity, thanks to the lower number of features fed to the indexing system [40], [44].

6.3.6. Biomedical information retrieval

Participants: Vincent Claveau, Ewa Kijak.

In collaboration with N. Grabar (STL), T. Hamon (LIMSI), and S. Le Maguer (Univ. Saarland).

The right of patients to access their clinical health record is granted by the code of Santé Publique. Yet, this piece of content remains difficult to understand. We propose different IR experiments in which we use queries defined by patients in order to find relevant documents [3], [16]. We use the Indri search engine, based on statistical language modeling, as well as semantic resources. More precisely, our approaches are chiefly based on the terminological variation (e.g., synonyms, abbreviations) to link between expert and patient languages. Various combinations of resources and Indri settings are explored, mostly based on query expansion.

6.4. Linking, navigation and analytics

6.4.1. Sentiment analysis on social networks

Participants: Vincent Claveau, Christian Raymond, Vedran Vukotić.

In the framework of our participation to the DeFT 2015 text-mining challenge, we have developped sentiment-analysis methods for tweets [34]. Several sub-tasks have been considered: i) valence classification of tweets and ii) fine-grained classification of tweets (which includes two sub-tasks: detection of the generic class of the information expressed in a tweet and detection of the specific class of the opinion/sentiment/emotion. For all three problems, we adopt a standard machine learning framework. More precisely, three main methods are proposed and their feasibility for the tasks is analyzed: i) decision trees with boosting (bonzaiboost), ii) naive Bayes with Okapi and iii) convolutional neural networks (CNNs). Our approaches are voluntarily knowledge free and text-based only, we do not exploit external resources (lexicons, corpora) or tweet metadata. It allows us to evaluate the interest of each method and of traditional bag-of-words representations vs. word embeddings. Methods using simple ML frameworks and IR-based similarity metrics have been demonstrated to yield the best results.

6.4.2. A multi-dimensional data model for personal photo browsing

Participant: Laurent Amsaleg.

Work performed in the framework of the CNRS PICS MMAnalytics, and in collaboration with Marcel Worring, Univeristy of Amsterdam (The Netherlands)

Digital photo collections—personal, professional, or social—have been growing ever larger, leaving users overwhelmed. It is therefore increasingly important to provide effective browsing tools for photo collections. Learning from the resounding success of multi-dimensional analysis (MDA) in the business intelligence community for on-line analytical processing (OLAP) applications, we proposed a multi-dimensional model for media browsing, called M3, that combines MDA concepts with concepts from faceted browsing [21]. We present the data model and describe preliminary evaluations, made using server and client prototypes, which indicate that users find the model useful and easy to use.

6.4.3. NLP-driven hyperlink construction in broadcast videos

Participants: Rémi Bois, Guillaume Gravier, Pascale Sébillot, Anca-Roxana Şimon.

In collaboration with Sien Moens (Katholieke Universiteit Leuven, Belgium), Éric Jamet and Martin Ragot (Univ. Rennes 2, France).

In the context of the the CominLabs project "Linking media in acceptable hypergraphs" dedicated to the creation of explicit and meaningful links between multimedia documents or fragments of documents, we have introduced a typology of possible links between contents of a multimedia news corpus [32]. While several typologies have been proposed and used by the community, we argue that they are not adapted to rich and large corpora which can contain texts, videos, or radio stations recordings. We have defined a new typology, as a first step towards automatically creating and categorizing links between documents' fragments in order to create new ways to navigate, explore, and extract knowledge from large collections.

We also investigated video hyperlinking based on speech transcripts, leveraging a hierarchical topical structure to address two essential aspects of hyperlinking, namely, serendipity control and link justification [26]. We proposed and compared different approaches exploiting a hierarchy of topic models as an intermediate representation to compare the transcripts of video segments. These hierarchical representations offer a basis to characterize the hyperlinks, thanks to the knowledge of the topics which contributed to the creation of the links, and to control serendipity by choosing to give more weights to either general or specific topics. Experiments have been performed on BBC videos from the Search and Hyperlinking task at MediaEval. Link precisions similar to those of direct text comparison have been achieved however exhibiting different targets along with a potential control of serendipity.

The Search and Anchoring in Video Archives task at MediaEval addressed two issues: The Search part aims at returning a ranked list of video segments that are relevant to a textual user query; The Anchoring part focuses on identifying video segments that would encourage further exploration within the archive. Capitalizing on the experience acquired in previous participations, we implemented a two step approach for both sub-tasks [27]. The first step, common to both, consists in generating a list of potential anchor segments and response-query segments relying on a hierarchical topical structuring technique. In the second step, for each query, the best 20 segments are selected according to content-based comparisons, while for the anchor detection sub-task, the segments are ranked based on a cohesion measure. The use of a hierarchical topical structure helps to propose segments of variable length at different levels of details with precise jump-in points for them. More, the algorithm deriving the structure relies on the burstiness phenomenon in word occurrences which gives an advantage over the classical bag-of-words model.

6.4.4. Information extraction

Participants: Vincent Claveau, Ewa Kijak.

In collaboration with X. Tannier (LIMSI), A. Vilnat (LIMSI) and B. Arnulphy (ANR).

Identifying events from texts is an information extraction task necessary for many NLP applications. Through the TimeML specifications and TempEval challenges, it has received some attention in the last years; yet, no reference result is available for French. In [12], we try to fill this gap by proposing several event extraction systems, combining for instance Conditional Random Fields, language modeling and k-nearest-neighbors. These systems are evaluated on French corpora and compared with state-of-the-art methods on English. The very good results obtained on both languages validate our whole approach.

6.5. 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 3D tracking

Participants: Marie-Odile Berger, Jaime Garcia Guevara, Nazim Haouchine, Gilles Simon, Frédéric Sur.

Pose initialization Automating the camera pose initialization is still a problem in non instrumented environments. Difficulties originate in the possibly large viewpoint changes between the data stored in the model and the current view. In this context, Pierre Rolin's PhD work concerns viewpoint simulation techniques for localization. The idea is to generate keypoint descriptors from simulated views in order to enrich the model and to ease the matching of the current view to the model. We have demonstrated the effectiveness of this technique in several situations, either under an affine or a perspective camera model [17], [21]. The computed pose is more stable when it is difficult to obtain reliable correspondences between the model and the current view. In addition, several examples show that our method successfully computes the camera pose whereas the traditional methods fail. Our recent work concerns a progressive sampling strategy to speed up the search of correspondences when confronted to a large outlier rate, which is inherent to viewpoint simulation. We also currently investigate the localization of the virtual camera from which viewpoints should be simulated.

AR in urban environments

Pose initialization is especially difficult in urban scenes due to the presence of repeated patterns. Another difficulty originates in the fact that a pedestrian is free of his motion in the scene and can therefore adopt uncontrolled viewpoints (close or distant views) with respect to the model. As a result, the set of 2D/3D correspondence hypotheses may contain a high ratio of outliers which may lead to erroneous pose computation. In order to improve the matching / recognition stage, we investigated how facades in calibrated images can be orthorectified and delimited by considering prior information about the scene and the camera relevant to AR in urban context [20]. We provide a Bayesian framework to detect vanishing points in Manhattan worlds, which incorporate priors about the Manhattan frame by imposing a near-vertical direction as well as orthogonality constraints. Second, we propose to detect right-angle corners due to windows or doors using a SVM-based machine learning technique. Rectangular facade hypotheses are then generated through min-cuts techniques with the idea to identify rectangles with high density of right-angle corners. Our algorithm performs better or as well as state-of-the-art techniques and is much faster, mainly as a result of using a suitable prior.

Tracking 3D deformable objets

3D augmentation of deformable objects is a challenging problem with many potential applications in computer graphics, augmented reality and medical imaging. Most existing approaches are dedicated to surface augmentation and are based on the inextensibility constraint, for sheet-like materials, or on the use of a model built from representative samples. However, few of them consider in-depth augmentation which is of utmost importance for medical applications. Since the beginning of N. Haouchine's PhD thesis, we have addressed several important limitations that currently hinder the use of augmented reality in the clinical routine of minimally invasive procedures. In collaboration with the MIMESIS team, our main contribution is the design and the validation of an augmented reality framework based on a mechanical model of the organ and guided by features extracted and tracked on the video at the surface of the organ [12]. Specific models which best suit the considered organs, such as a vascularized model of the liver, have been introduced in this framework. Experiments conducted on ex-vivo data of a porcine liver show that the localization error of a virtual tumor were less than 6mm, and thus below the safety margin required by surgery. To our knowledge, we were the first to produce such evaluation for deformable objects.

This work has been extended to augment highly elastic objects in a monocular context. Shape recovery from a monocular video sequence is an underconstrained problem. State-of-the art solutions enforce smoothness or geometric constraints, consider specific deformation properties such as inextensibility or resort to shading constraints. However, few of them can handle properly large elastic deformations. We have proposed [13] a real-time method that uses a mechanical model and is able to handle highly elastic objects. The problem is formulated as an energy minimization problem accounting for a non-linear elastic model constrained by external image points acquired from a monocular camera. This method prevents us from formulating restrictive assumptions and specific constraint terms in the minimization. In addition, we propose to handle self-occluded regions thanks to the ability of mechanical models to provide appropriate predictions of the shape.

The work conducted during N. Haouchine's PhD thesis allowed us to build a complete framework for the use of AR in liver surgery. We now want to focus on specific points to improve the accuracy and the robustness of the augmented process and to facilitate the clinical use of such AR systems. The PhD thesis of Jaime Garcia Guevara started in October on this topic with the aim to build more realistic mechanical models of organs during the surgery (taking into account liver deformation due to insuflation of air during surgery) and to improve the robustness of visual tracking through the use of multiple visual cues and improved methods for outlier detection.

6.2. Image-based modeling

Participants: Marie-Odile Berger, Charlotte Delmas, Antoine Fond, Erwan Kerrien, Gilles Simon, Pierre-Frédéric Villard, Brigitte Wrobel-Dautcourt.

Finding Manhattan directions in uncalibrated images

Finding orthogonal vanishing points (VPs) in a photography has many potential applications in computer vision and computer graphics, including perspective correction, image-based reconstruction and texture extraction. Surprisingly, while this problem has been extensively studied in the literature, manual solutions are still used in most software. Existing algorithms generally follow two steps. First, lines are grouped into pencils, whose centers are considered as potential VPs. Then, an orthogonality measure is evaluated for every triplet of VPs and the most plausible triplet is used for camera calibration. A drawback of this approach is that complex and time-consuming techniques have to be used to solve the general problem of VP detection, while only three particular VPs are finally used. By contrast, we propose an effective and easy-to-implement algorithm that estimates the zenith and the horizon line before detecting the VPs, using simple properties of the central projection and exploiting accumulations of oriented segments around the horizon. Our method is fast and yields an accuracy comparable, and even better in some cases, to that of state-of-the-art algorithms. This work was submitted to Eurographics 2016.

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 but the vasculature and no other brain tissue. 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. She first devised a solution in a two-view reconstruction context. A paper reporting on this work was published and an oral presentation was made at SPIE Medical Imaging 2015 [19]. The focus was made this year on reconstructing the first coil, a single wire that tangles into a complex pattern when deployed in an aneurysm. The challenge is to produce a 3D reconstruction with as little X-ray dose and acquisition time as possible. Two paths were simultaneously followed this year: 1) design and compare various configurations to rapidly shoot 6 X-ray views from different viewpoints with a biplane (stereo) system; 2) compensate the lack of information (small number of images) with a prior, incorporated in the tomographic reconstruction algorithm, to express the sparsity in space of the shape to be reconstructed. Preliminary work sets forward a simultaneous fast rotation of both planes around the patient's head. A database is currently being acquired for a full validation in a view to submitting this work for publication early next year.

Individual-specific heart valve modeling
Mitral valve surgical repair is a complex procedure where the outcome largely depends on the surgeons' experience. Predicting a good coaptation of the two leaflets post-operatively is one of the most difficult sub-tasks in the procedure. We worked on a biomechanical simulation framework [25] that computes the leaflet deformation and coaptation based on individual-specific microtomography data as an initial step toward patient-based mitral valve surgical repair assistance through simulation. Results from FEM analysis on three explanted porcine hearts showed that it is possible to obtain the real shape of the leaflets during systolic peak. We also measured the influence of the positions of chordae tendineae on the simulation and showed that marginal chordae have a greater influence on the final shape than intermediary chordae.

Quasi-periodic noise removal
Images may be affected by quasi-periodic noise. This undesirable feature manifests itself by spurious repetitive patterns covering the whole image, well localized in the Fourier domain. While notch filtering permits to get rid of this phenomenon, this however requires to first detect the resulting Fourier spikes, and, in particular, to discriminate between noise spikes and spectrum patterns caused by spatially localized textures or repetitive structures. Several approaches have been investigated. First, we have reviewed the available methods, most of them requiring expert tuning, with applications to experimental mechanics in view [11]. We have also proposed two automated patch-based approaches. A parametric approach has been investigated in [14] (more information available in [26]), based on the detection of noise spikes as statistical outliers to the distribution expected from natural non-noisy patches, which is known to follow the inverse of a power of the frequency. A non-parametric approach, based on a-contrario detection, was also proposed in [22].

6.3. Parameter estimation

Participants: Frédéric Sur, Erwan Kerrien, Raffaella Trivisonne.

Metrological performance assessment in experimental mechanics
A problem of interest in experimental solid mechanics is to estimate displacement and strain maps on the surface of a specimen subjected to a load or a tensile test. Two contactless approaches are available in the literature, based on depositing on the surface of the specimen either a pseudo-periodic grid or a random speckle. Analyzing images taken before and after deformation permits to estimate strain maps. While periodicity permits to make use of Fourier analysis in the first case, digital image correlation (DIC) is used in the second case. Concerning pseudo-periodic grids, we have investigated noise reduction techniques. While averaging a series of images is certainly the most basic option to reduce the noise, it is impossible to get rid of residual vibrations carried for instance by concrete floor slabs. We have shown in [16] that, while these vibrations indeed blur grid images, they still permit to reduce the noise amplitude in the displacement and strain maps. Concerning DIC-based techniques, we have investigated the effect of sensor noise on the measurement resolution. Since displacement of interest are most of the time below one pixel, interpolation plays a major role. We have proposed a new resolution formula which takes interpolation into account. Besides, this formula has been the subject of an experimental assessment on real data, in the presence of signal-dependent noise [24], [18].

Sensor noise measurement
We have investigated in [15] (additional information available in [27]) the problem of sensor parameter estimation from a series of images, under illumination flickering and vibrations. Illumination flickering is indeed a natural assumption for indoor artificial lights. It is also involved by slight variations in the opening time of a mechanical shutter. We have proposed a model of the pixel intensity based on a Cox process, together with an algorithm which, taking benefit of flickering, gives an estimation of every sensor parameter, namely the gain, the readout noise, and the offset.

Image driven simulation

The IDeaS ANR project targets image-driven simulation, applied to interventional neuroradiology: 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). Our guideline is to follow a sequential statistical filtering approach to fuse such heterogeneous data.

Christo Gnonnou was hired as an engineer (located at Inria Lille in Defrost team (ex-Shacra), contract started on January 1st and ended on October 31st). He continued the work to specify which parameters the simulation is sensitive to, in a view to design a reduced parametric model of the device, and associate covariances to its parameters. He also worked on inverting the mechanical parameters of any device, using our experimental setup based on a high speed stereo rig.

Maxime Malgras worked on his Master's thesis in the team. His investigations aimed at designing a particle filter where the transition function is approximated by a polynomial chaos (PC) instead of particles. It appeared that PC is adapted to compute the posterior in a particle filter but it is not clear whether the number of samples required to estimate the PC coefficients is smaller than the number of particles required for the filter to be accurate, which questions the capacity of PC to reduce the computational burden of particle filters in high dimensions. Raffaella Trivisonne started her PhD thesis in November (co-supervised by Stéphane Cotin, from MIMESIS team in Strasbourg) to investigate deeper on this subject and apply data assimilation to image driven simulation of endovascular interventions.

MORPHEO Project-Team

7. New Results

7.1. QuickCSG: Arbitrary and Faster Boolean Combinations of N Solids

While studied over several decades, the computation of boolean operations on polyhedra is almost always addressed by focusing on the case of two polyhedra. For multiple input polyhedra and an arbitrary boolean operation to be applied, the operation is decomposed over a binary CSG tree, each node being processed separately in quasilinear time. For large trees, this is both error prone due to intermediate geometry and error accumulation, and inefficient because each node yields a specific overhead. We introduce a fundamentally new approach to polyhedral CSG evaluation, addressing the general N-polyhedron case. We propose a new vertex-centric view of the problem, which both simplifies the algorithm computing resulting geometric contributions, and vastly facilitates its spatial decomposition. We then embed the entire problem in a single KD-tree, specifically geared toward the final result by early pruning of any region of space not contributing to the final surface. This not only improves the robustness of the approach, it also gives it a fundamental speed advantage, with an output complexity depending on the output mesh size instead of the input size as with usual approaches. Complemented with a task-stealing parallelization, the algorithm achieves breakthrough performance, one to two orders of magnitude speedups with respect to state-of-the-art CPU algorithms, on boolean operations over two to several dozen polyhedra. The algorithm is also shown to outperform recent GPU implementations and approximate discretizations, while producing a topologically exact output without redundant facets. This algorithm was published as Inria research report [16].

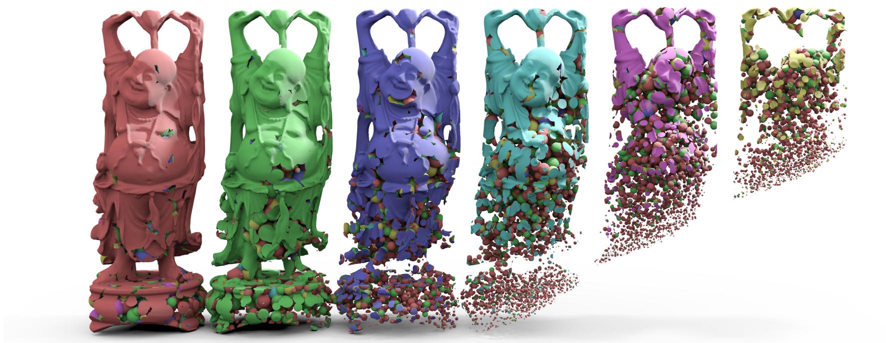


Figure 3. Intersection of 6 Buddhas with the union of 100,000 spheres (total 24 million triangles). Computed in 8 seconds on a desktop machine [16]

7.2. An Efficient Volumetric Framework for Shape Tracking

Recovering 3D shape motion using visual information is an important problem with many applications in computer vision and computer graphics, among other domains. Most existing approaches rely on surface-based strategies, where surface models are fit to visual surface observations. While numerically plausible, this paradigm ignores the fact that the observed surfaces often delimit volumetric shapes, for which deformations are constrained by the volume inside the shape. Consequently, surface-based strategies can fail when the observations define several feasible surfaces, whereas volumetric considerations are more restrictive with respect to the admissible solutions. In this work, we investigate a novel volumetric shape parametrization to track shapes over temporal sequences. In constrast to Eulerian grid discretizations of the observation space, such as voxels, we consider general shape tesselations yielding more convenient cell decompositions, in particular the Centroidal Voronoi Tesselation. With this shape representation, we devise a tracking method that exploits volumetric information, both for the data term evaluating observation conformity, and for expressing deformation constraints that enforce prior assumptions on motion. Experiments on several datasets demonstrate similar or improved precisions over state-of-the-art methods, as well as improved robustness, a critical issue when tracking sequentially over time frames. This work was accepted as **oral** at CVPR 2015 (less than 3% acceptance rate) [8].

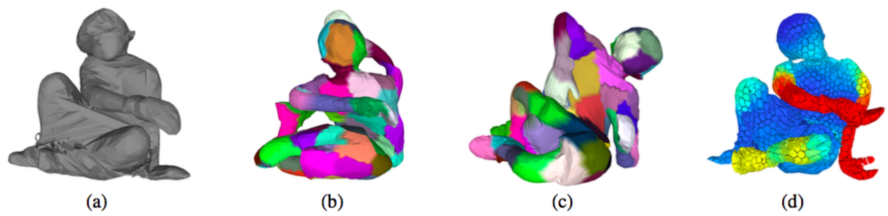


Figure 4. Frames of the GOALKEEPER *dataset acquired on the Kinovis platform. (a) Visual hull input. (b) Tracking result of Cagniart et al. 2010. (c) Allain et al. 2014. (d) This method [8]. Note the improved angular shapes and the improved robustness.*

7.3. Sparse Multi-View Consistency for Object Segmentation

Multiple view segmentation consists in segmenting objects simultaneously in several views. A key issue in that respect and compared to monocular settings is to ensure propagation of segmentation information between views while minimizing complexity and computational cost. In this work, we first investigate the idea that examining measurements at the projections of a sparse set of 3D points is sufficient to achieve this goal. The proposed algorithm softly assigns each of these 3D samples to the scene background if it projects on the background region in at least one view, or to the foreground if it projects on foreground region in all views. Second, we show how other modalities such as depth may be seamlessly integrated in the model and benefit the segmentation. The paper exposes a detailed set of experiments used to validate the algorithm, showing results comparable with the state of art, with reduced computational complexity. We also discuss the use of different modalities for specific situations, such as dealing with a low number of viewpoints or a scene with color ambiguities between foreground and background. This work was published as article in the PAMI journal [3].

7.4. Building Statistical Shape Spaces for 3D Human Modeling

Statistical models of 3D human shape and pose learned from scan databases have developed into valuable tools to solve a variety of vision and graphics problems. Unfortunately, most publicly available models are of limited expressiveness as they were learned on very small databases that hardly reflect the true variety in human body shapes. In this paper, we contribute by rebuilding a widely used statistical body representation from the largest commercially available scan database, and making the resulting model available to the community (visit http://humanshape.mpi-inf.mpg.de). As preprocessing several thousand scans for learning the model is a challenge in itself, we contribute by developing robust best practice solutions for scan alignment that quantitatively lead to the best learned models. We make implementations of these preprocessing steps also publicly available. We extensively evaluate the improved accuracy and generality of our new model, and show its improved performance for human body reconstruction from sparse input data. This work was published as Max Planck research report [17].

Figure 5. Three views of the PLANT *dataset as processed by our method for mutli-view silhouette extraction [3].*

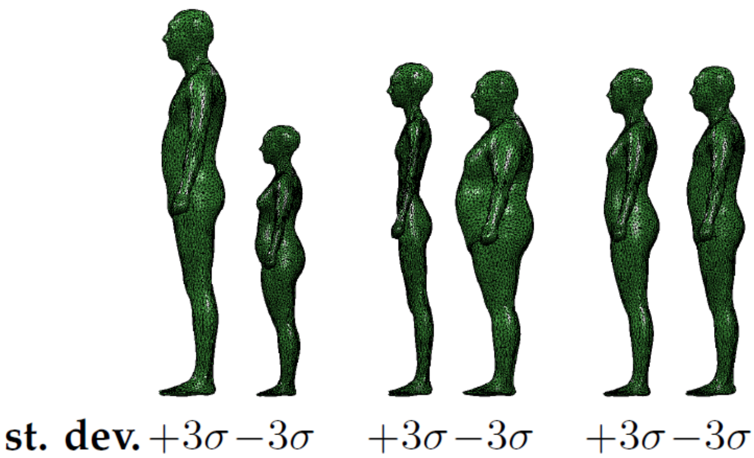


Figure 6. Visualization of the first three principal components learned from a large database of posture-normalized 3D human body scans [17].

7.5. A Groupwise Multilinear Correspondence Optimization for 3D Faces

Multilinear face models are widely used to model the space of human faces with expressions. For databases of 3D human faces of different identities performing multiple expressions, these statistical shape models decouple identity and expression variations. To compute a high-quality multilinear face model, the quality of the registration of the database of 3D face scans used for training is essential. Meanwhile, a multilinear face model can be used as an effective prior to register 3D face scans, which are typically noisy and incomplete. Inspired by the minimum description length approach, we propose the first method to jointly optimize a multilinear model and the registration of the 3D scans used for training. Given an initial registration, our approach fully automatically improves the registration by optimizing an objective function that measures the compactness of the multilinear model, resulting in a sparse model. We choose a continuous representation for each face shape that allows to use a quasi-Newton method in parameter space for optimization. We show that our approach is computationally significantly more efficient and leads to correspondences of higher quality than existing methods based on linear statistical models. This allows us to evaluate our approach on large standard 3D face databases and in the presence of noisy initializations. This work was published at the ICCV conference [9].

7.6. A statistical shape space model of the palate surface trained on 3D MRI scans of the vocal tract

We describe a minimally-supervised method for computing a statistical shape space model of the palate surface. The model is created from a corpus of volumetric magnetic resonance imaging (MRI) scans collected from 12 speakers. We extract a 3D mesh of the palate from each speaker, then train the model using principal component analysis (PCA). The palate model is then tested using 3D MRI from another corpus and evaluated using a high-resolution optical scan. We find that the error is low even when only a handful of measured coordinates are available. In both cases, our approach yields promising results. It can be applied to extract the palate shape from MRI data, and could be useful to other analysis modalities, such as electromagnetic articulography (EMA) and ultrasound tongue imaging (UTI). This work was published at the 18th International Congress of Phonetic Sciences [11].

7.7. Toward User-specific Tracking by Detection of Human Shapes in Multi-Cameras

Human shape tracking consists in fitting a template model to temporal sequences of visual observations. It usually comprises an association step, that finds correspondences between the model and the input data, and a deformation step, that fits the model to the observations given correspondences. Most current approaches find their common ground with the Iterative-Closest-Point (ICP) algorithm, which facilitates the association step with local distance considerations. It fails when large deformations occur, and errors in the association tend to propagate over time. In this paper, we propose a discriminative alternative for the association, that leverages random forests to infer correspondences in one shot. It allows for large deformations and prevents tracking errors from accumulating. The approach is successfully integrated to a surface tracking framework that recovers human shapes and poses jointly. When combined with ICP, this discriminative association proves to yield better accuracy in registration, more stability when tracking over time, and faster convergence. Evaluations on existing datasets demonstrate the benefits with respect to the state-of-the-art. This work was published at CVPR 2015 [12].

7.8. Video based Animation Synthesis with the Essential Graph

We propose a method to generate animations using video-based mesh sequences of elementary movements of a shape. New motions that satisfy high-level user-specified constraints are built by recombining and interpolating the frames in the observed mesh sequences. The interest of video based meshes is to provide real full shape information and to enable therefore realistic shape animations. A resulting issue lies, however, in the difficulty to combine and interpolate human poses without a parametric pose model, as with skeleton based animations. To address this issue, our method brings two innovations that contribute at different levels: Locally between two motion sequences, we introduce a new approach to generate realistic transitions using dynamic time warping; More globally, over a set of motion sequences, we propose the essential graph as an efficient structure to encode the most realistic transitions between all pairs of input shape poses. Graph search in the essential graph allows then to generate realistic motions that are optimal with respect to various user-defined constraints. We present both quantitative and qualitative results on various 3D video datasets. They show that our approach compares favourably with previous strategies in this field that use the motion graph. This work was published at the 3DV 2015 conference [10].

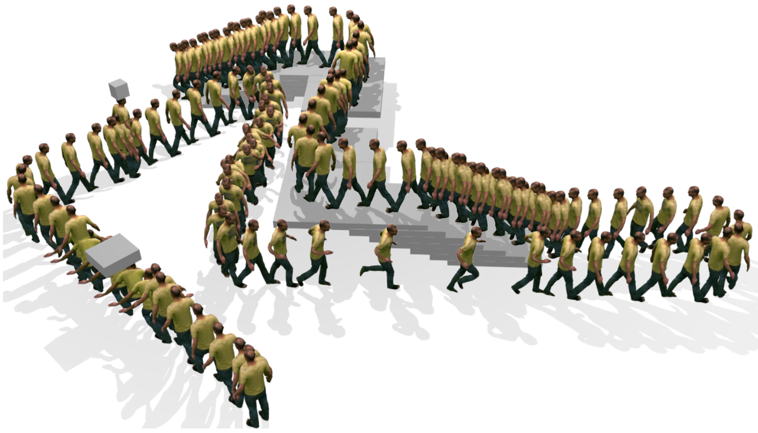


Figure 7. Example of 4D animation generated using by combining recorded 4D sequences [10].

7.9. Implicit B-Spline Surface Reconstruction

This paper presents a fast and flexible curve/surface reconstruction technique based on implicit b-spline. This representation does not require any parameterization and it is locally supported. This fact has been exploited in this paper to propose a reconstruction technique through solving a sparse system of equations. This method is further accelerated to reduce the dimension to the active control lattice. Moreover, the surface smoothness and user interaction are allowed for controlling the surface. Finally, a novel weighting technique has been introduced in order to blend small patches and smooth them in the overlapping regions. The whole framework is very fast and efficient and can handle large cloud of points with low computational cost. The experimental results show the flexibility and accuracy of the proposed algorithm to describe objects with complex topologies. Comparisons with other fitting methods highlight the superiority of the proposed approach in the presence of noise and missing data. This work was published as journal article in IEEE Transactions on Image Processing [6].

7.10. A Bayesian Approach to Multi-view 4D Modeling

This paper considers the problem of automatically recovering temporally consistent animated 3D models of arbitrary shapes in multi-camera setups. An approach is presented that takes as input a sequence of frame-wise reconstructed surfaces and iteratively deforms a reference surface such that it fits the input observations. This approach addresses several issues in this field that include: large frame-to-frame deformations, noise, missing data, outliers and shapes composed of multiple components with arbitrary geometries. The problem is cast as a geometric registration with two major features. First, surface deformations are modeled using mesh decomposition into elements called patches. This strategy ensures robustness by enabling flexible regularization priors through inter-patch rigidity constraints. Second, registration is formulated as a Bayesian estimation that alternates between probabilistic datal-model association and deformation parameter estimation. This accounts for uncertainties in the acquisition process and allows for noise, outliers and missing geometries in the observed meshes. In the case of marker-less 3D human motion capture, this framework can be specialized further with additional articulated motion constraints. Extensive experiments on various 4D datasets show that complex scenes with multiple objects of arbitrary nature can be processed in a robust way. They also demonstrate that the framework can capture human motion and provides visually convincing as well as quantitatively reliable human poses. This work was published as journal article in International Journal on Computer Vision (IJCV) [4].

7.11. A Hierarchical Approach for Regular Centroidal Voronoi Tessellations

In this paper we consider Centroidal Voronoi Tessellations (CVTs) and study their regularity. CVTs are geometric structures that enable regular tessellations of geometric objects and are widely used in shape modeling and analysis. While several efficient iterative schemes, with defined local convergence properties, have been proposed to compute CVTs, little attention has been paid to the evaluation of the resulting cell decompositions. In this paper, we propose a regularity criterion that allows us to evaluate and compare CVTs independently of their sizes and of their cell numbers. This criterion allows us to compare CVTs on a common basis. It builds on earlier theoretical work showing that second moments of cells converge to a lower bound when optimising CVTs. In addition to proposing a regularity criterion, this paper also considers computational strategies to determine regular CVTs. We introduce a hierarchical framework that propagates regularity over decomposition levels and hence provides CVTs with provably better regularities than existing methods. We illustrate these principles with a wide range of experiments on synthetic and real models.

This work was published as a journal article in Computer Graphics Forum [7].

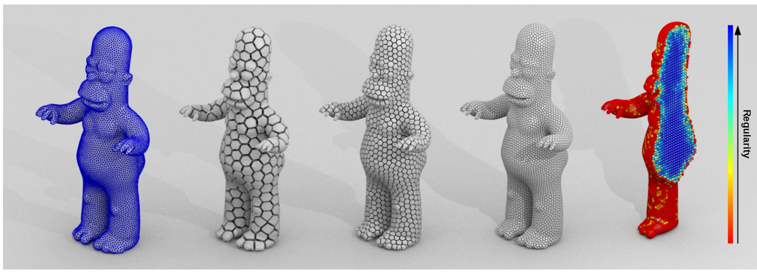


Figure 8. Hierarchical computation of a centroidal Voronoi tessellation from a 3D mesh [7]. Inside cells are very regular.

7.12. Just Noticeable Distortion Profile for Flat-Shaded 3D Mesh Surfaces

It is common that a 3D mesh undergoes some lossy operations (e.g., compression, watermarking and transmission through noisy channels), which can introduce geometric distortions as a change in vertex position. In most cases the end users of 3D meshes are human beings; therefore, it is important to evaluate the visibility of introduced vertex displacement. In this paper we present a model for computing a Just Noticeable Distortion (JND) profile for flat-shaded 3D meshes. The proposed model is based on an experimental study of the properties of the human visual system while observing a flat-shaded 3D mesh surface, in particular the contrast sensitivity function and contrast masking. We first define appropriate local perceptual properties on 3D meshes. We then detail the results of a series of psychophysical experiments where we have measured the threshold needed for a human observer to detect the change in vertex position. These results allow us to compute the JND profile for flat-shaded 3D meshes. The proposed JND model has been evaluated via a subjective experiment, and applied to guide 3D mesh simplification as well as to determine the optimal vertex coordinates quantization level for a 3D model.

This work was published as a journal article in IEEE Transactions on Visualization and Computer Graphics [5].

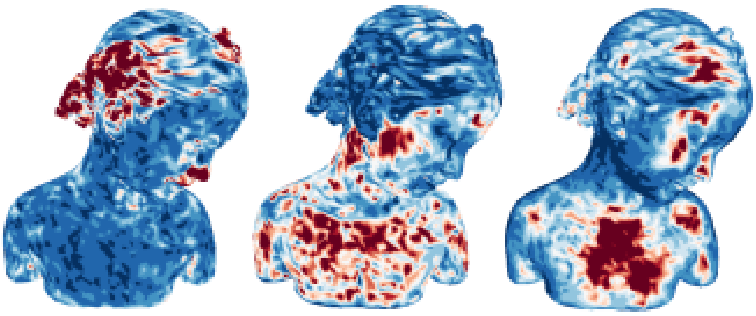


Figure 9. Just noticeable distortion profile in a light independent mode (left, middle) or with a light fixed in front of the model (right), for vertex displacements in the normal direction (left, right) or in the tangent direction (middle) [5].

PERCEPTION Project-Team

6. New Results

6.1. Supervised Audio-Source Localization

We addressed the problem of localizing audio sources using binaural measurements. After proposing an unsupervised method [20], we proposed a supervised formulation that simultaneously localizes multiple sources at different locations [22]. The approach is intrinsically efficient because, contrary to prior work, it relies neither on source separation, nor on monaural segregation. The method starts with a training stage that establishes a locally-linear Gaussian regression [21] between the directional coordinates of all the sources and the auditory features extracted from binaural measurements. 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 release 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. More recently the method has been extended to an arbitrary number of microphones [35], [34]. Moreover, we have started to develop a method that extracts the direct path on an acoustic wave in order to enable robust audio-source localization in reverberant environments [40].

Websites:
https://team.inria.fr/perception/research/acoustic-learning/
https://team.inria.fr/perception/research/binaural-ssl/
https://team.inria.fr/perception/research/local-rtf/

6.2. Multichannel 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. 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 learn in advance, e.g. [37], 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). It received the best student paper award at WASPAA'15 [31]. An extended version has been submitted to IEEE Transactions on Audio, Speech, and Language Processing [39].

6.3. Audio-Visual Speaker Tracking and Recognition

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 diarization and speaker recognition 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 [38], 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 trackers [29], [30] are tested on challenging data-sets that are available from recent contributions which are used as baselines for comparison. Currently we are developing a variational framework for the on-line tracking of multiple persons [36].

Websites:
https://team.inria.fr/perception/research/speakerloc/
https://team.inria.fr/perception/research/speechturndet/
https://team.inria.fr/perception/research/avdiarization/

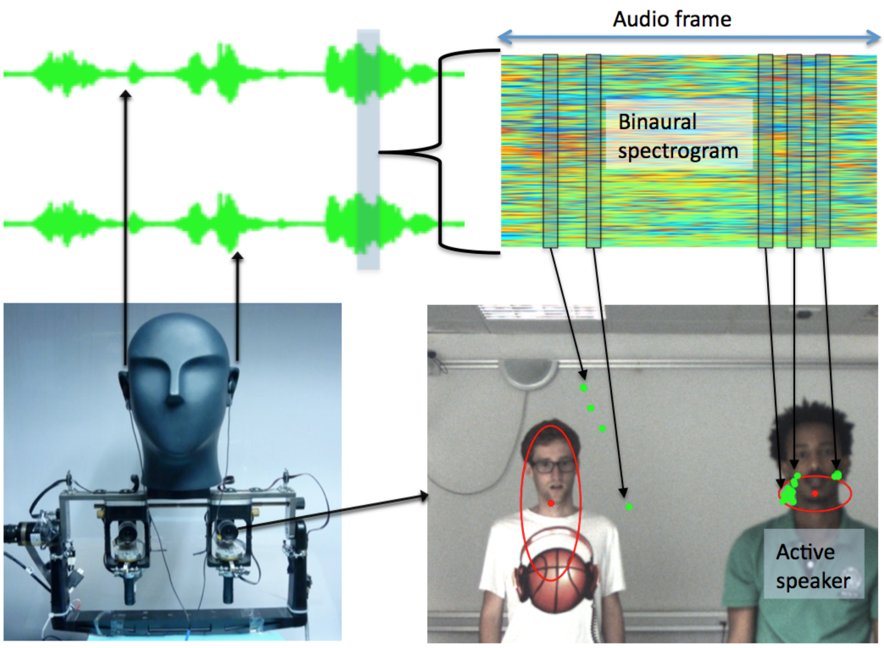


Figure 6. This figures illustrates the general principle of our audio-visual speaker tracking and diarization method. The auditory and visual data are recorded with two microphones and one camera. The audio signals are segmented into frames and each frame (vertical grey rectangle) is transformed into a binaural spectrogram [20]. This spectrogram is composed of a sequence of binaural vectors (vertical rectangles) and each binaural vector is mapped onto a sound-source direction which corresponds to a point in the image plane (green dots) [22]. The proposed audio-visual tracker associates people detected in the image sequence with these sound directions via audio-visual clustering [38] that is combined with an active-speaker transition model.

6.4. Head Pose Estimation

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 [21]. 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. 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 [28]. Currently we investigate a temporal extension of this model.

Website:
https://team.inria.fr/perception/research/head-pose/

6.5. High-Resolution Scene Reconstruction

We addressed the problem of range-stereo fusion for the construction of high-resolution depth maps. In particular, we combine low-resolution depth 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 [23]. This work is funded by the ANR project MIXCAM.

Website:
https://team.inria.fr/perception/research/dsfusion/

6.6. Hyper-Spectral Image Analysis

As an extension to our work on high-dimensional regression [21] we addressed the problem of analyzing hyper-spectral data. In particular we addressed the problem of recovering physical properties (parameters) form hyper-spectral low-resolution images, i.e. at large planetary scales. This involves resolving inverse problems which can be addressed within machine learning, with the advantage that, once a relationship between physical parameters and spectra has been established in a data-driven fashion, the learned relationship can be used to estimate physical parameters for new hyper-spectral observations. Within this framework, we propose a spatially-constrained and partially-latent regression method which maps high-dimensional inputs (hyper-spectral images) onto low-dimensional responses (physical parameters such as the local chemical composition of the soil). The proposed regression model comprises two key features. Firstly, it combines a Gaussian mixture of locally-linear mappings (GLLiM) with a partially-latent response model. While the former makes high-dimensional regression tractable, the latter enables to deal with physical parameters that cannot be observed or, more generally, with data contaminated by experimental artifacts that cannot be

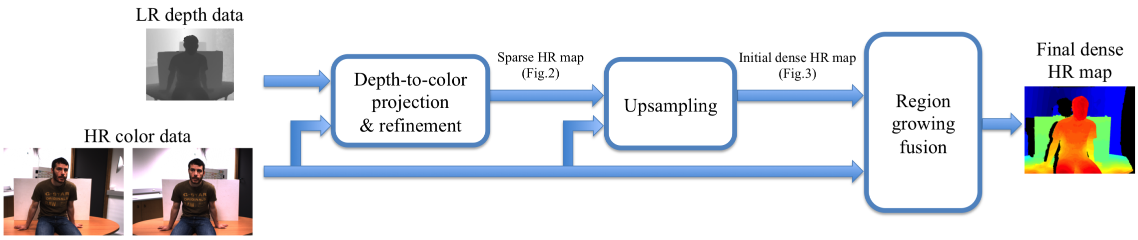


Figure 7. The pipeline of the proposed depth-stereo fusion method [23]. The low-resolution (LR) depth data are projected onto the color data and refined to yield a high-resolution (HR) sparse disparity map. Starting from these disparity seeds, an upsampling process provides an initial HR dense disparity map. Both the HR seeds and the initial dense disparity map are then used by the region-growing depth-stereo fusion to produce the final HR depth map. A prominent feature of our method is that fusion takes place at several data processing stages.

explained with noise models. Secondly, spatial constraints are introduced in the model through a Markov random field (MRF) prior which provides a spatial structure to the Gaussian-mixture hidden variables [19]. Experiments conducted on a database composed of remotely sensed observations collected from the Mars planet by the Mars Express orbiter demonstrate the effectiveness of the proposed model.

6.7. Gaussian Mixture Regression for Acoustic-Articulatory Inversion

The team expertise in latent-variable mixture models was applied to the problem of adaptation of an acoustic-articulatory model of a reference speaker to the voice of another speaker, using a limited amount of audio-only data [25]. In the context of pronunciation training, a virtual talking head displaying the internal speech articulators (e.g., the tongue) could be automatically animated by means of such a model using only the speaker's voice. In this study, the articulatory-acoustic relationship of the reference speaker is modeled by a gaussian mixture model (GMM). To address the speaker adaptation problem, we propose a new framework called cascaded Gaussian mixture regression (C-GMR), and derive two implementations. The first one, referred to as Split-C-GMR, is a straightforward chaining of two distinct GMRs: one mapping the acoustic features of the source speaker into the acoustic space of the reference speaker, and the other estimating the articulatory trajectories with the reference model. In the second implementation, referred to as Integrated-C-GMR, the two mapping steps are tied together in a single probabilistic model. For this latter model, we present the full derivation of the exact EM training algorithm, that explicitly exploits the missing data methodology of machine learning. Other adaptation schemes based on maximum-a posteriori (MAP), maximum likelihood linear regression (MLLR) and direct cross-speaker acoustic-to-articulatory GMR are also investigated. Experiments conducted on two speakers for different amount of adaptation data show the interest of the proposed C-GMR techniques. This work was done in collaboration with Thomas Hueber and Gérard Bailly from Gipsa Lab and with Xavier Alameda-Pineda from University of Trento and former team member.

PRIMA Project-Team

6. New Results

6.1. Attention-Based Navigation

Participants: Thierry Fraichard, Remi Paulin, Patrick Reignier.

The domain of service-robots is growing fast and has become the focus of many researchers and industrials alike. Application areas are extremely broad, from logistics to handicap assistance. A large proportion of such robots are expected to share humans' living space and thus must be endowed with navigation capabilities that exceed the standard requirements pertaining to autonomous navigation such as motion safety. In a human populated environment, optimality does not boil down to minimizing resources such as time or distance traveled anymore, the robot motion must abide by social rules and move in a manner which is **appropriate**.

Most of the approaches proposed so far rely upon the definition of so-called **social spaces**, *i.e.* regions in the environment that, for different reasons, the persons consider as psychologically theirs. Such social spaces are primarily characterized using either the position of the person, *e.g.* "Personal space" [36], or the activity he is currently engaged in, *e.g.* "Interaction Space" [41] and "Activity Space" [45]. The most common approach is then to define costmaps on such social spaces: the higher the cost, the less desirable it is for the robot to be at the corresponding position. The costmaps are ultimately used for motion planning and navigation purposes.

While improving upon the standard "non social" navigation methods, this type of approach intrinsically ignores the correlations between interactions as well as the influence of the robot on those interactions. It thus fails to capture several important features of social navigation, such as the distraction and surprise caused to the surrounding individuals. To overcome those limits, we suggest using the psychological concept of attention, which plays a central role when humans navigate around each other. This concept brings a new degree of control over the motion of the robot, namely the invasive and distracting character of the robot motion, which have so far proven hard to tackle with the conventional tools such as social spaces. Beside leading appropriate motion, attention-based navigation enables interaction through motion by predicting the quantity of attention the human will give to the robot.

Building upon a computational model of attention that was earlier proposed in [47], we have developed the novel concept of **attention field**. The attention field is straightforward to define: it is a measure of the amount of attention that a given person would allocate to the robot, should the robot be in a given position/state. It is a mapping from the state space of the robot to $\mathbb{R}$. We use this attention field in order to carefully control the degree of distraction caused by the robot to the individuals in its surroundings. By monitoring the variations of attentional resources that it causes, we also control the amount of surprise caused by the robot which must be kept to a minimum since it is a cause of discomfort. Furthermore this approach enables us to tackle more complex situations where more than one person is involved such as the task of delivering a private message to an individual, or else joining a group (an example of interaction through motion). Rather than navigating on a single global costmap, this new approach provides for each path several measures of the distraction and surprise caused by the robot on a given individual. Those quantities are then multi-optimised in order to find a path that satisfies all the given requirements for fulfilling the robot's task as well as minimizing the discomfort for individuals who are not directly involved in an interaction with the robot.

In 2015, we have developed a variant of the well-known differential evolution algorithm which deals with optimizing continuous trajectories under multiple constraints. The performance of our approach is now being compared with trajectories obtained by relying only on social spaces. Besides the traditional qualitative approaches to evaluate the discomfort caused by the robot motion, we work on defining more quantitative measures that would enable us to further validate our approach.

6.2. SPOK: End User Programming for Smart Homes

Participant: Alexandre Demeure.

As part of the CATRENE project AppsGate, we have developed SPOK, an End User Development Environment, that enables inhabitants to control and program their smart Homes via a web interface. The current version of SPOK includes an editor for editing programs using a pseudo-natural language and an interpreter. A multi-syntax editor as well as additional services such as a debugger and a simulator are expected for the second version.

A multi-syntax editor will allow users to build syntactically correct programs using the syntax that is most appropriate to them or by using a combination of them. These syntaxes include pseudo-natural language (i.e. a constrained natural language) and graphical iconic syntax (as exemplified by Scratch [Maloney et al. 2010]). The interaction techniques used to enter programs may be menu-based, free typing, as well as by demonstration in the physical home or by the way of the simulator. The simulator is the dual digital representation of the real home. It is intended to serve also as a debugger for testing and correcting end-user programs.

Whatever syntax used by end-users, programs are translated into syntactic abstract trees whose leaves reference services provided by the Core HMI and/or by the Extended HMI Middleware. The interpreter, executes end-user programs, using the corresponding syntactic abstract trees as input.

In order to support a dynamically extensible grammar as well as to provide end-users with feedforward at the user interface of the editor, the grammar used by the editor is split into 2 parts: the root grammar and the device specific grammars. The root grammar specifies the generic structures of an end-user program: loops, conditions, etc. The device specific grammars are separated from the root grammar to be able to dynamically build the final grammar to be compliant with what is currently installed and detected by the AppsGate server. Each device type brings with it its own events, status and actions. These grammatical elements are injected into the root grammar when generating the parser and for compiling end-user programs.

The language used by end-users to express their programs is a pseudo-natural language using the rule-based programming paradigm. The left hand side of a rule is composed of events and conditions, and the right hand side specifies the actions to be taken when the left hand-side is true or becomes true. A program may include several rules that can be executed either in parallel or sequentially. Once entered, programs are translated into syntactic abstract trees. The interpreter, executes end-user programs, using the corresponding syntactic abstract trees as input. SPOK is implemented as a mix of OSGi and ApAM components where ApAM is in turn a middleware that runs on top of OSGi.

6.3. Qualitative approaches for building energy management

Participant: Patrick Reignier.

Reducing housing energy costs is a major challenge of the 21st century. In the near future, the main issue for building construction is the thermal insulation, but in the longer term, the issues are those of renewable energy (solar, wind, etc.) and smart buildings. Home automation system basically consists of household appliances linked via a communication network allowing interactions for control purposes. Thanks to this network, a load management mechanism can be carried out: it is called distributed control. An optimal home energy management system is still a goal to aim for, because lots of aspects are still not completely fulfilled. Most of the energy systems respect only the energy needs, but they don't tackle the user needs or satisfaction. Energy systems also have a lack when it comes to the dynamicity of the environments (the system ability to adapt).The problem is similar for the existing HMI (Human User Interface) of those Home Automation Systems where only experts can understand the data coming from the sensors and most important, the energy plan coming from management system (How? and Why?). The goal of this study is to propose a house energy model that can be both used to predict at some level energy evolution and that can be understood by the end user. The house energy model is based on Fuzzy Cognitive Maps representing cause-effects relations. It is first designed by an expert and then automatically tuned to a particular house using machine learning approaches. Preliminary experiments have been done this year using the Predis datasets.

6.4. Situation Aware Services on Mobile Devices

Participants: James Crowley, Thibaud Flury.

Modern mobile devices, such as smart phones and tablets, combine a rich set of sensors, internet connectivity, with embedded computational power and memory. The PRIMA group has recently demonstrated that it is possible to construct embedded software that uses the full suite of mobile sensors to recognise activities and learn the daily routines of users.

A first proof of concept has recently been constructed using recognition of places and activities. The system was trained by having student volunteers carry a cell phone running a data acquisition program that recorded signals from accelerometer, gyroscope, ambient sound, ambient light, Cell tower, wifi, bluetooth, and GPS based geolocalisation. The data were labeled by the students with ground truth data about transportation modes, places, and activities. This data was then used to learn recognition routines. Recognition of places, activities, and transportation was used to construct probabilistic models of daily routines using PRIMA's situation modelling techniques, previous demonstrated in constructing situation aware services. The system was demonstrated by constructing a Twitter Bot (a robot that publishes on twitter) that published information about volunteers during their daily activity.

A professional quality software system named CAM - Context Aware Manager - is currently under construction, and will be licensed to the PRIMA startup SItu8ed, for use in context aware mobile services.

6.5. Perceiving mass in mixed reality

Participants: Sabine Coquillart, Paul Issartel.

In mixed reality, real objects can be used to interact with virtual objects. However, unlike in the real world, real objects do not encounter any opposite reaction force when pushing against virtual objects. The lack of reaction force during manipulation prevents users from perceiving the mass of virtual objects. Although this could be addressed by equipping real objects with force-feedback devices, such a solution remains complex and impractical. In this work, we present a technique to produce an illusion of mass without any active force-feedback mechanism. This is achieved by simulating the effects of this reaction force in a purely visual way. A first study demonstrates that our technique indeed allows users to differentiate light virtual objects from heavy virtual objects. In addition, it shows that the illusion is immediately effective, with no prior training. In a second study, we measure the lowest mass difference (JND) that can be perceived with this technique. The effectiveness and ease of implementation of our solution provides an opportunity to enhance mixed reality interaction at no additional cost.

6.6. Pseudo-haptic feedback

Participants: Sabine Coquillart, Jingtao Chen.

"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 Jingtao Chen PhD thesis, a novel pseudo-haptic experiment has been set up. The aim of this experiment is to study the 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 are under way. They will be followed by the tests with subjects.

SIROCCO Project-Team

7. New Results

7.1. Analysis and modeling for compact representation and navigation

3D modelling, multi-view plus depth videos, Layered depth images (LDI), 2D and 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

7.1.1. Visual attention

Participants: Pierre Buyssens, 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 deals with the understanding and modeling of overt attention as well as saliency-based image editing. These research activities are described in the following sections.

Saccadic model: Most of the computation models of visual attention output a 2D static saliency map. This single topographic saliency map which encodes the ability of an area to attract our gaze is commonly computed from a set of bottom-up visual features. Although the saliency map representation is a convenient way to indicate where we look within a scene, these models do not completely account for the complexities of our visual system. One obvious limitation concerns the fact that these models do not make any assumption about eye movements and viewing biases. For instance, they implicitly make the hypothesis that eyes are equally likely to move in any direction.

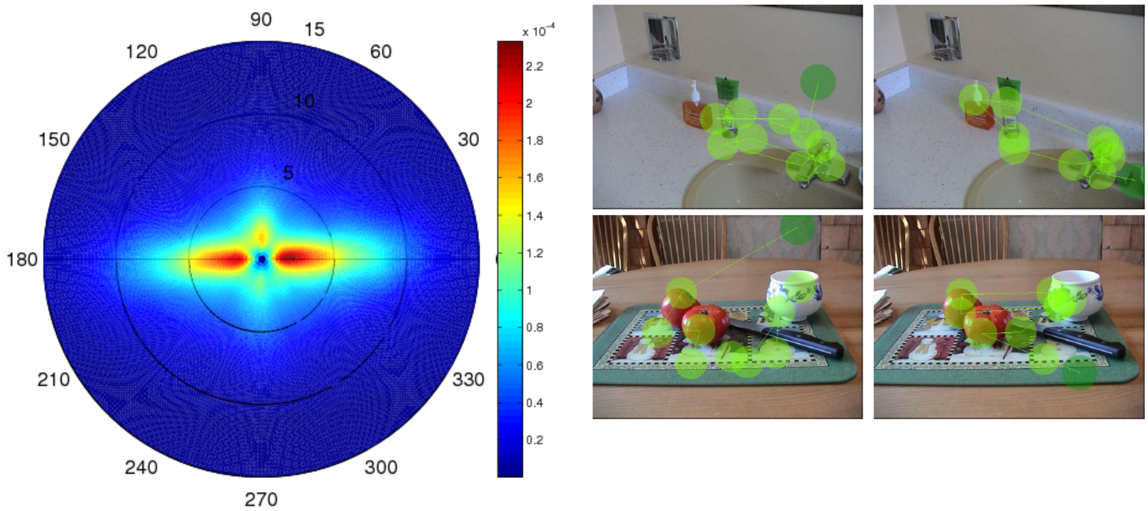
There is evidence for the existence of systematic viewing tendencies. Such biases could be combined with computational models of visual attention in order to better predict where we look. Such a model, predicting the visual scanpath of observer, is termed as saccadic model. We recently propose a saccadic model ([20]) that combines bottom-up saliency maps, viewing tendencies and short-term memory. The viewing tendencies are related to the fact that most saccades are small (less than 3 degrees of visual angle) and oriented in the horizontal direction. Figure 1 (a) illustrates the joint probability distribution of saccade amplitudes and orientations. Examples of predicted scanpaths are shown in Figure 1 (b). We demonstrated that the proposed model outperforms the best state-of-the-art saliency models.

In the future, the goal is to go further by considering that the joint distribution of saccade amplitudes and orientations is spatially variant and depends on the scene category.

Perceptual-based image editing: Since the beginning of October, we have started new studies related to perceptual-based image editing. The goal is to combine the modelling of visual attention with image/video editing methods. More specifically it aims at altering images/video sequences in order to attract viewers attention over specific areas of the visual scene. We intend to design new computational editing methods for emphasizing and optimizing the importance of pre-defined areas of the input image/video sequence. There exist very few studies in the literature dealing with this problem. Current methods simply alter the content by using blurring operation or by recoloring the image locally so that the focus of attention falls within the pre-defined areas of interest. One avenue for improving current methods is to minimize a distance computed between a user's defined visual scanpath and predicted visual scanpath. The content would be edited (i.e. recoloring, region rescaling, local contrast/resolution adjustment, removing disturbing object, etc) in an iterative manner in order to move the focus of attention towards the regions selected by the user.

7.1.2. Epitome-based video representation

Participants: Martin Alain, Christine Guillemot.



(a) (b)

Figure 1. (a) Joint probability distribution of saccade amplitudes and orientations shown on a polar plot. Radial position indicates saccadic amplitudes expressed in degree of visual angle. (b) Predicted scanpaths composed of ten fixations represented by green circles. The dark green circle corresponds to the first fixation which is randomly chosen.

In 2014, we have developed fast methods for constructing epitomes from images. An epitome is a factorized texture representation of the input image, and its construction exploits self-similarities within the image. Known construction methods are memory and time consuming. The proposed methods, using dedicated list construction on one hand and clustering techniques on the other hand, aim at reducing the complexity of the search for self-similarities.

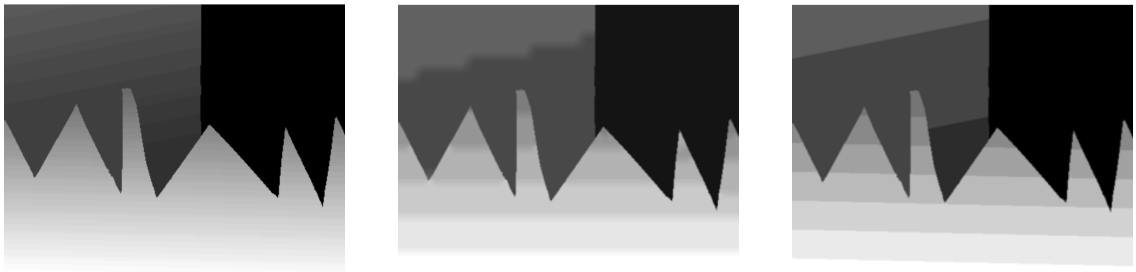
In 2015, we have developed methods for quantization noise removal (after decoding) exploiting the epitome representations together with local learning of either LLE (locally linear embedding) weights, which has proved to be a powerful tool for prediction [14], or using linear mapping functions between original and noisy patches. Compared to classical denoising methods which, most of the time, assume additive white Gaussian noise, the quantization turns out to be correlated to the signal which makes the problem more difficult. The methods have been experimented both in the contexts of single layer encoding and scalable encoding. The same methodology has been applied to super-resolution learning this time mapping functions between the low resolution and high resolution spaces in which lie the patches of the epitome [32].

7.1.3. 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. Our goal is to transmit "just enough" geometry information for accurate representation of a given set of views, and hence better control the effect of geometry lossy compression.



(a) (b) (c)

Figure 2. (a) original depth map, (b) depth map compressed with edge-adaptive method at 10kb with compression artifacts (c) depth image retrieved from the graph of our proposed GBR transmitted at 10kb keeping the original scene structure.

More particularly, in [23], we proposed a new 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 (Fig. 2). This work has first been carried out for multi-view configurations, in which cameras are parallel. We are currently investigating the extension of this promising GBR to complex camera transitions. An algorithm has already been implemented for two views and is being extended for multiple views. The next steps will be to develop color coding tools adapted to these graph structures.

7.2. Rendering, inpainting and super-resolution

image-based rendering, inpainting, view synthesis, super-resolution

7.2.1. Color and light transfer

Participants: Hristina Hristova, Olivier Le Meur.

Color transfer aims at modifying the look of an original image considering the illumination and the color palette of a reference image. It can be employed for image and video enhancement by simulating the appearance of a given image or a video sequence. It can also be applied to hallucinations of particular parts of the day. Current state-of-the-art methods focus mainly on the global transfer of the light and color distributions. Unfortunately, the use of a global distribution is questionable since the light and color of image can significantly vary within the same scene. In [27], we proposed a new method to deal with the limitations of existing methods. The proposed method aims at performing the partitions of the input and reference images into Gaussian distributed clusters by considering the main style of input and reference images. From this clustering, several novel policies are defined for mapping the clusters of the input and reference images. To complete the style transfer, for each pair of corresponding clusters, we apply a parametric color transfer method (i.e. Monge-Kantorovitch transformation) and a local chromatic adaptation transform. Results, subjective user evaluation as well as objective evaluation show that the proposed method obtains visually pleasing and artifact-free images, respecting the reference style. Some results are illustrated in Figure 3 .

Figure 3. From left to right: input image, reference image and the result of the proposed method.

In [34], we extended the method presented in [27] to deal with a color transfer between two HDR images. One limitation of the two proposed methods is that we are still considering that the distributions of color and light follow a Gaussian law. We are currently investigating a more general approach by considering multivariate generalized Gaussian distribution.

7.2.2. Image guided inpainting

Participants: Christine Guillemot, Thomas Maugey.

Inpainting of images has been intensively studied in the past few years, especially for applications such as image restoration and editing [16]. Another application where inpainting techniques are useful is view synthesis, where holes are to be filled corresponding to areas that are no longer occluded. In the particular cases where one has access to ground truth images (like for example in multiview video coding where view synthesis is used for predicting the captured views from a reference one), auxiliary information can be generated to help inpainting, which leads to the concept of *guided inpainting*.

In [29], we have proposed a new auxiliary information that is used to refine the set of candidate patches for the hole filling step of the inpainting. Assuming that the patches of an image lie in a union of subspaces, *i.e.*, the images have different regions with different color textures, these patches are first clustered using a new recursive spectral clustering algorithm that extends existing sparse subspace clustering and replaces the sparse approximation by locally linear embedding, better suited for the inpainting context. Dictionaries are then built from these clusters and used for the hole filling process. However, the inpainting is not always able to "guess" in which cluster the patches of the hole belong to (especially around discontinuities). The auxiliary information that is built from the ground truth image may help to find the right cluster. We thus propose a new guided inpainting algorithm that forces the patch reconstruction to be done in one cluster only, if no auxiliary information is available, or in the cluster pointed by the auxiliary information, if it is available. Experiments (Fig. 4) show that auxiliary information helps to significantly improve the inpainting quality for a reasonable coding cost.

We are currently working on the extension of this technique in order to place the guided inpainting problem in an information theoretic framework, and better answer the following questions: when additional information is actually needed? What type of auxiliary information is needed? how to optimize in a rate-distortion sense the guided inpainting problem?.

(a) (b) (c)

Figure 4. (a) input image to inpaint, (b) filled image using baseline not guided inpainting (c) filled image using proposed guided inpainting with an auxiliary information cost of 0.018 *bpp bitrate.*

7.2.3. Clustering on manifolds for image restoration

Participants: Julio Cesar Ferreira, Christine Guillemot, Elif Vural.

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. We have developed 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.

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. Manifold learning and low dimensional embedding for classification

Participants: Christine Guillemot, Elif Vural.

Typical supervised classifiers such as SVM are designed for generic data types and do not make any particular assumption about the geometric structure of data, while data samples have an intrinsically low-dimensional structure in many data analysis applications. Recently, many supervised manifold learning methods have been proposed in order to take the low-dimensional structure of data into account when learning a classifier. Unlike unsupervised manifold learning methods which only take the geometric structure of data samples into account when learning a low-dimensional representation, supervised manifold learning methods learn an embedding that not only preserves the manifold structure in each class, but also enhances the separation between different classes.

An important factor that influences the performance of classification is the separability of different classes in the computed embedding. We have done a theoretical analysis of separability of data representations given by supervised manifold learning. In particular, we have focused on the nonlinear supervised extensions of the Laplacian eigenmaps algorithm and have examined the linear separation between different classes in the learned embedding. We have shown that, if the graph is such that the inter-group graph weights are sufficiently small, the learned embedding becomes linearly separable at a dimension that is proportional to the number of groups. These theoretical findings have been confirmed by experimentation on synthetic data sets and image data.

We have then considered the problem of out-of-sample generalizations for manifold learning. Most manifold learning methods compute an embedding in a pointwise manner, i.e., data coordinates in the learned domain are computed only for the initially available training data. The generalization of the embedding to novel data samples is an important problem, especially in classification problems. Previous works for out-of-sample generalizations have been designed for unsupervised methods. We have studied this problem for the particular application of data classification and proposed an algorithm to compute a continuous function from the original data space to the low-dimensional space of embedding. In particular, we have constructed an interpolation function in the form of a radial basis function that maps input points as close as possible to their projections onto the manifolds of their own class. Experimental results have shown that the proposed method gives promising results in the classification of low-dimensional image data such as face images.

7.3.2. Adaptive clustering with Kohonen self-organizing maps for second-order prediction

Participants: Christine Guillemot, Bihong Huang.

The High Efficiency Video Coding standard (HEVC) supports a total of 35 intra prediction modes which aim at reducing spatial redundancy by exploiting pixel correlation within a local neighborhood. However the correlation remains in the residual signals of intra prediction, leading to some high energy prediction residuals. In 2014, we have studied several methods to exploit remaining correlation in residual domain after intra prediction. These methods are based on vector quantization with codebooks learned and dedicated to the different prediction modes in order to model the directional characteristics of the residual signals. The best matching code vector is found in a rate-distortion optimization sense. Finally, the index of the best matching code vector is sent to the decoder and the vector quantization error, the difference between the intra residual vector and the best matching code vector, is processed by the conventional operations of transform, scalar quantization and entropy coding.

In a first approach called MDVQ (Mode Dependent Vector Quantization), the codebooks were learned using the k-means algorithm [26]. More recently, we have developed a variant of the approach, called AMDVQ (Adaptive MDVQ) by adding a codebook update step based on Kohonen Self-Organized Maps which aims at capturing the variations of the residual signal statistical charateristics. The Kohonen algorithm uses previously reconstructed residual vectors to continuously update the code vectors during the encoding and decoding of the video sequence [12].

7.3.3. 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 Envivio, 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 close to some "aesthetic" a priori. Taking into account the aesthetic of the scene in video compression is novel, since video compression is traditionally optimized to deliver the smallest distortion with the input data at the minimum datarate.

7.3.4. *Local Inverse Tone Curve Learning for HDR Image Scalable Compression*

Participants: Christine Guillemot, Mikael Le Pendu.

In collaboration with Technicolor, we have developed local inverse tone mapping operators for scalable high dynamic range (HDR) image coding. The base layer is a low dynamic range (LDR) version of the image that may have been generated by an arbitrary Tone Mapping Operator (TMO). No restriction is imposed on the TMO, which can be either global or local, so as to fully respect the artistic intent of the producer. The method which has been developed successfully handles the case of complex local TMOs thanks to a block-wise and non-linear approach [28]. A novel template based Inter Layer Prediction (ILP) is designed in order to perform the inverse tone mapping of a block without the need to transmit any additional parameter to the decoder. This method enables the use of a more accurate inverse tone mapping model than the simple linear regression commonly used for blockwise ILP [21]. In addition, this paper shows that a linear adjustment of the initially predicted block can further improve the overall coding performance by using an efficient encoding scheme of the scaling parameters. Our experiments have shown an average bitrate saving of 47% on the HDR enhancement layer, compared to previous local ILP methods.

7.3.5. *HEVC-based UHD video coding optimization*

Participants: Nicolas Dhollande, Christine Guillemot, Olivier Le Meur.

The HEVC (High Efficiency Video Coding) standard brings the necessary quality versus rate performance for efficient transmission of Ultra High Definition formats (UHD). However, one of the remaining barriers to its adoption for UHD content is the high encoding complexity. We address the reduction of HEVC encoding complexity by investigating different strategies: First we have proposed to infer UHD coding modes and quad-tree from a first encoding pass which consists in encoding a lower resolution version of the input video. In the context of our study, the first encoding pass encodes a HD video sequence. A speed-up by a factor of 3 is achieved compared to directly encoding the UHD format without compromising the final video quality. The second strategy focuses on the block partitioning of intra frame coding. The Coding Tree Unit (CTU) is the root of the coding tree and can be recursively split into four square Coding Unit (CU), given that the smallest block size is 8×8. Once the partitioning procedure is fully completed, the final quad-tree can be obtained by choosing the configuration leading to the best rate-distortion trade-off. Rather than performing an exhaustive partitioning, we aim to predict the quad-tree partition into coding units (CU). This prediction is based on low-level visual features extracted from the video sequences. The low-level features are related to gradient-based statistics, structure tensors statistics or entropy etc. From these features, we trained a probabilistic model on a set of UHD training sequences in order to determine whether the coding unit should be further split or not. The proposed methods yield a significant encoder speed-up ratio (up to 5.3 times faster) with a moderate loss in terms of compression efficiency [33].

7.4. Distributed processing and robust communication

Information theory, stochastic modelling, robust detection, maximum likelihood estimation, generalized likelihood ratio test, error and erasure resilient coding and decoding, multiple description coding, Slepian-Wolf coding, Wyner-Ziv coding, information theory, MAC channels

7.4.1. *Information theoretical bounds of Free-viewpoint TV*

Participants: Thomas Maugey, Aline Roumy.

Free-viewpoint television FTV is a new system for viewing video where the user can choose its viewpoint freely and change it at anytime. The goal is to propose an immersive sensation without the disadvantage of Three-dimensional (3D) television (TV). Indeed, the conventional 3D displays (with or without glasses) occur, by construction, an accommodation-vergence conflict: since the eye tend to focus on the display screen (accommodation), whereas the brain perceives the depth of 3D images due to the different views seen by each eye (vergence). Instead, with FTV, a look-around effect is produced without any visual fatigue since the displayed images remain 2D. Therefore, FTV presents nice properties that makes it a serious competitor for 3DTV. Existing compression algorithms for FTV consider to send all the views, which would require about 100 Mbits/s (for 100 views, as needed to propose a true navigation within the scene). Since this amount does not fit the current datarate for transmission in a streaming scenario, we investigate a solution where the server only send the request. In [31], [30], we have shown a very surprising and positive result: if all the views are compressed once and if the server extracts from the compressed bitstream the request (i.e. one view at a time), the datarate is exactly the same as if the whole database was entirely decoded, and the requested views reencoded. This very positive result shows that it is possible to send FTV with the same datarate as single view television with very limited computational cost at the server (only extraction from the bistream). This result is an information theoretical result and the goal is now to build a practical system that can achieve this performance.

7.4.2. Compressed Sensing : a probabilistic analysis of the orthogonal matching pursuit algorithm

Participant: Aline Roumy.

Compressed sensing (CS) is an efficient acquisition scheme, where the data are projected onto a randomly chosen subspace to achieve data dimensionality reduction. The projected data are called measurements. The reconstruction is performed from these measurements, by solving underdetermined linear systems under a sparsity a priori constraint. It is generally believed that the greedy algorithm Orthogonal Matching pursuit performs well and can determine which variables are active (i.e. non zero). In contrast, we showed that this is not the case even in the noiseless context. We derived an exact probabilistic analysis of the iterative algorithm in the large system regime, when all dimensions tend to infinity. We showed that as the number of iterations grows, the algorithm will make errors with probability one.

STARS Project-Team

7. New Results

7.1. Introduction

This year Stars has proposed new results related to its three main research axes : perception for activity recognition, semantic activity recognition and software engineering for activity recognition.

7.1.1. *Perception for Activity Recognition*

Participants: Julien Badie, Slawomir Bak, Piotr Bilinski, François Brémond, Duc Phu Chau, Etienne Corvée, Antitza Dancheva, Kanishka Nithin Dhandapani, Carolina Garate, Furqan Muhammad Khan, Michal Koperski, Thi Lan Anh Nguyen, Javier Ortiz, Ujjwal Ujjval.

The new results for perception for activity recognition are:

- Pedestrian Detection using Convolutional Neural Networks (see 7.2)
- Head detection for eye tracking application (see 7.3)
- Minimizing hallucination in Histogram of Oriented Gradients (see 7.4)
- Hybrid approaches for Gender estimation (see 7.5)
- Automated Healthcare: Facial-expression-analysis for Alzheimer's patients in musical mnemotherapy (see 7.6)
- Robust Global Tracker based on an Online Estimation of Tracklet Descriptor Reliability (see 7.7)
- Optimizing people tracking for a video-camera network (see 7.8)
- Multi-camera Multi-object Tracking and Trajectory Fusion (see 7.9)
- Person Re-Identification in Real-World Surveillance Systems (see 7.10)
- Human Action Recognition in Videos (see 7.11)

7.1.2. *Semantic Activity Recognition*

Participants: Vasanth Bathrinarayanan, François Brémond, Duc Phu Chau, Serhan Cosar, Alvaro Gomez Uria Covella, Carlos Fernando Crispim Junior, Ramiro Leandro Diaz, Giuseppe Donatielo, Baptiste Fosty, Carolina Garate, Alexandra Koenig, Michal Koperski, Farhood Negin, Thanh Hung Nguyen, Min Kue Phan Tran, Philippe Robert.

For this research axis, the contributions are :

- Evaluation of Event Recognition without using Ground Truth (see 7.12)
- Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition (see 7.13)
- Semi-supervised activity recognition using high-order temporal-composite patterns of visual concepts (see 7.14)
- From activity recognition to the assessment of seniors' autonomy (see 7.15)
- Serious Games Interfaces using an RGB-D camera (see 7.16)
- Assistance for Older Adults in Serious Game using an Interactive System (see 7.17)
- Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition (see 7.18)

7.1.3. *Software Engineering for Activity Recognition*

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Ines Sarray, Imane Khalis, Nazli Temur, Daniel Gaffé, Rachid Guerchouche, Matias Marin, Etienne Corvée, Carolina Da Silva Gomes Crispim, Anais Ducoffe, Jean Yves Tigli, François Brémond.

The contributions for this research axis are:

- Run-time Adaptation of Video Systems (see 7.19)
- Scenario Description Language (see 7.20)
- Scenario Recognition (see 7.21)
- The Clem Workflow (see 7.22)
- Safe Composition in WComp Middleware for Internet of Things (see 7.23)
- Design of UHD panoramic video camera (see 7.24)
- Brick & Mortar Cookies (see 7.25)
- Monitoring Older People Experiments (see 7.26)

7.2. Pedestrian Detection using Convolutional Neural Networks

Participants: Ujjwal Ujjwal, François Brémond.

Keywords: Pedestrian detection, CNN

The objective of the work was to perform pedestrian detection in different settings. The settings corresponded to different types of camera-views as well as different types of camera settings (e.g- moving camera vs. static camera). The work followed a wide range of experiments using different public implementations of convolutional neural networks and on different types of datasets. We detail the experiments one by one in the following subsections :

Experiments on CNN architectures

We started with an evaluation of different CNN architectures for pedestrian detection. Towards this end, we implemented three important and famous architectures - LeNet [72], AlexNet [69] and CifarNet [68]. For the purpose of training and validation we extracted patches from the public datasets of Inria [55], Daimler [58], TUD-Brussels [92], Caltech [57], ViPer [62], USC [93] and MIT [78]. The breakup of the dataset used for training was as shown in table 1 .

Table 1. Training, testing and validation patch details for CNN training

	Pedestrian Patches	Non-Pedestrian Patches
Training	131,183	61,500
Validation	65,591	30,700
Testing	65,591	30,600

Implementation of all the three models for pedestrian detection which gave a very high accuracy (94.2% (LeNet), 98% (AlexNet) and 98.2% (CifarNet)) for classification at patch level. Though these results were good at the patch level, more thorough understanding was needed to determine the effect of network architecture on classification. This was important because the three architectures vary greatly in terms of number of layers and other parameters. Moreover the practical problem in pedestrian detection chiefly involves detecting pedestrians in an image (i.e. when full-scale images instead of pre-defined patches are available).

The first set of experiments was done using sliding windows. This had to be abandoned soon, since for each image this was taking an impractical time ($> 3minutes/image$). Moreover sliding window is less suited in its naive setting due to the fact that each candidate window had to be rescaled to meet the network input size and tested individually by extracting features over it. This was followed by efforts to understand and implement a wide range of other techniques for full-scale detection using CNN. This is still an open problem though some encouraging advancements through R-CNN [61] and OverFeat [83] have been made. A major difficulty lies in lack of robust implementations of CNN which allow for integrated training and testing with object localization. Moreover existing implementations are less flexible and often make it difficult to carry out modifications required to implement new techniques independently.

We settled with the R-CNN which uses region proposals extracted using selective search [87] to extract object proposals and then train a CNN using those proposals and subsequently classifying using a SoftMax classifier or a SVM.

The evaluation was done on both moving cameras and static cameras and the evaluation showed that the network was performing a little satisfactorily, though below the state-of-art performance standards. The performance metric was Average Miss Rate (AMR) Vs. False Positives Per Image (FP/I). A good detector must exhibit a very low AMR alongwith very low FP/I. Table 2 summarizes the detection results, with table 3 summarizing state-of-art results on different pedestrian detection datasets.

Table 2. Our R-CNN results on different pedestrian datasets.

Dataset	#Images	AMR	FP/I
Inria	741	0.27	0.36
DAIMLER	15K	0.38	0.40
Caltech	16K	0.46	0.43
USC	584	0.02	0
PETS 2009 S2.L1	5565	0.42	0.29
PETS 2009 S2.L2	1744	0.35	0.19

Table 3. State-of-art results on different pedestrian datasets.

Dataset	#Images	AMR	FP/I
Inria	741	0.14	0.1
DAIMLER	15K	0.29	0.21
Caltech	16K	0.12	0.1
PETS 2009 S2.L1	5565	0.22	0.1
PETS 2009 S2.L2	1744		

It was felt that more work is needed to organize a CNN library and subsequently work to improve the above results.

Conclusion

Pedestrian detection finds its applications in different settings. It is also highly influenced by a wide variety of variations which have many practical ramifications in areas such as surveillance. It is important to develop a robust and high-performance system for pedestrian detection that is able to take into account a very wide range of such variations such as occlusion and poor visibility. CNNs have shown great promise in object detection and recognition lately and this inspires it growing applications in pedestrian detection. While the current results of our R-CNN experiments do not match the state-of-art it has shown some promise by providing consistent numbers across datasets which shows that CNNs are a good way to transcend a system beyond dataset-specific restrictions. An important factor is the instance of moving cameras vis-a-vis static cameras. While the present experiments show that decent performance is obtained on moving camera databases, consistent and similar performance is also obtained in the context of static camera databases such as PETS. This shows that with better training and improved practices of dataset handling such as augmentation and dataset structuring by clustering based methods can help in pushing the performance to acceptable levels for applications in automated surveillance and driving applications.

Further Work We intend to take this study forward, by looking into novel approaches to gather more information about a pedestrian dataset from CNN, while further increasing the detection results.

7.3. Head Detection for Eye Tracking Application

Participants: Thanh Hung Nguyen, Antitza Dantcheva, François Brémond.

Keywords: computer vision, head detection, eye tracking

Head detection [77] uses RGBD sensor (Kinect 2 sensor) which is supported by SUP platform of STARS team. For the eye tracking, we use the open-source library (OpenBR) which has good performance in our test.

Until now, the head detector was working well when people were standing but not good enough when people were lying or bending as you can see on Figure 5 and Figure 6 . This experiment was realised on simple datasets where mostly people was close to camera and walking. The main reason is the lack of samples in the learning process for the challenges cases (lying, bending). So at this moment we are collecting the head samples for it.

Number of true positives	209
Number of false positives	28
Number of false negatives	25

Figure 5. Performance of head detection on simple dataset

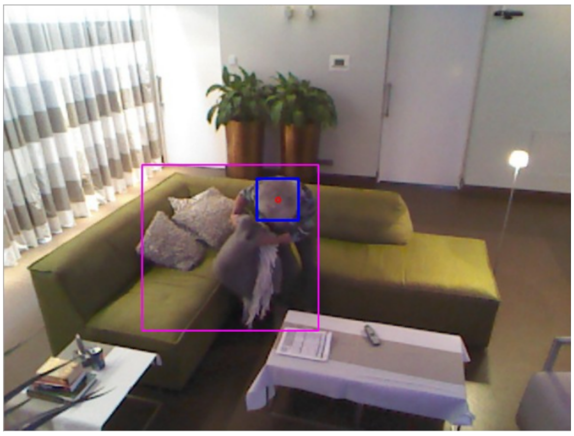


Figure 6. An example of experiment

7.4. Minimizing Hallucination in Histogram of Oriented Gradients

Participants: Javier Ortiz, Slawomir Bak, Michal Koperski, François Brémond.

keywords: computer vision, action recognition, re-identification,
Challenges in histogram representation

In computer vision very often histogram of values is used as a feature representation. For instance HOG descriptor is in fact histogram of gradient orientations. Also histogram of codewords in Bag of Words representation is a very popular action recognition representation. In such a case we are not interested in absolute values for given bin, but rather in shape of histogram to find the patterns. To make histogram representation independent from absolute values we use L1 or L2 normalization. In the normalization process we treat histogram as a vector and we transform it in the way that it should have unit length according to given norm (L1, L2). The drawback of such approach is that normalization process may amplify the noise for histograms where absolute values are very small (no pattern or only noise). Such histogram after normalization can be very similar to histogram with strong absolute values. This situation is showed in figure 7 . Although original histogram in second row of figure 7 contains almost no information after normalization the values are amplified and the result is exactly the same as for histogram in the first row. Such behavior is called hallucinations.

Proposed normalization method

We propose to add an artificial bin with given value which would prevent small noisy values from being amplified during normalization process. In figure 7 in first row we show artificial bin in pink color. Thanks to that histograms after normalization (last column) are different. If we analyze cumulative sum of histogram values across data (we sum values of all bins in whole histogram and we draw distribution), we can find that in some data-sets we obtain bi-nomial distribution see figure 8 . The gap between two Gaussians indicates convenient border between noise and meaningful data. On the other hand many data-sets do not have this bi-nomial feature and for that we do not have formalization to find the value of artificial bin. This problem would be a subject of further studies.

The following method was successfully applied to person re-identification and action recognition problem. Further details can be found in the paper [42]

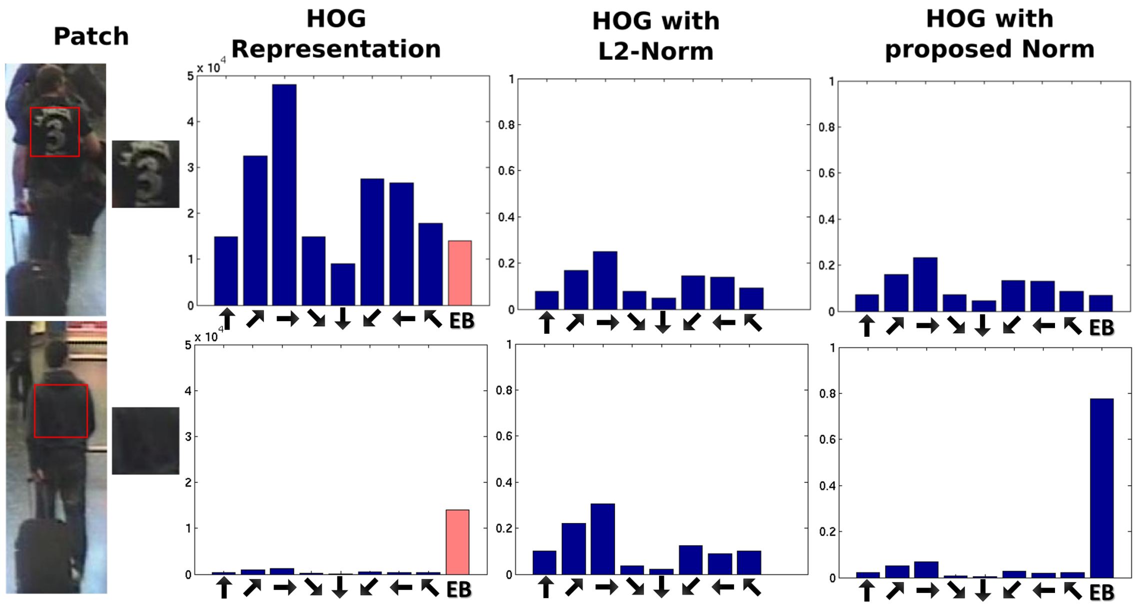


Figure 7. HOG representations of two patches with different amount of texture. Each bin in blue represents the sum of magnitudes of edges for a particular orientation. The bin marked as EB represents the proposed extra bin

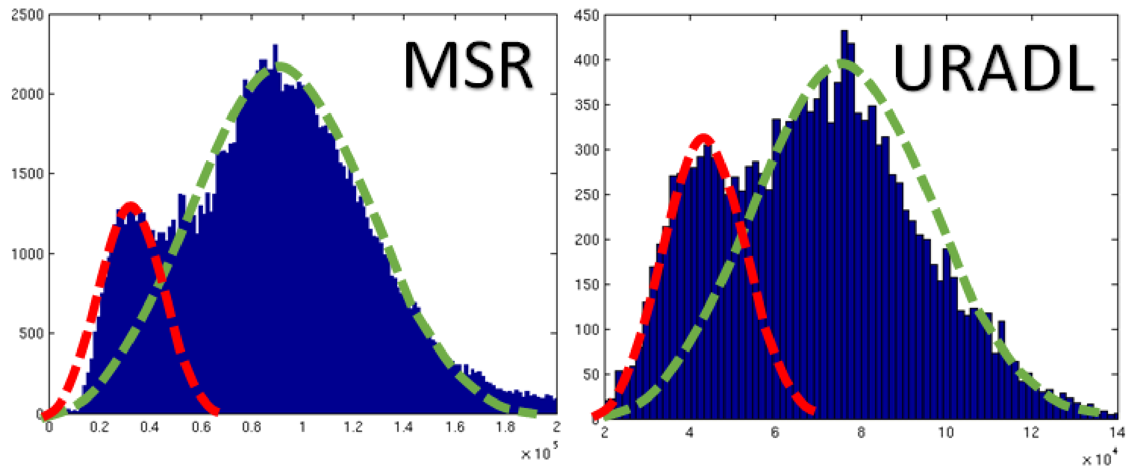


Figure 8. Distribution of the sum of bins for randomly sampled patches for URADL and MSRDailActivity3D data-set.

7.5. Hybrid Approaches for Gender estimation

Participants: Antitza Dantcheva, François Brémond, Philippe Robert.

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. Gender estimation remains a challenging task, which is inherently associated with different biometric modalities including fingerprint, face, iris, voice, body shape, gait, as well as clothing, hair, jewelry and even body temperature [31]. Recent work has sought to further the gains of single-modal approaches by combining them, resulting into hybrid cues that offer a more comprehensive gender analysis, as well as higher resilience to degradation of any of the single sources.

Can a smile reveal your gender?

In this work we propose a novel method for gender estimation, namely the use of dynamic features gleaned from smiles and show that (a) facial dynamics can be used to improve appearance-based gender-estimation, (b) that while for adults appearance features are more accurate than dynamic features, for subjects under 18 years old 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 proposed system, fusing a state-of-the-art appearance and dynamics-based features (see Figure 9), improves the appearance based algorithm from 78.0% to 80.8% in video-sequences of spontaneous smiles and from 80% to 83.1% in video-sequences of posed smiles for subjects above 18 years old (see Table 4). These results show that smile-dynamics include pertinent and complementary information to appearance gender information.

While this work studies video sequences capturing frontal faces expressing human smiles, we can envision that additional dynamics, such as other facial expressions or head and body movements carry gender information as well.

Distance-based gender prediction: What works in different surveillance scenarios?

Table 4. True gender classification rates. Age given in years.

Age	< 10	$10 - 19$	$20 - 29$	$30 - 39$	$40 - 49$	> 49
Subj. amount	48	95	60	49	72	33
OpenBR	58.33%	50.53%	81.67%	75.51%	75%	81.82%

Combined Age-Groups	< 20	> 19
Subj. amount	143	214
OpenBR	52.45%	78.04%
Dynamics (SVM+PCA)	60.1%	69.2%
Dynamics (AdaBoost)	59.4%	61.7%
OpenBR + Dynamics (Bagged Trees)	60.8%	80.8%

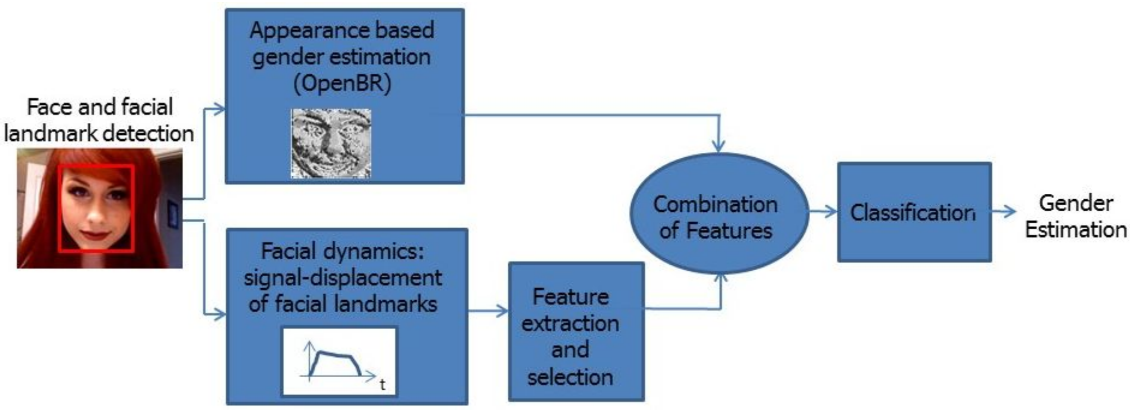


Figure 9. General Scheme of the facial appearance and dynamics framework.

In this work we fuse features extracted from face, as well as from body silhouette towards gender estimation. Specifically, for face, a set of facial features from the OpenBR library, including histograms of local binary pattern (LBP) and scale-invariant feature transform (SIFT) are computed on a dense grid of patches. Subsequently, the histograms from each patch are projected onto a subspace generated using PCA in order to obtain a feature vector, followed by a Support Vector Machine (SVM) used for the final face-based-gender decision. Body based features include geometric and color based features, extracted from body silhouettes, obtained by background subtraction, height normalization and SVM-classification for the final body-based-gender-decision. We present experiments on images extracted from video-sequences, emphasizing on three distance-based settings: close, medium and far from the TunnelDB dataset (see Figure 10). As expected, while face-based gender estimation performs best in the close-distance-scenario, body-based gender estimation performs best when the full body is visible - in the far-distance-scenario (see Table 5). A decision-level-fusion of face and body-based features channels the benefits of both approaches into a hybrid approach, providing results that demonstrate that our hybrid approach outperforms the individual modalities for the settings medium and close.

Table 5. True gender classification rates. Age given in years.

textbfSystem Distance	**FAR**	**MEDIUM**	**CLOSE**
FBGE	57.14	79.29	89.29
BBGE 89.3 85 79.3			
Fusion BBGE & FGBE	82.9	88.6	95

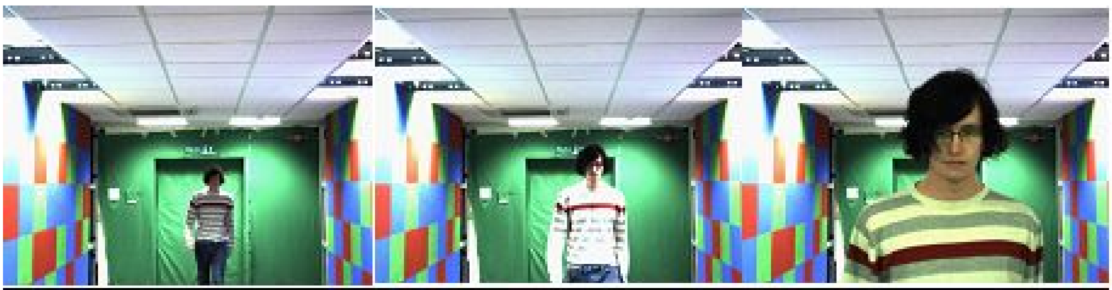


Figure 10. Three distance-based settings in the TunnelDB dataset: far, medium and close.

While the dataset is relatively unconstrained in terms of illumination, body and face are captured facing relatively frontally towards the camera. Future work will involve less constraints also towards the pose of humans.

7.6. Automated Healthcare: Facial-expression-analysis for Alzheimer's Patients in Musical Mnemotherapy

Participants: Antitza Dantcheva, François Brémond, Philippe Robert.

keywords: automated healthcare, healthcare monitoring, expression recognition

In this work we seek to apply computer vision towards increasing the life quality of patients with Alzheimer's disease (AD), and particularly in applying computer vision towards interventions to delay functional decline and to decrease the burden of the most disturbing behavioral symptoms. Towards this we design a smart interaction tool, that "reads" emotions of AD patients. This approach is becoming necessary now, because the increasing prevalence of chronic disorders and its impact on functional decline is challenging the sustainability of healthcare systems. Firstly, we have assembled a dataset of video-sequences acquired in the Alzheimer's Disease - clinique Fondation GSF Noisiez. Multiple patients and sessions have been captured during musical mnemotherapy. We then have annotated several sequences per one of four facial expressions, that occur in the recorded dataset including: neutral, talking, smile and sad. We then proceed to classify these expressions for 10 patients based on two approaches, that we study individually, as well as fused. The first approach contains face detection, facial landmark localization and signal displacement analysis for different facial landmarks, which are ranked based on categorization-pertinence, fused and classified into one of the four expression-categories (see Figure 11). In the second approach, we use face detection, eyes-detection, face normalization and HOG-features, which we classify into one of the four expression-categories.

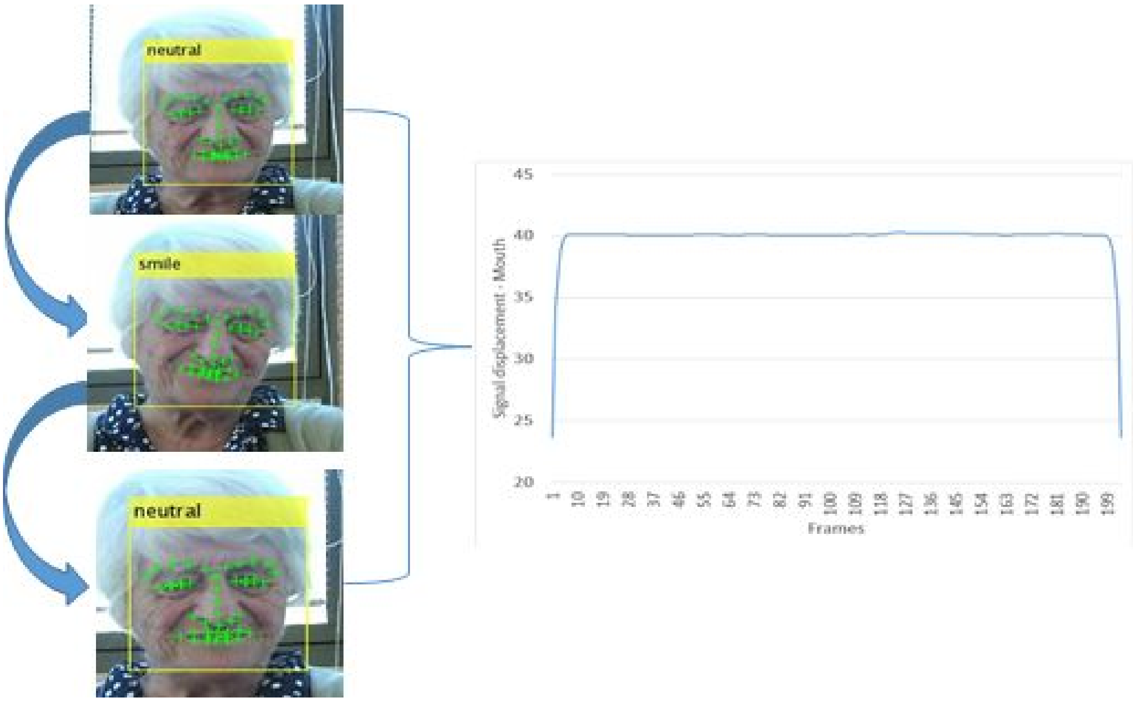


Figure 11. Smile detection based on signal displacement in the mouth region in the dataset collected at the Fondation GSF Noisiez in Biot, France.

The here used real-world-data challenges, as expected, all utilized computer vision algorithms (from face detection - due to no constraints of pose and illumination, to classifiers - due to a large intra-class-variation of facial expressions). Nevertheless, we obtain promising results that we envision improving by analyzing 2D and depth, as well as infrared data.

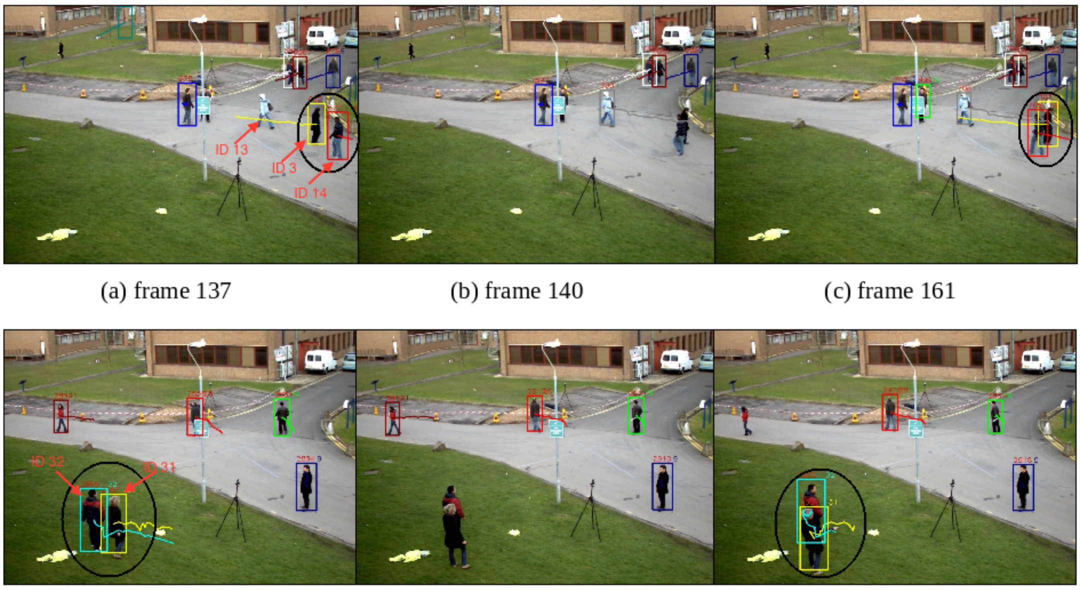
7.7. Robust Global Tracker Based on an Online Estimation of Tracklet Descriptor Reliability

Participants: Thi Lan Anh Nugyen, Chau Duc Phu, François Brémond.

Keywords: Tracklet fusion, Multi-object tracking

Object tracking - the process of locating a moving object (or multiple objects) in one camera or in a camera network over time - is an important part in surveillance video processing. However, the video context variation requires trackers to face plenty of challenges. For example, objects change their movement direction or their appearances, poses; objects are occluded by other objects or background; illumination is changed... In order to overcome above challenges, calculating the object appearance model overtime to adapt tracker to context variation is a necessary work.

In the state of the art, some online learning approaches [52], [48] have been proposed to track objects in various video scenes in each frame. These approaches learn online discriminative object descriptors to the current background as in [52] or learn an object appearance model which discriminates objects overtime as in [48]. However, the limitation of these approaches is that the reliability of object descriptor computed on only current frame is sensitive. False positives can reduce tracking quality. Furthermore, these algorithms try to find the discriminative descriptors or signatures of one object compared to its neighborhood but not considering to the correlation of this object with its can-match candidates. Meanwhile, global tracking methods [91], [98] show their dominant ability over previous methods in noisy filtering. The approach in [98] proposes an algorithm that recovers fragmentation of object trajectories by using enhanced covariance-based signatures and an online threshold learning. The approach in [91] proposes a hierarchical relation hypergraph based tracker. These global tracking algorithms have significant results in matching short trajectories and filtering some noise. However, object descriptor weights are fixed for the whole video. Therefore, their tracking performances can be reduced if the scene changes.



(a) frame 137 (b) frame 140 (c) frame 161

(d) frame 553 (e) frame 563 (f) frame 865

Figure 12. PETs2009 dataset: The online computation of discriminative descriptor weights depending on each video scene.

In this work, we propose a new approach to improve the tracking quality by a global tracker which merges all tracklets belonging to an object in the whole video. Particularly, we compute online descriptor reliability over time based on their discrimination. Based on the computed discriminative descriptor weights, the global matching score over descriptors of 2 tracklets is given. Then, we apply Hungarian algorithm to optimize

Table 6. Tracking performance.The best values are printed in bold, the second best values are printed in italic.

Dataset	Method	MOTA	MOTP	GT	MT	PT	ML	FG
PETS2015	Chau *et Al.* [53]	–	–	2	0.0	100.0	**0.0**	2
	Ours (Proposed approach + [53])	–	–	2	**100.0**	0.0	**0.0**	**1**
PETS2009	Chau *et Al.* [53]	0.62	0.63	21	–	–	–	8
	Bae *et Al.with all* [48]	0.83	0.69	23	**100**	0	**0.0**	**4**
	***Zamir* et Al.** [95]	**0.90**	0.69	21	–	–	–	–
	***Bae* et Al.**-*global association* [48]	0.73	0.69	23	**100**	0	**0.0**	12
	***Badie* et Al.** [47]	**0.90**	**0.74**	21	–	–	–	–
	***Badie* et Al.** [47] + [53]	0.85	0.71	21	66.6	23.9	9.5	*6*
	Ours (Proposed approach + [53])	*0.86*	*0.72*	21	*76.2*	14.3	9.5	**4**
TUD-Stadtmitte	Milan *et Al.* [74]	**0.71**	**0.65**	9	**70.0**	20.0	**0.0**	–
	Yan *et Al.* [94]	–	–	10	**70.0**	30.0	**0.0**	–
	Chau *et Al.* [53]	0.45	0.62	10	60.0	40.0	**0.0**	13
	Ours (Proposed approach + [53])	0.47	**0.65**	10	**70.0**	30.0	**0.0**	**7**
TUD-Crossing	Tang *et Al.* [84]	–	–	–	**53.8**	38.4	7.8	–
	Chau *et Al.* [53]	0.69	0.65	11	46.2	53.8	**0.0**	14
	Ours (Proposed approach + [53])	**0.72**	**0.67**	11	**53.8**	46.2	**0.0**	**8**

tracklet matching. On the other hand, a motion model is also combined with appearance descriptors in a flexible way to improve the tracking quality. Figure12 shows the visual explanation . In frame 137, two objects have similar appearance but move with different direction. In this case, motion descriptor is more reliable. Inversely, in frame 553, two objects go consistently together but their coat and hair's colors are different. Therefore, the appearance descriptors are more reliable than motion one.

The proposed approach gets results of tracker in [53] as input and is tested on challenge datasets. The comparable results of this tracker with other trackers from the state of the art are shown in Table 6 . This paper is accepted in PETs workshop [41].

7.8. Optimizing People Tracking for a Video-camera Network

Participants: Julien Badie, François Brémond.

Keywords: tracking quality estimation, error recovering, tracklet matching

This work addresses the problem of improving tracking quality during runtime. Most state-of-the-art tracking or high-level algorithms such as event recognition have difficulties to handle erroneous inputs. This framework detects and repairs detection or tracking errors. It works in an online situation and even in the case where prior knowledge of the scene (such as contextual information or training data) is not available.
The Global Tracker (figure 13) uses tracking results (tracklets) as input and produces corrected tracklets as output.

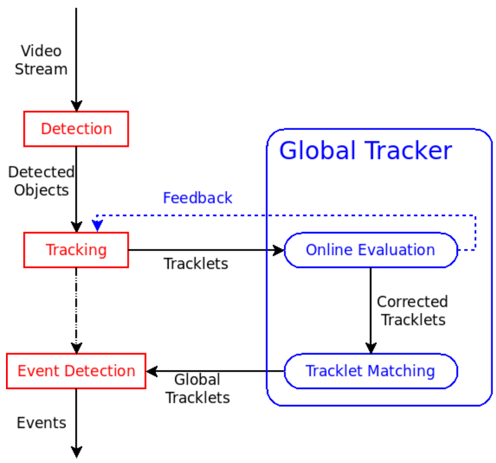


Figure 13. The Global Tracker framework, combining online evaluation and tracklet matching to improve tracking results.

The Global Tracker framework is divided into two main modules:

- **Online evaluation of tracking results**: the quality of the tracking results is computed by analyzing the variation of each tracklet feature. A significant feature variation is considered as a potential error, an anomaly. To determine if this anomaly is a real error or a natural phenomenon, we use information given by the object neighborhood and the context. Finally the errors are corrected either by removing the erroneous nodes (basic approach) or by sending a signal to the tracking algorithm in order to tune its parameters for the next frames (feedback approach).

- **Tracklets matching over time**: tracklets representing the same object are merged together in a four-step algorithm. First we select key frames (frames that are close to the mean value of the features) for each tracklet. Then a visual signature is computed based on these key frames. The distance between each pair of signature is then computed. Finally the tracklet merging is done using unsupervised learning and a constrained clustering algorithm where all tracklets representing the same object are put in the same cluster.

This approach has been tested on several datasets such as PETS2009 (table 7), CAVIAR (table 8), TUD, I-LIDS and VANAHEIM and with different kinds of scenarios (tracking associated with a controller, 3D camera, camera network with overlapping or distant cameras). In each case, we are able to reach or outperform the state-of-the-art results.

Table 7. Tracking results on sequence S2.L1.View1 of the PETS2009 dataset

Methods	MOTA	MOTP
Zamir et al. [95]	**0.90**	0.69
Milan et al. [75]	**0.90**	**0.74**
Online evaluation	0.90	0.74
Tracklet matching	0.83	0.68
Global Tracker	***0.92***	***0.76***

Table 8. Tracking results on the Caviar dataset

Methods	MT (%)	PT (%)	ML (%)
Li et al. [73]	84.6	14.0	1.4
Kuo et al. [70]	84.6	14.7	***0.7***
Online Evaluation	82.6	11.7	5.7
Tracklet matching	84.6	9.5	5.9
Global Tracker	***86.4***	***8.3***	5.3

This approach is described more in detail in the PhD manuscript [27].

7.9. Multi-camera Multi-object Tracking and Trajectory Fusion

Participants: Kanishka Nithin Dhandapani, Thi Lan Anh Nguyen, Julien Badie, François Brémond.

Keywords: Multicamera, Tracklet association, Trajectory fusion, Object Tracking.

In spite of number of solutions that exist for multi-object tracking, it is still considered most challenging and unsolved computer vision problems, mainly due to inter and intra-occlusions, inferior visibility in crowed scenes, object re-entry, abrupt movement of object, placement of cameras and other detection inaccuracies that occur in single camera. Such drawbacks in single-camera multi-target tracking can be solved to an extent by obtaining more visual information on the same scene (more cameras). Few works done in the past years are [50], [79], [65], [59]. However they have their own problems such as not real run time performance, complex optimizations, hypothesizing 3D reconstruction and data association together might lead to suboptimal solutions.

We present a multi-camera tracking approach that associates and performs late fusion of trajectories in a centralized manner from distributed cameras. We use multiple views of the same scene to recover information that might be missing in a particular view. For detection we use background subtraction followed by discriminatively Trained Part Based Models . For object tracking, we use an object appearance-based tracking algorithm introduced by Chau et al [54] that combines a large set of appearance features such as 2D size, 3D displacement, colour histogram, and dominant colour to increase the robustness of the tracker to manage occlusion cases. Each camera in the network runs the detection and tracking chain independent of each other in a distributed manner. After a batch of frames, the data from each camera is gathered to a central node by projecting the trajectories of people to the camera with the most inclusive view through a planar homography technique and then global association and fusion are performed. Unlike the temporally local (frame to frame) data-association method, global data association has ability to deal with challenges posed by noisy detections. Global association also increases the temporal stride under optimization, therefore more stable and discriminative properties of targets can be used. Trajectory similarities are calculated as heuristically weighted combination of individual features based on geometry, appearance and motion. Association is modeled as a complete K-partite graph (all pairwise relationships inside the temporal window are taken into account) K corresponds to number of cameras in network. For simplicity purpose, we use K=2. Since we use complete K-partite graph, we have an optimal solution. Whereas methods that model association as complex multivariate optimization, upon scaling, face the problem of being stuck at local minima and may provide sub-optimal solutions. Fusion is performed using adaptive weighting method. Where the weights are derived from reliability attribute of each tracker. This enables correct and consistent trajectories after fusion even if the individual trajectories have inherent noises, occlusion and false positives

Our approach is evaluated on the publicly available PETS2009 dataset. PETS2009 is a challenging dataset due to its low FPS and interobject occlusions . We choose View1,View3 and View5 in S2.L1 scenario to evaluate. The results can be seen in Figure 14

The results are encouraging and are very raw and preliminary with lot of scope and room for improvement. With more fine tuning, error rate can be improved. However too significant errors in people detection to build on top of it. Thus, we need training detector on specific datasets to improve the approach. As future work, we will study if we can improve the optimization stage with a more complex optimization using minimal graph flow would improve the results drastically.

Scenario	Method	Camera ID's	MOTA(%)	MOTP(%)	MT(%)	PT(%)	ML(%)
PETS 2009 S2.L1	Berclaz et al.[1]	1,3,5,6,8	82	56	-	-	-
	Leal-Taixe et al.[2]	1,5	76	60	-	-	-
	Leal-Taixe et al.[2]	1,5,6	71.4	53.4	-	-	-
	Murray Evans et al[4]	2 Cameras	63	55			
	Martin Hofmann et al[3]	1,5	99.4	82.9	100	0	0
	Martin Hofmann et al.[3]	1,5,7	99.4	83.0	100	0	0

Scenario	Method	Camera ID's	MOTA(%)	MOTP(%)	MT(%)	PT(%)	ML(%)
PETS2009 S2.L1	Our Approach	1,3	76.33	65.28	92.59	0.035	0.714

Figure 14. Result of our approach on PET2009 dataset.

7.10. Person Re-identification in Real-World Surveillance Systems

Participants: Furqan Mohammad Khan, François Brémond.

Keywords: re-identification, long term visual tracking, signature modeling

Cost of supervised metric learning Person re-identification problem has recently received a lot of attention and the recent focus is to use supervised model training to learn cross camera appearance transformation. In general, $O(n^2)$ models are trained in a surveillance network with n cameras, one for each camera pair. $2p$ tracks are required to train one model with p person identities. In a real-world surveillance network with non-overlapping fields of view, a person appears only in a subset of cameras (see figure 15 , courtesy of [51]). This puts the requirement of number of tracks to train all models at $O(n^2p)$, or more precisely, $n(n-1)p$. That is, to train each model with 100 people in a 10 camera network we need 9000 tracks. For supervised training, these tracks need to be given consistent identities, and worse, have their bounding boxes marked. This is a significant burden on human annotators for deployment in real-world. Further, the annotation cost has to be repaid at a significant fraction if only one new camera is added to the system (may be due to failure of an existing camera), or if the lighting changes significantly (in case of outdoor surveillance). In our opinion, this is a significant bottleneck for supervised metric learning based re-identification in real-world.

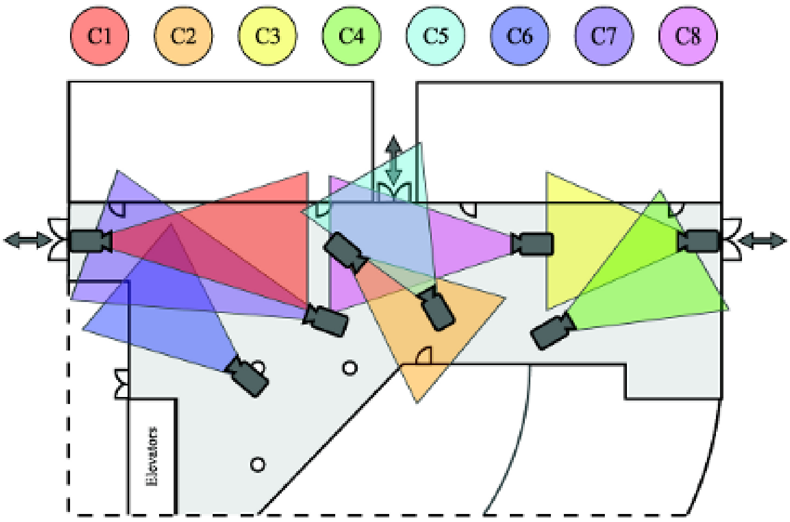


Figure 15. Camera arrangement in multi-camera surveillance scenario of SAIVT-SoftBio dataset [51]

Improved re-identification through signature modeling Re-identification is challenging because variance is intra-class appearance in often higher than inter-class appearance due to varying lighting conditions and viewpoints, and non-uniqueness of clothing. More importantly, in real-world when re-identification is fed by automated human detectors and trackers, significant mis-alignment or partial visibility of the person within proposed bounding box make it difficult to extract relevant features. Our work focuses on improving signature construction from low level features for multi-shot re-identification. We explicitly model multi-modality of person appearance using a feature mixture (corresponding publication is under review at this moment). This improves state-of-the-art re-identification performance on SAIVT-SoftBio [51] dataset and performs equally well as state-of-the-art metric learning methods on iLIDS-VID [88] and PRID2011 [64] datasets. The performance comparison of our method with state-of-the-art is presented using CMC in figure 16 (our results are denoted by MCAM).

7.11. Human Action Recognition in Videos

Participants: Piotr Bilinski, François Brémond.

Keywords: Action Recognition, Video Covariance Matrix Logarithm, VCML, Descriptor

Video Covariance Matrix Logarithm for Human Action Recognition in Videos

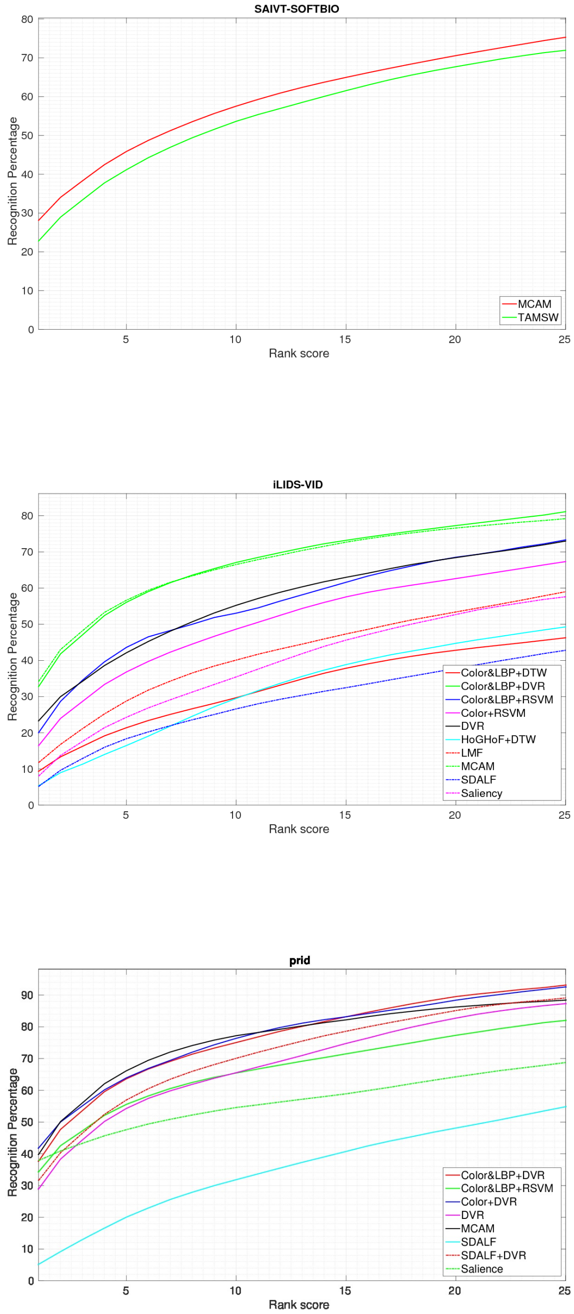


Figure 16. Performance comparison of our MCAM approach using CMC curves on different datasets. ***Top****: Comparison with TAMSW [49] on SAIVT-SoftBio dataset;* ***middle****:Comparison with Color+RSVM [88], Color&LBP+DTW [88], Color&LBP+DVR [88], Color&LBP+RSVM [88], DVR [88], HoG-HoF+DTW [88], LMF [97], Salience [96], and SDALF [60] on iLIDS-VID dataset;* ***bottom****: Comparison with Color+DVR [88], Color&LBP+DVR [88], Color&LBP+RSVM [88], DVR [88], Salience [96], and SDALF [60] on PRID2011 dataset.*

In this work, we propose a new local spatio-temporal descriptor for videos and we propose a new approach for action recognition in videos based on the introduced descriptor. Overview of the proposed action recognition approach based on the introduced descriptor is presented in Figure 17 . The new descriptor is called the Video Covariance Matrix Logarithm (VCML). The VCML descriptor is based on a covariance matrix representation, and it models relationships between different low-level features, such as intensity and gradient. We apply the VCML descriptor to encode appearance information of local spatio-temporal video volumes, which are extracted by the (Improved) Dense Trajectories. Then, we present an extensive evaluation of the proposed VCML descriptor with the (Improved) Fisher Vector encoding and the Support Vector Machines on four challenging action recognition datasets (*i.e.* URADL, MSR Daily Activity 3D, UCF50, and HMDB51 datasets). We show that the VCML descriptor achieves better results than the state-of-the-art appearance descriptors. In comparison with the most popular visual appearance descriptor, *i.e.*the HOG descriptor, the VCML achieves superior results. Moreover, we present that the VCML descriptor carries complementary information to the HOG descriptor and their fusion gives a significant improvement in action recognition accuracy (*e.g.* the VCML improves the HOG by 15% on the HMDB51 dataset). Finally, we show that the VCML descriptor improves action recognition accuracy in comparison to the state-of-the-art (Improved) Dense Trajectories, and that the proposed approach achieves superior performance to the state-of-the-art methods. The proposed VCML based technique achieves 94.7% accuracy on the URADL dataset, 85.9% on the MSR Daily Activity 3D dataset, 92.1% on the UCF50 dataset, and 58.6% on the HMDB51 dataset. More results and comparisons with the state-of-the-art are presented in Table 9 and Table 10 . To the best of our knowledge, this is the first time covariance based features are used to represent the trajectories. Moreover, this is the first time they encode the structural information and they are applied with the (Improved) Fisher Vector encoding for human action recognition in videos. This work has been published in [40].

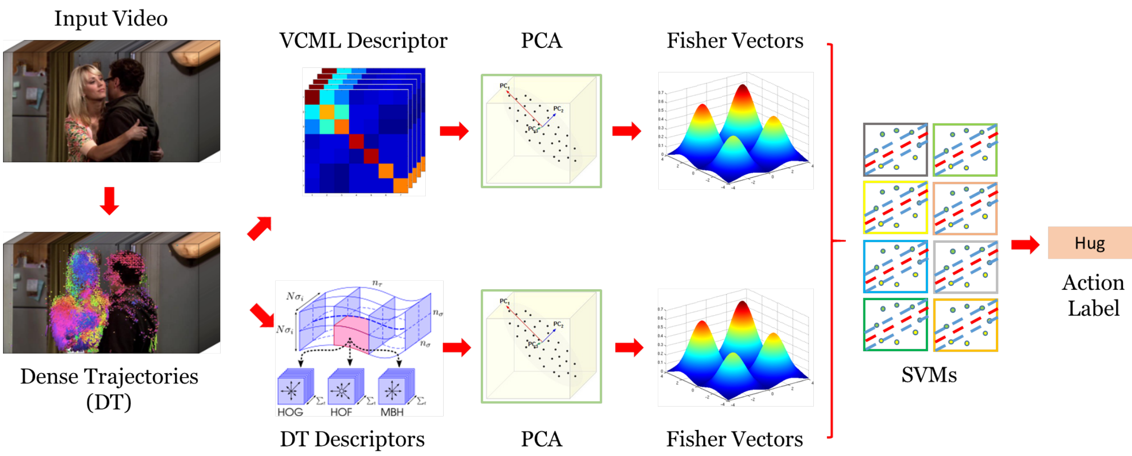


Figure 17. Overview of our action recognition approach based on the introduced VCML descriptor.

7.12. Evaluation of Event Recognition without Using Ground Truth

Participants: Ramiro Diaz, Carlos Fernando Crispim Junior, François Brémond.

Keywords: Computer Vision, Event Recognition, Video Summarization.

The main goal of the work is to improve the Event Recognition process and to improve the way we build the event models as well. The work concerns the Valrose Nursing Home, it consists in monitoring older people with health issues like Dementia and who are in need of care and stimulation.

Table 9. Comparison with the state-of-the-art on the URADL and MSR Daily Activity 3D datasets. The table presents the accuracy of our approach using Dense Trajectories (DT) and Improved Dense Trajectories (IDT).

URADL		MSR Daily Activity 3D	
Benabbas *et Al.*, 2010	81.0	Koperski *et Al.*, 2014	72.0
Raptis and Soatto, 2010	82.7	JPF – Wang *et Al.*, 2012	78.0
Messing *et Al.*, 2009	89.0	Oreifej and Liu, 2013	80.0
Bilinski and Bremond, 2012	93.3	AE – Wang *et Al.*, 2012	85.7
Dense Trajectories	94.0	Dense Trajectories	76.2
Our Approach (DT)	**94.0**	**Our Approach (DT)**	**78.1**
Our Approach (IDT)	**94.7**	**Our Approach (IDT)**	**85.9**

Table 10. Comparison with the state-of-the-art on the UCF50 and HMDB51 datasets. The table presents the accuracy of our approach using Dense Trajectories (DT) and Improved Dense Trajectories (IDT).

UCF50		HMDB51	
Kantorov and Laptev, 2014	82.2	Kantorov and Laptev, 2014	46.7
Shi *et Al.*, 2013	83.3	Jain *et Al.*, 2013	52.1
Oneata *et Al.*, 2013	90.0	Oneata *et Al.*, 2013	54.8
Wang and Schmid, 2013	91.2	Wang and Schmid, 2013	57.2
Dense Trajectories	84.2	Dense Trajectories	47.0
Our Approach (DT)	**88.1**	**Our Approach (DT)**	**52.9**
Our Approach (IDT)	**92.1**	**Our Approach (IDT)**	**58.6**

Since the video dataset contains data for about 8 months, a new evaluation method is required to properly analyze the whole dataset and gain a deeper understanding of it. Our approach consists in displaying the data in a way that can be useful either for doctors, as well for engineers to detect failures and to improve the event recognition process in an efficient way. Because of this need, a new evaluation tool has been developed and named Event Plotter.

This tool provides a new method for event evaluation. First of all, as we do not have ground truth information for the total duration of the 8 months, but just for one week, so another method is needed to check the event model efficiency. To address this issue, the tool displays all the events in the desired time period (as clusters on a timeline) and single events (or time intervals) can be selected to quickly check the video and visualize the results of the event recognition working on the fly -see Figure 18 -. The goal of this work is to understand how event recognition works, change the models on the fly, import them, and see how the recognition changes in real time.

Also, to compare new event models with the old ones, video summarization is implemented as well. Event based video summarization is utilized here to check how the recognition of one particular event type changes globally on the whole video and to display the recognition results. Also video summarization can be useful for doctors to check the way patients behave, for example playing all the videos of event ”Get-up from bed”, trying to predict patterns.

The data processed to address this issue was 1 week, because it was the time corresponding to the ground truth data. With these processed data, we have tested the efficiency of the Event Plotter tool, and we are currently improving the event recognition process by changing event models, adding new zones, and testing them on the fly.

7.13. Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition

Participants: Carlos Fernando Crispim-Junior, François Brémond.

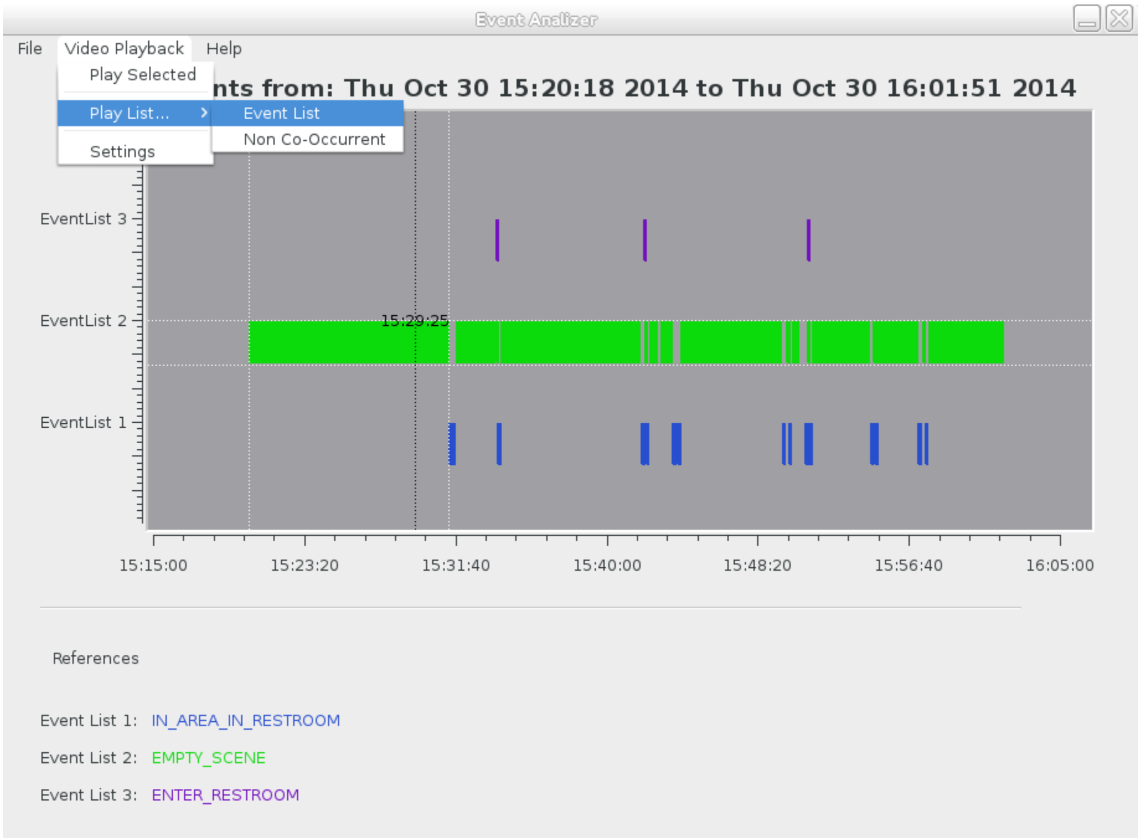


Figure 18. GUI of Event Plotter with 3 loaded Event Lists.

Keywords: Knowledge representation formalism and methods, Uncertainty and probabilistic reasoning, Concept synchronization, Activity recognition, Vision and scene understanding, Multimedia Perceptual System,

Combining multimodal concept streams from heterogeneous sensors is a problem superficially explored for activity recognition. Most studies explore simple sensors in nearly perfect conditions, where temporal synchronization is guaranteed. Sophisticated fusion schemes adopt problem-specific graphical representations of events that are generally deeply linked with their training data and focus on a single sensor. In this work we have proposed a hybrid framework between knowledge-driven and probabilistic-driven methods for event representation and recognition. It separates semantic modeling from raw sensor data by using an intermediate semantic representation, namely concepts. It introduces an algorithm for sensor alignment that uses concept similarity as a surrogate for the inaccurate temporal information of real life scenarios (Fig. 20). Finally, it proposes the combined use of an ontology language, to overcome the rigidity of previous approaches at model definition, and a probabilistic interpretation for ontological models, which equips the framework with a mechanism to handle noisy and ambiguous concept observations, an ability that most knowledge-driven methods lack (Fig. 19). We evaluate our contributions in multimodal recordings of elderly people carrying out instrumental activities of daily living (Table 11). Results demonstrated that the proposed framework outperforms baseline methods both in event recognition performance and in delimiting the temporal boundaries of event instances

This work has been developed as a collaboration between different teams in Dem@care consortium (Inria, University of Bordeaux, and CERTH). We thank the other co-authors for their contributions and support in the development of this work up to its submission for publication.

Table 11. Comparison to baseline methods in the test set

mean F_1-score	Fusion approach		
	Baselines		Ours
IADL	SVM	OSF	
S. bus line	44.19	31.36	73.10
M.finances	43.99	0.00	43.73
P. pill box	45.86	49.11	65.02
P. drink	20.02	24.29	64.03
Read	90.18	91.82	95.22
T.telephone	72.12	0.00	75.58
W. TV	2.32	0.00	35.80
W. Plant	0.00	0.00	100.00
Average	39.83	24.57	69.06

OSF: Ontology-based Semantic Fusion

7.14. Semi-supervised Activity Recognition Using High-order Temporal-composite Patterns of Visual Concepts

Participants: Carlos Fernando Crispim-Junior, Michal Koperski, Serhan Cosar, François Brémond.

Keywords: visual concepts, semi-supervised activity recognition, complex activities, cooking composite activities

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 activities - a composition of actions and scene objects - is still an open problem due to the complex temporal, composite structure of this category of events. Existing methods either still focus on simple activities or oversimplify the modeling of complex activities by only targeting whole-part relations between activity components, like actions. In this work, we have investigated a hierarchical, semi-supervised approach that unsupervisely learns actions from the composite patterns of atomic concepts (*e.g.*, slice, tomato), and complex activities from the

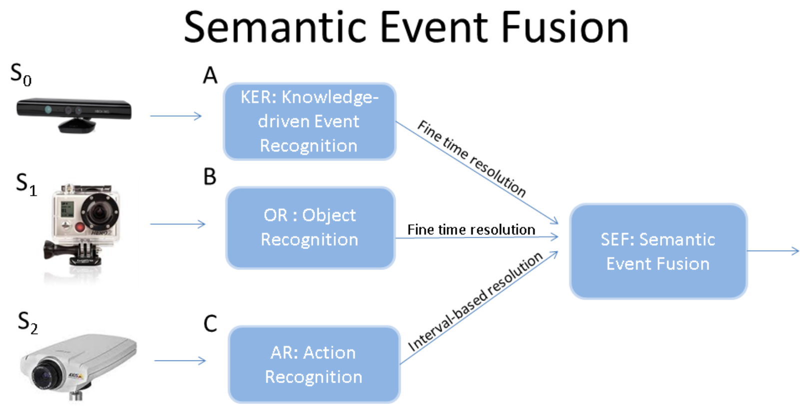


Figure 19. Semantic event fusion framework: detector modules (A-C) process data from their respective sensors (S0-S2) and output concepts (objects and low-level events). Semantic Event Fusion uses the ontological representation to initialize concepts to event models and then infer complex, composite activities. Concept fusion is performed on millisecond temporal resolution to cope with instantaneous errors of concept recognition.

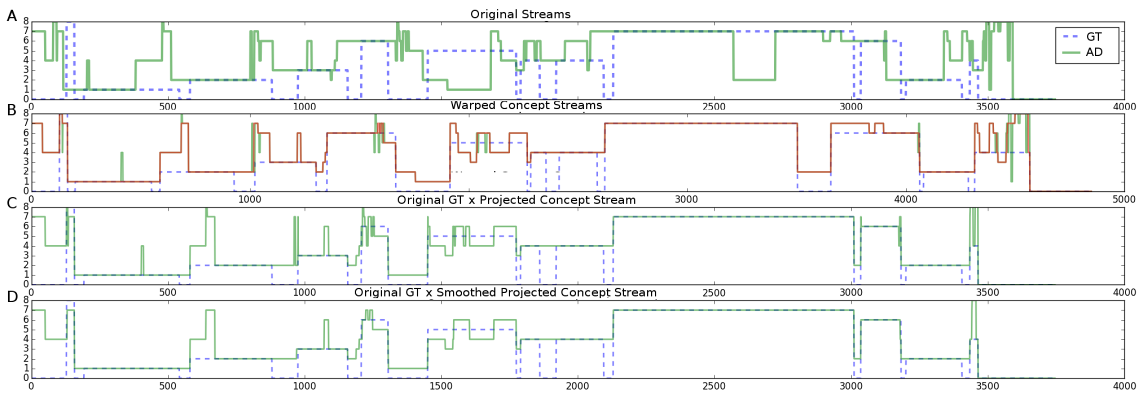


*Figure 20. Semantic alignment between the concept stream of the action recognition detector (AR) and a concept stream (GT) generated from events manually annotated by domain experts using the time axis of the color-depth camera. X-axis denotes time in frames, and Y-axis denotes activity code (1-8), respectively, search bus line on the map, establish bank account balance, prepare pill box, prepare a drink, read, talk on the telephone, watch TV, and water the plant. From top to bottom, images denote: (**A**) original GT and AR streams, (**B**) GT and AR streams warped, AR stream warped and smoothed (in red), (**C**) original GT and AR stream warped and then backprojected onto GT temporal axis, (**D**) original GT and AR warped, backprojected, and then smoothed with median filtering.*

temporal patterns of concept compositions (actions). On a first step, an unsupervised, inductive approach iteratively builds a multi-scale, temporal-composite model of the concept occurrences during the activity taking place (Fig. 22). Then, activity recognition is performed by comparing the similarity of the generated model of a given video and *a priori* learned and labeled unsupervised models. We have evaluated the proposed method in the MPII Cooking Composite Activities dataset (Fig. 21), a video collection where people perform a set of complex activities related to cooking recipes. To tackle this dataset it is necessary to recognize a large variety of visual concepts (*e.g.*, from actions, such as cutting and stirring, to objects, such as tomato and cutting board). Moreover, the detection of cooking activities is a very challenging problem since we observe a low inter-class variance between activity classes, and a high intra-class variance within an activity due to person to person differences in performing them. The proposed approach presents a mean average precision (mAP) of 56.36% ± 5.1%, and then outperforms previous methods ([81], mean AP 53.90%). This improvement is devoted to the modeling of deeper composite and temporal relations between visual concepts (from 2nd to 5th order compositions). The performance of the proposed method is mostly limited by the performance of low-level concept detectors. Future work will investigate ways to extend the current probabilistic model to handle more efficiently the differences in concept detector performance.

Figure 21. Illustation of one of the cooking recipes of Cooking Composite dataset [81]

7.15. From Activity Recognition to the Assessment of Seniors' Autonomy

Participants: Carlos Fernando Crispim-Junior, Carola Strumia, Alvaro Gomez Uria Covella, Alexandra Konig, François Brémond.

Activity recognition plays a fundamental role in several research fields as a way to extract semantic meaning from images and videos, to find more accurate matches for textual queries in video search engines, and to analyze long-term activity patterns in assisted living scenarios, such as seniors living at home. In this sense, we have continued our work on activity monitoring by proposing a novel knowledge-based event monitoring system that combines the observations of a vision system with expert knowledge and scene semantics, to recognize daily living activities in assisted living scenarios.

The approach's novelty lies in the combination of a flexible constraint-based ontology language for event modeling with efficient and robust algorithms to detect, track and re-identify people using color-depth sensing (low-level vision). The robust low-level vision promotes the modeling of longer and more complex events,

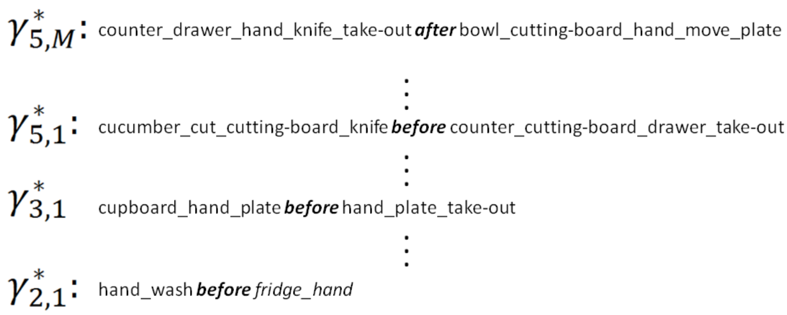


Figure 22. Example of temporal concept composite patterns of different arities extracted from an unlabeled activity.

while the ontology language provides a flexible way to describe event and incorporate domain knowledge, and ease knowledge transfer across different scenes. The proposed approach has been investigated for two assisted living scenarios: a) the monitoring of physical tasks and daily living activities in observation rooms of hospital and clinics, and b) daily and nightly activities of seniors living in nursing home apartments. To evaluate our approach performance compared to state of art methods, we have computed its results for GAADRD dataset. This is public dataset, which is composed of videos of seniors performing physical tasks and activities of daily living. Evaluation results (Table 12) have demonstrated that our approach achieves an average $F_1 - score$ 20 % higher than the baseline method [89].

Table 12. Recognition of IADLs - GAADRD data set - F 1 -score

Event	DT-HOG	DT-HOF	DT-MBHy	Proposed
Account Balance	44.96	34.71	42.98	66.67
Prepare Drink	81.66	44.87	52.00	100.00
Prepare Drug Box	14.19	0.00	0.00	57.14
Read Article	52.10	42.86	33.91	63.64
Talk on telephone	82.35	0.00	33.76	100.00
Turn on radio	85.71	42.52	58.16	94.74
Water Plant	0.00	0.00	0.00	52.63
Average $\pm$ SD	51.8 $\pm$ 34.4	23.6 $\pm$ 22.3	31.5 $\pm$ 23.3	76.4 $\pm$ 21.0

Given the satisfactory performance of the proposed activity recognition framework we have also investigated it as a method to automatically measure a seniors' autonomy in quantitative and objective fashion. To do so, we have developed a probabilistic model that takes as input the recognized activities and gait-patterns from the period of time the person performs physical tasks. The proposed autonomy model has presented an average performance of 83.67 %, which suggests that the use of such technologies may provide clinicians with diagnostic relevant information, and decrease observer's biases when compared to clinical scales. The results of this investigation have been published in [33].

7.16. Serious Games Interfaces Using an RGB-D Camera : Results and Perspectives

Participants: Baptiste Fosty, François Brémond.

Keywords: RGB-D camera analysis, walking speed, serious games, startup project

Within the context of the development of serious games for people suffering from Alzheimer disease (Az@Game project), we have developed algorithms to interact with the virtual environment through simple gesture recognition and walking speed computation. We have shown in previous work that the walking speed measured by our system is accurate enough within this context and reproducible. A paper has been submitted in Gait and Posture journal (now in reviewing process).
Concerning the gesture recognition algorithm, it consists in recognizing three basic gestures (right arm left on the side, left arm left on the side, right or left arm left on top). We performed a small experimentation to test the robustness of the system in detecting these gestures where participants (10 in total) had to perform 10 times each gesture while walking at 2.5km/h on the treadmill (see Figure 23). The results are shown in Table 13 .

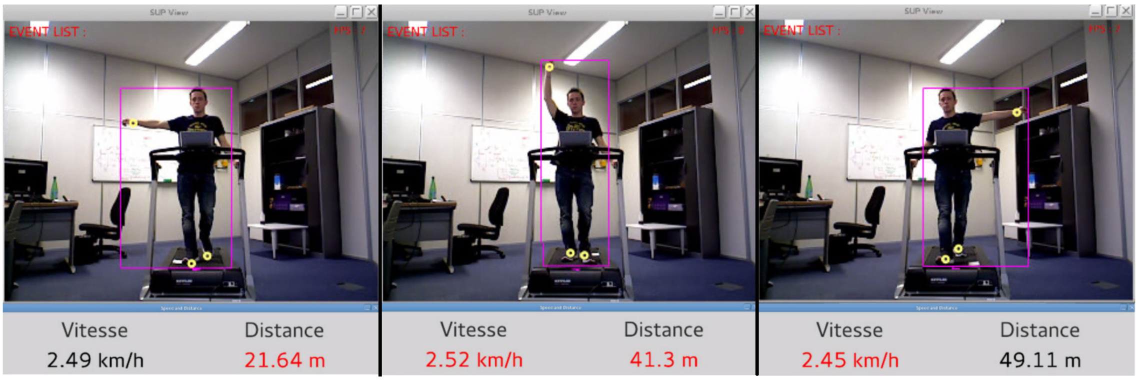


Figure 23. Display of gesture recognition and speed computation. Distance is red when right arm is left, speed is red when left arm is left, both are red when one arm is on top. Yellow dots correspond to feet and arm detection.

Table 13. Gesture recognition results.

	Right	Left	Top	TOTAL
Recall	97.0	95.0	100.0	97.3
Precision	100.0	97.9	94.3	97.3
F-Score	**98.5**	**96.4**	**97.1**	**97.3**

Following that, we decided to study whether this system would be useful in rehabilitation. Some experts of this domain in a rehabilitation center (Centre Hélio Marin in Vallauris, France) have been interviewed and they were very enthusiasts about using this type of system to get objective gait parameters. To go further on the market opportunity and evaluate the feasability of a technology transfer, we studied the concurrent products and contacted more than thirty other rehabilitation centers to have a deeper understanding of the needs and validate our idea. This investigation lead to the proposition of a startup project(BOMOTECH) to Inria which has been accepted and funded for the next 7 months, during what the goal is to get closer from a marketable tool.

7.17. Assistance for Older Adults in Serious Game Using an Interactive System

Participants: Minh Khue Phan Tran, François Brémond, Philippe Robert.

Keywords: Older Adults, Assistance, Serious Games

Serious Games offer a new way to older adults to improve their abilities such as vision, balance or memory. However, cognitive impairment causes a lot of difficulties to them when actively practising these games. Their engagement and motivation are reduced rapidly when encountering successive problems without any help. Our hypothesis is that this problematic situation can be handled if they are assisted regularly. We propose then an interactive system which can determine dynamically the situations and provide different helps in real-time. We focuse on two main problems that the older players encounter regularly :

- they forget how to continue to play the game.
- they make a lot of errors.

The system determines the above problems by computing various characteristics of the player (skeleton, postures, gestures ...) along the game states. This process is presented in Figure 24 . The characteristics of the player, which are collected by the Recognition Module thanks to Kinect Camera and the related SDK, are sent to the Interaction Module. This module associates these data with the game states provided by the game in order to recognize the problem and interacts with the player through a 3D-animated avatar.

The system is tested with 3 groups of patients described by 3 different cognitive states : mnesic plaints, MCI and Alzheimer. The patients are invited to play a concentration-based game with a Kinect camera. Each patient plays 3 phases : playing with therapist, playing alone and playing with the avatar. The playing time and the final score of each game phase are recorded. Here, the system takes into account the player's gestures and the game states for recognizing two situations :

- the player reacts too late and too slowly to the current game task.
- the player makes many mistakes.

The experimental results confirmed our hypothesis. Most of the patients have the best performance in phase "playing with the avatar". Their playing time is shorter and their final score is higher in phase "playing with assistance" than in phase "playing alone". The results are presented in the publication [36] accepted by the Games and Learning Alliance Conference in December 2015. Future work aims at improving the system and compare its efficiency with the one of "humans assistances".

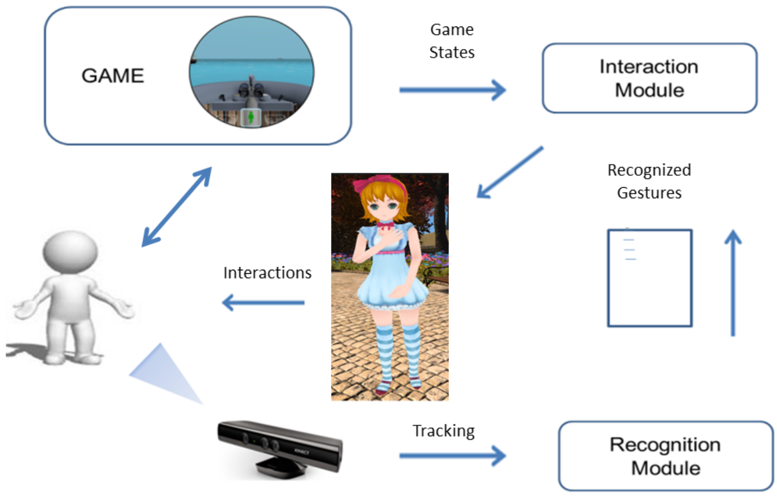


Figure 24. Assisting older adults in serious game playing

7.18. Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition

Participants: Farhood Negin, Serhan Coşar, Michal Koperski, François Brémond.

Keywords: Unsupervised Activity Recognition

Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition

In this work, we propose an unsupervised approach that offers a comprehensive representation of activities by modeling both global and body motion of people. Compared to existing supervised approaches, our approach automatically learns and recognizes activities in videos without user interaction. First, the system learns important regions in the scene by clustering trajectory points. Then, a sequence of primitive events is constructed by checking whether people are inside a region or moving between regions. This enables to represent the global movement of people and automatically split the video into clips. After that, using action descriptors [90], we represent the actions occurring inside each region. Combining action descriptors with global motion statistics of primitive events, such as time duration, an activity model that represents both global and local action information is constructed. Since the video is automatically clipped , our approach performs online recognition of activities. The contributions of this work are twofolds: (i) generating unsupervised human activity models that obtains a comprehensive representation by combining global and body motion information, (ii) recognizing activities online and without requiring user interaction. Experimental results show that our approach increases the level of accuracy compared to existing approaches. Figure 25 illustrates the flow of the system.

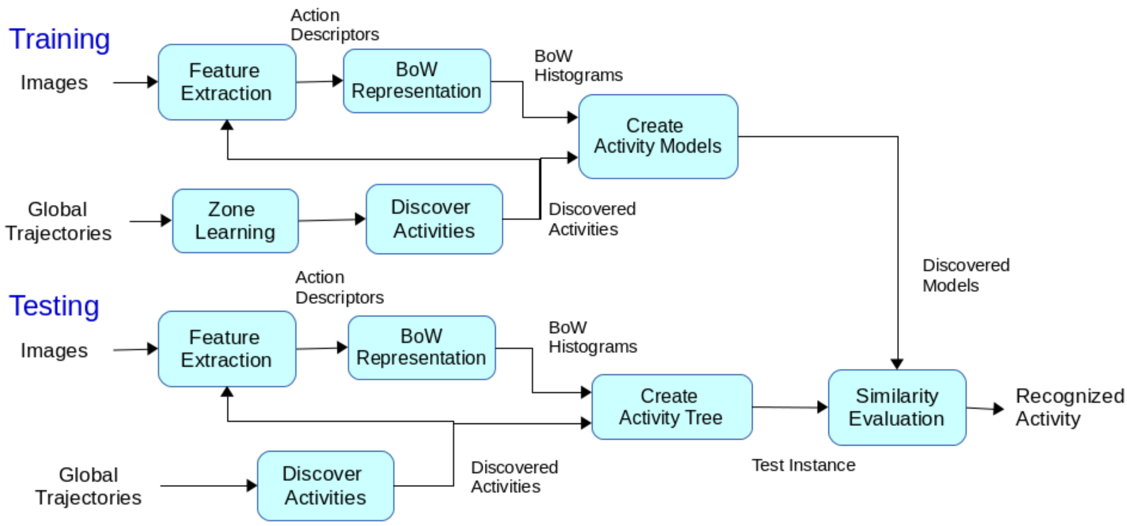


Figure 25. Architecture of the framework: Training and Testing phases

The performance of the proposed approach has been tested on the public GAADRD dataset [67] and CHU dataset (http://www.demcare.eu/results/datasets) that are recorded under EU FP7 Dem@Care Project1 in a clinic in Thessaloniki, Greece and in Nice, France, respectively. The datasets contain people performing everyday activities in a hospital room. The activities considered in the datasets are listed in Table 1 and Table 2. Each person is recorded using RGBD camera of 640x480 pixels of resolution. The GAADRD dataset contains 25 videos and the CHU dataset contains 27 videos. Each video lasts approximately 10-15 minutes.

We have compared our approach with the results of the supervised approach in [90]. We did also a comparison with an online supervised approach that follows [90]. For doing this, we train the classifier on clipped videos and perform the testing using sliding window. In the online approach, a SVM is trained using the action descriptors extracted from groundtruth intervals. We have also tested different versions of our approach that i) only uses global motion features and ii) which only uses body motion features. We have randomly selected 3/5 of the videos in both datasets for learning the activity models. The codebook size is set to 4000 visual words for all the methods.

The performance of the online supervised approach and our approach in GAADRD dataset are presented in Table 1. In all approaches that use body motion features, HoG descriptors are selected since they give the best results. It can be clearly seen that, using models that represent both global and body motion features, our unsupervised approach enables to obtain high sensitivity and precision rates. Compared to the online version of [90], thanks to the learned zones from positions and discovered activities, we obtain better activity localization, thereby better precision. However, since the online version of [90] utilizes only dense trajectories (not global motion), it fails to localize activities. Hence, it detects the intervals that does not include an activity (e.g. walking from radio desk to phone desk) and for ”prepare drug box“, ”watering plant“, and ”reading“ activities, it cannot detect the correct intervals of the activities. Compared to the unsupervised approach that either use global motion features or body motion features, we can see that, by combining both features, our approach achieves more discriminative and precise models, thereby improves both sensitivity and precision rates. By combining global and body motion features, our approach benefits from discriminative properties of both feature types. Table 1 also presents the results of the supervised approach in [90]. Although the supervised approach uses groundtruth intervals in test videos in an offline recognition scheme, it fails to achieve accurate recognition. As our approach learns the zones of activities, we discover the places where the activities occur, thereby we achieve precise and accurate recognition results. Since this information is missing in the supervised approach, it detects ”turning on radio“ while the person is inside drink zone preparing drink.

Table 2 shows the results of the online supervised approach and our approach in CHU dataset. MBH descriptor along y axis and HoG descriptor gives the best results for our approach and the online supervised approach, respectively. In this dataset, since people tend to perform activities in different places (e.g. preparing drink at phone desk), it is not easy to obtain high precision rates. However, compared to the online version of [90], our approach detects all activities and achieves a much better precision rate. The online version of [90] again fails to detect activities accurately, thereby misses some of the ”preparing drink“ and ”reading“ activities and gives many false positives for all activities.

Thanks to the activity models learned in unsupervised way, we accurately perform online recognition. In addition, the zones learned in an unsupervised way help to model activities accurately, thereby most of the times our approach achieves more accurate recognition compared to supervised approaches. This paper has been published in third Asian Conference on Pattern Recognition (ACPR 2015) [35].

7.19. Run-time Adaptation of Video Systems

Participants: Sabine Moisan, Jean-Paul Rigault, François Brémond.

In the framework of our research on model engineering techniques for video-surveillance systems, we have focused this year on run-time reconfiguration of such systems. The goal is to follow the “model at run-time” approach and to obtain context-aware self-adaptive video systems. In this approach models are kept and used at run-time. In our case, these models describe all the possible run-time configurations. They are specified using *Feature Models*.

Run-time reconfiguration means to react to context changes by tuning, adding, removing, or replacing components of the video chain, and possibly changing the chain itself.

So far, we have defined a run-time architecture consisting of three layers. The lower level describes the video analysis components and the context events. The upper one handles feature model adaptation. The middle layer is an adapter: it transforms lower level context event occurrences into upper level feature reconfiguration; in the other direction, it transforms the corresponding feature reconfigurations into video components reconfigurations.

Table 14. The activity recognition results for GAADRD dataset. Bold values represent the best sensitivity and precision results for each class.

	Supervised Approach [90]		Online Version of [90]		Unsupervised (Only Global Motion)		Unsupervised (Only Body Motion)		Proposed Approach	
ADLs	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)
Answering Phone	**100**	88	**100**	70	**100**	**100**	57	**100**	**100**	**100**
Establish Acc. Bal.	67	**100**	**100**	29	**100**	86	50	**100**	**100**	86.67
Preparing Drink	**100**	69	**100**	69	78	**100**	**100**	**100**	**100**	**100**
Prepare Drug Box	**58.33**	**100**	11	20	33.34	**100**	33.34	**100**	33.34	**100**
Watering Plant	**54.54**	**100**	0	0	44.45	57	33	**100**	44.45	**100**
Reading	**100**	**100**	88	37	**100**	**100**	38	**100**	**100**	**100**
Turn On Radio	60	86	**100**	75	89	89	44	**100**	89	**100**
AVERAGE	77.12	91.85	71.29	42.86	77.71	90.29	50.71	**100**	**80.97**	98.10

Table 15. The activity recognition results for CHU dataset. Bold values represent the best sensitivity and precision results for each class.

	Supervised Approach [90]		Online Version of [90]		Proposed Approach	
ADLs	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)
Answering Phone	57	78	**100**	**86**	**100**	65
Preparing Drink	78	**73**	92	43	**100**	58
Prepare Drug Box	**100**	83	**100**	43	**100**	**100**
Reading	35	**100**	92	36	**100**	**78**
Using Bus Map	90	**90**	**100**	50	**100**	47
AVERAGE	72.0	**84.80**	90.95	48.76	**100**	70.00

This year we focused on the upper layer. We first formalized the run-time feature reconfiguration rules. First, any reconfiguration should respect the semantics of feature models and their attached constraints. Second, the reconfiguration should satisfy the requests from the middle layer, essentially selections and deselections of features. From there two requirements, we identified three possible outcomes: successful reconfiguration, impossible reconfiguration (selection/deselection conflicts), and “undefined” reconfiguration (not enough information to get through the process). We also determined the actions to take in these cases. In particular, in the last two cases, we decided to let the component configuration unchanged.

To implement this upper layer, we first attempted to rely on an existing feature model manipulation framework, namely FAMILIAR [45]. However, this approach suffers from a number of drawbacks. First, FAMILIAR is a standalone Java program, whereas the rest of the system is written in C++, for performance reasons and library availability. Hence, using FAMILIAR implies superfluous back and forth inter-module communications and data transformations. Second, we confirmed that FAMILIAR is more a system deployment tool than a run-time reconfiguration one. In particular it cannot fulfill all the reconfiguration rules that we have formalized. Therefore, we are completing a full re-implementation of the upper layer.

The programming language homogeneity permits a more efficient integration of the three layers. It particular, it becomes easier to incorporate our extensions to feature models such as quality metrics [34].

7.20. Scenario Description Language

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Nazli Temur, François Brémond.

Last year, we developed a scenario recognition engine based on the Synchronous Model of reactive systems. We now need a scenario description language friendly to our end users who are not computer scientists in general. In fact, Stars has already defined such a language. However, it is a declarative language based on (temporal) constraints. This is certainly not the most natural and the simplest way for end users to express their domain specific scenarios.

Consequently, we started this year a comparative study of different means to express scenarios in various domains (video understanding but also games, movies, music, criminology, military strategy...). We investigated 16 formalisms covering these domains. We defined a comparison grid based on criteria relevant for our video understanding goals. We retained 9 such criteria: application domain scope, ease of use, representation of scenario basic elements (background, scene, roles...), modularity (possibility of scenario hierarchy), time representation (absolute, logical, multi-clocks, no clocks...), expression of temporal constraints, representation of repetitive patterns, support for concurrency and parallelism, and finally formal foundations.

To complete the study, we conducted an experiment: describing a case study scenario using some of these formalisms to concretely estimate their advantages and drawbacks, especially their ease of use.

At this time, none of the studied languages fulfills completely our needs. Many languages are graphical ones. While this may appear as user friendly, scalability and automatic analysis become an issue. Some languages lack formal semantics, which is not acceptable in our case; others are merely extensions of computer languages, hence dedicated to specialists.

We plan to define our own version, which will rely on solid semantic foundations. (see section 7.21). To enforce user-friendliness, we started to collaborate with ergonomists.

7.21. Scenario Recognition

Participants: Annie Ressouche, Sabine Moisan, Jean-Paul Rigault, Ines Sarray, Daniel Gaffé.

Keywords: Synchronous Modeling, Model checking, Mealy machine, Cognitive systems.

For a long time, Stars strategy has been to favor the easy generation of activity recognition systems. These systems correspond to a succession of 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). Due to the large range of application domains (surveillance, safety, health care, ...), we propose a generic approach to design activity recognition engines. Moreover, such domains require high dependability due to possible safety issues. Thus, our approach should also rely on formal methods to describe, analyze, verify, and generate effective recognition engines. We consider activity recognition engines as reactive systems that react to input events from their environment and produce output events in the form of alarms or notifications. Such engines are intrinsically real time, reactive and they evolve in discrete time. As a consequence, to recognize scenarios, we adapt the usual techniques of synchronous modeling approach to express scenario behaviors. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario.

Our previous developments, on top of existing synchronous languages as Lustre and LE (see secrion 7.22), were convenient for rapid prototyping. However, even if LE is not a closed environment, it appeared as difficult as Lustre to customize, for efficiency reasons. This year, in the framework of Ines Sarray PhD thesis, we began to define a synchronous semantics for the future scenario language (see section 7.20). The idea is to generate automatically recognition engines at compilation time. The compilation itself is totally based on the semantics. To complete this approach we will rely on both our experiment with the LE language last year and on the LE compilation process.

7.22. The Clem Workflow

Participants: Annie Ressouche, Daniel Gaffé, Imane Khalis.

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 (see Figure 26) around the language to design, simulate, verify, and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality.

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 (see 7.21 and 7.23). We also improved the design of a new simulator for LE programs which integrates our new approach. In CLEM we generate an independent intermediate code (LEC) before specific target generations. This code represents the semantics of programs with 4-valued equation systems. In our design flow, we need to simulate programs at this level. Last year, we begun to develop such a simulator in order to integrate the data part of the language. The simulator GUI has been designed in Qt and the simulator takes into account the values carried by signals. This year, during her internship, Imane Khalis has completed the simulator to allow an external computation of data values and a communication with the simulator through a socket mechanism. With this last development, the LEC simulator is complete and is integrated in the CLEM toolkit.

7.23. Safe Composition in WCOMP Middleware for Internet of Things

Participants: Annie Ressouche, Daniel Gaffé, Ines Sarray, Jean-Yves Tigli.

Keywords: Synchronous Modeling, Ubiquitous Computing, middleware, internet of things

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 6.7) with a service-oriented middleware approach dealing with dynamic evolutions of system infrastructure in ubuquitous computing, and particularly in the Internet of Things (IoT). The Rainbow team (Nice-Sophia Antipolis University) proposes a component-based adaptive middleware (WComp [86], [85], [66]) to dynamically adapt and recompose assemblies of components.

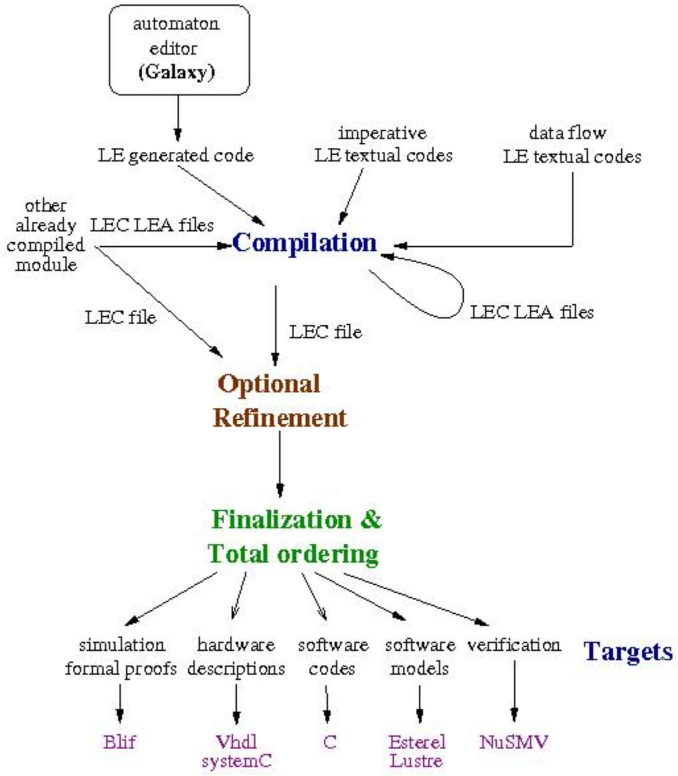


Figure 26. The Clem Toolkit

IoT is a way to combine computation and communication capabilities, sometimes in large scale information systems, with a huge number of complex devices connected to the physical world. Such infrastructures are often dedicated to the deployment of multiple applications, running concurrently. These applications are using shared devices from a common environment through different network middleware and numerous IoT protocols. Indeed, “Things”, also called the *Entities of Interest* [63], are the part of the real world in which devices are interacting and which must not be neglected. We aimed to model and validate concurrent accesses to shared devices without neglecting their associated *Entity of Interest*, their common physical context. One of the main challenge is then how to guarantee and validate some safety and integrity properties throughout the system’s evolution. In WComp middleware, we use synchronous models to facilitate the study and the validation of new composition mechanisms between applications at runtime. Then key problems to solve are: (1) how to specify and respect the "Thing" behavior? (2) how to ensure a safe combination of these multiple accesses when several services accesses a same entry of an *Entity of Interest* ? (3) how to manage multiple uses when applications simultaneously use a same service ?

This year, we addressed these problems by relying on formal method to model device behaviors as synchronous automata, taking into consideration their impact on the Entity Of Interest. Such an approach allows applying model-checking techniques to verify safety properties of applications. The main contribution is the definition of a sound way to compose models allowing context change adaptation. This composition relies on synchronous parallel composition paradigm. We proved that this operation preserves safety properties. However, it is not sufficient to obtain a global model of this composition because some devices may interact with the same Entity Of Interest. Moreover, several applications may use the same device services and then they can have concurrent accesses to their entries, so it can have an unexpected impact on our Entity Of Interest. Therefore, we added constraints to the device models composition and to applications level. We defined a generic way to express these constraints, independently of the knowledge about the devices and the applications, only their type is sufficient. We proposed the Description Constraint Language (DCL) to express these generic constraints and their compilation into LE Mealy machines. Thus we rely on CLEM model-checking facilities (see section7.22) to validate the constraints. As a consequence, this approach ensures the adaptation to a context change and offers a means to formally perform validation.

These results have been published in [43]

7.24. Design of UHD Panoramic Video Camera

Participants: Carolina Da Silva Gomes Crispim, Rachid Guerchouche, Daniel Gaffé, François Brémond.

The goal of this work is to investigate the possibilities of designing a new camera-based system for retail. This work was carried in the context of a collaboration between STARS and Neosensys. The system is composed of several high definition cameras placed in a configuration such as it makes it possible to obtain a panoramic vision with 360° of field of view. The work was divided into 2 parts: theoretical part and practical part.

In the theoretical part, the different characteristics of the desired system were studied, such as: number of cameras, resolution of each camera, the different characteristics of the sensors (WDR, HDR, exposure) etc. Depending on these characteristics, data transmission through an IP network was addressed. In addition to the hardware characteristics, the possibility of embedding stitching capabilities was studied. After spending some time understanding the background behind such techniques, an existing implementation of the stitching was adopted. Simulations were then made in order to estimate the characteristics of an FPGA capable of handling 5 cameras with 12MP resolution each. An existing FPGA architecture extensively used in the industry was chosen and a mathematical model was provided in order to estimate the characteristics of such FPGA according to the different parameters of the camera-based system.

In the practical part, an implementation of the two first steps of the stitching algorithm (homograph estimation and warping) was performed on a FPGA using 2 cameras. The problems of code optimisation were addressed in order to achieve a functioning implementation with respect to the memory and computation capabilities of the FPGA.

7.25. Brick & Mortar Cookies

Participant: Anaïs Ducoffe.

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. This final system is designed to be used without changing current camera network of the customer store, dedicated to security purpose. Analysis should be given at different time and space resolutions. For instance, day attendance can be as interesting as year attendance. 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*. Following Bosch *tasks* were selected and studied:

- **Loitering and idle object in a field tasks** enable to detect stop actions in a zone, when they happened and the stop positions.
- **Entering and leaving field tasks** enable to know when a person enters or leaves a zone.
- **Detect people in a field task** enables counting people in a zone.
- **Crossing lines tasks** for counting people entering or leaving shop. We are able to know when the line was crossed and in what sense.

It is not possible to get people trajectory when metadata from Bosch cameras are acquired in offline mode. Then we studied live connection to get metadata directly from the camera stream using a RTSP connection. Metadata information is saved in XML format.

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. Moreover these algorithms can be run on any type of video cameras (live and offline modes) whereas Bosch IVA can only be run in live mode on compatible Bosch cameras. Stars algorithms can also combine several cameras at the same time in order to track people across the camera network. We need those algorithms to sell a system that doesn't need a new camera network but can be used with existing ones. They will be integrated in the final product.

Tests in real conditions

A system for testing cameras and our software was installed in partner store (Super U). Cameras were installed and configured to process all our use cases and test our mechanism to extract the metadata. We used only Bosch camera with embedded IVA. We successfully acquire 2 hours of the desired metadata. The results of embedded algorithms are reliable on realistic data : we get good results in counting people and trajectories are accurate.

Metadata storage in database

Metadata have to be stored in a hierarchical way as request of the metadata by the application should be easy and quick. We choose to store metadata in a database. This database design was constraint by data storage speed and a quick access for live computation. Different parts of database (store information, devices description etc...) were designed to be as much independent as possible.

Web interface (GUI)

The graphic interface design is in progress. The interface will be a web based one to narrow compatibility problems: the application should be used as well with a computer as a tablet.

7.26. Monitoring Older People Experiments

Participants: Matias Marin, Etienne Corvée, François Brémond.

This year we have conducted many experiments, especially in Nice and partially in Thessaloniki, Lulea, Taiwan and Dublin, to validate our studies on monitoring older people suffering from various behavioral disorders in the framework of several projects.

DEM@CARE PROJECT

For the project Dem@care (see section 9.2.1.3), we use PCs with ASUS cameras, for monitoring and collecting a video dataset associated with metadata [0]. The software CAR is installed to automatically annotate the videos [0]. Data is recorded locally, and backups are made automatically and remotely: one on the server (LAB at the nursery home) and another backup at Inria. These data can be reached locally at the nursery home, thanks to the server located in the lab; also, they are all accessible from Inria network by ssh.

SafEE PROJECT

SafEE project (see section 9.1.1.2) experiments in the nursing home have started at the end of 2015, and are made up with different technologies (wifi, wired network, smart phones, Kinect, RFID, tablet...). In the nursing room, 2 PCs with KINECT2 are connected for monitoring the residents and are stored in a database. Another PC with Windows software is configured for SafEE serious game (cognitive games and music box, 1.7 version). Moreover, a Wifi access point will be used by medical staff at the nursery home to connect to a Graphical User Interface through a website designed by INDES team [0], to consult patient data (their daily activities), from the activity history stored in the database or in real-time. Another device called AromaCare is installed in the rooms, which is a connected aroma diffuser by Wi-Fi. With the app Aroma Therapeutics (smartphone or tablet) we can manage several diffusers, by scheduling different programs each day and change the intensity of the diffusion.

In patients' home the same devices than Nursing Home have to be configured. Today, only recordings are done and stored at Inria.

OTHER PROJECTS at ICP (Institut Claude Pompidou)

ICP has now a remote access by using rdesktop, which is safer than team-viewer screen sharing session. The installation of new experimentations (e.g. praxis, relaxation, serious games) is now in progress: the configuration expected includes PCs with KINECT2 connected at ICP network and accessible from Inria. Some experimentations will use wireless sensors (e.g. accelerometer, pressure), controlled by the app wireless tag (on smartphone or tablet) to measure fine patient activities: motion, kettle utilization, etc.

[0]https://team.inria.fr/stars/demcare-chu-dataset/

[0] https://team.inria.fr/stars/software/car-complex-activity-recognition-component-installation/

[0](http://webrobotics.inria.fr:8081/hop/events

WILLOW Project-Team

7. New Results

7.1. 3D object and scene modeling, analysis, and retrieval

7.1.1. The joint image handbook

Participants: Matthew Trager, Martial Hebert, Jean Ponce.

Given multiple perspective photographs, point correspondences form the “joint image”, effectively a replica of three-dimensional space distributed across its two-dimensional projections. This set can be characterized by multilinear equations over image coordinates, such as epipolar and trifocal constraints. In this work, we revisit the geometric and algebraic properties of the joint image, and address fundamental questions such as how many and which multilinearities are necessary and/or sufficient to determine camera geometry and/or image correspondences. Our new theoretical results answer these questions in a very general setting, and our work, published ICCV 2015 [17], is intended to serve as a “handbook” reference about multilinearities for practitioners.

7.1.2. 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 submitted to IJCV [25].

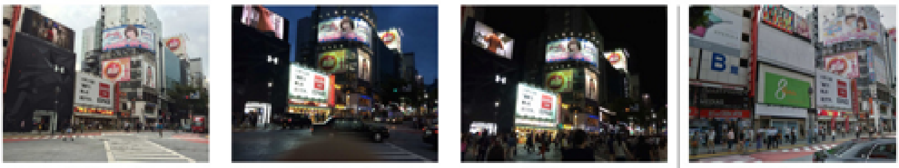
7.1.3. 24/7 place recognition by view synthesis

Participants: Akihiko Torii, Relja Arandjelović, Josef Sivic, Masatoshi Okutomi, Tomas Pajdla.

We address the problem of large-scale visual place recognition for situations where the scene undergoes a major change in appearance, for example, due to illumination (day/night), change of seasons, aging, or structural modifications over time such as buildings built or destroyed. Such situations represent a major challenge for current large-scale place recognition methods. This work has the following three principal contributions. First, we demonstrate that matching across large changes in the scene appearance becomes much easier when both the query image and the database image depict the scene from approximately the same viewpoint. Second, based on this observation, we develop a new place recognition approach that combines (i) an efficient synthesis of novel views with (ii) a compact indexable image representation. Third, we introduce a new challenging dataset of 1,125 camera-phone query images of Tokyo that contain major changes in illumination (day, sunset, night) as well as structural changes in the scene. We demonstrate that the proposed approach significantly outperforms other large-scale place recognition techniques on this challenging data. This work has been published at CVPR 2015 [16]. Figure 1 shows examples of the newly collected Tokyo 24/7 dataset.

7.1.4. NetVLAD: CNN architecture for weakly supervised place recognition

Participants: Relja Arandjelović, Petr Gronat, Akihiko Torii, Tomas Pajdla, Josef Sivic.



(a) Query 1. (b) Query 2. (c) Query 3. (d) Database image

Figure 1. Example query images from the newly collected 24/7 Tokyo dataset. Each place in the query set is captured at different times of day: (a) daytime, (b) sunset, and (c) night. For comparison, the database street-view image at a close-by position is shown in (d). Note the major changes in appearance (illumination changes in the scene) between the database image (d) and the query images (a,b,c)

In [21], 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 is under review. Figure 2 shows some qualitative results.

(a) Mobile phone query (b) Retrieved image of same place

Figure 2. 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).

7.2. Category-level object and scene recognition

7.2.1. Is object localization for free? – Weakly-supervised learning with convolutional neural networks

Participants: Maxime Oquab, Leon Bottou [MSR New York], Ivan Laptev, Josef Sivic.

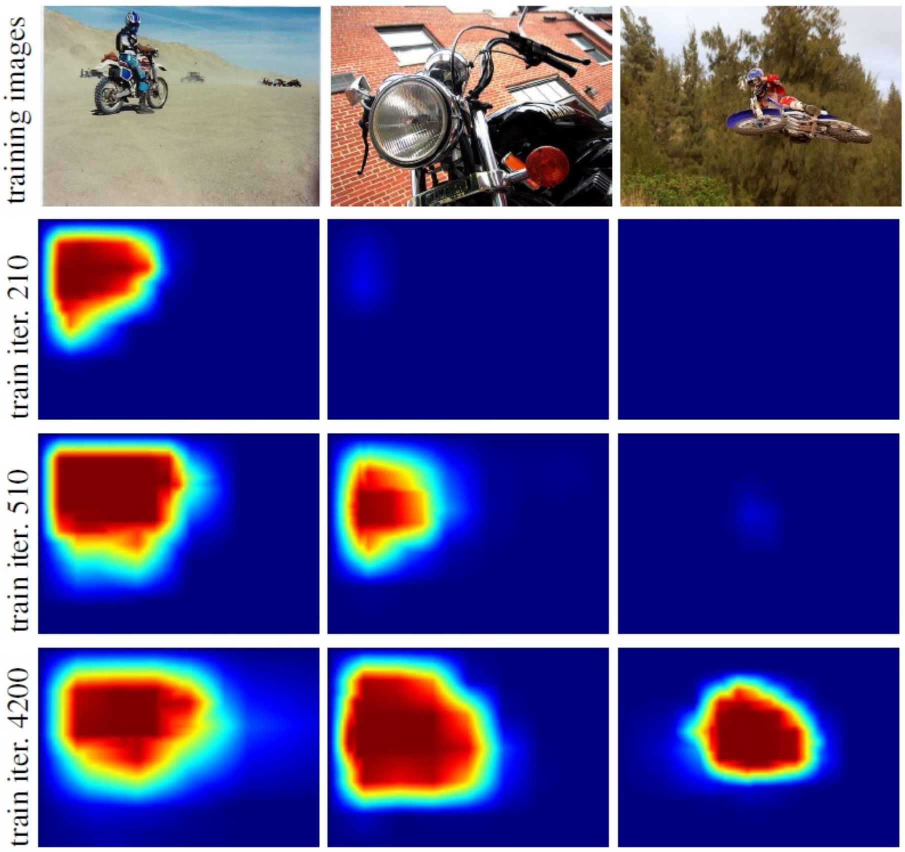


Figure 3. Evolution of localization score maps for the motorbike class over iterations of our weakly-supervised CNN training. Note that locations of objects with more usual appearance are discovered earlier during training.

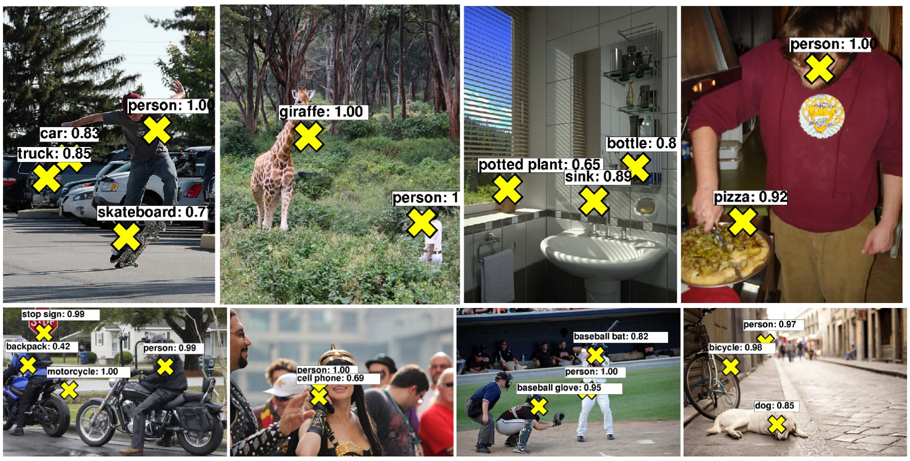


Figure 4. Example location predictions for images from the Microsoft COCO validation set obtained by our weakly-supervised method. Note that our method does not use object locations at training time, yet can predict locations of objects in test images (yellow crosses). The method outputs the most confident location for most confident object classes.

Successful methods for visual object recognition typically rely on training datasets containing lots of richly annotated images. Detailed image annotation, e.g. by object bounding boxes, however, is both expensive and often subjective. We describe a weakly supervised convolutional neural network (CNN) for object classification that relies only on image-level labels, yet can learn from cluttered scenes containing multiple objects (see Figure 3). We quantify its object classification and object location prediction performance on the Pascal VOC 2012 (20 object classes) and the much larger Microsoft COCO (80 object classes) datasets. We find that the network (i) outputs accurate image-level labels, (ii) predicts approximate locations (but not extents) of objects, and (iii) performs comparably to its fully-supervised counterparts using object bounding box annotation for training. This work has been published at CVPR 2015 [14] . Illustration of localization results by our method in Microsoft COCO dataset is shown in Figure 4 .

7.2.2. Unsupervised Object Discovery and Localization in the Wild: Part-based Matching with Bottom-up Region Proposals

Participants: Minsu Cho, Suha Kwak, Cordelia Schmid, Jean Ponce.

In [8], we address *unsupervised* discovery and localization of dominant objects from a noisy image collection of multiple object classes. The setting of this problem is fully unsupervised (Fig. 5), without even image-level annotations or any assumption of a single dominant class. This is significantly more general than typical colocalization, cosegmentation, or weakly-supervised localization tasks. We tackle the unsupervised discovery and localization problem using a part-based region matching approach: We use off-the-shelf region proposals to form a set of candidate bounding boxes for *objects* and *object parts*. These regions are efficiently matched across images using a probabilistic Hough transform that evaluates the confidence for each candidate correspondence considering both appearance similarity and spatial consistency. Dominant objects are discovered and localized by comparing the scores of candidate regions and selecting those that stand out over other regions containing them. Extensive evaluations on standard benchmarks (e.g., Object Discovery and PASCAL VOC 2007 datasets) demonstrate that the proposed approach significantly outperforms the current state of the art in colocalization, and achieves robust object discovery even in a fully unsupervised setting. This work has been published in CVPR 2015 [8] as oral presentation.

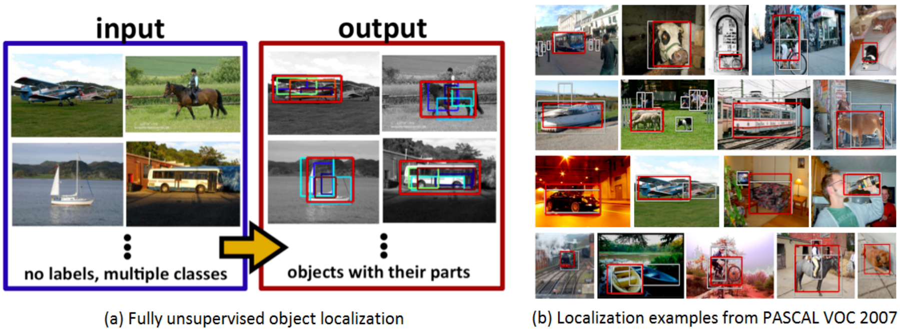


Figure 5. Unsupervised object discovery in the wild. We tackle object localization in an unsupervised scenario without any type of annotations, where a given image collection may contain multiple dominant object classes and even outlier images. The proposed method discovers object instances (red bounding boxes) with their distinctive parts (smaller boxes).

7.2.3. Unsupervised Object Discovery and Tracking in Video Collections

Participants: Suha Kwak, Minsu Cho, Ivan Laptev, Jean Ponce, Cordelia Schmid.

In [11], we address the problem of automatically localizing dominant objects as spatio-temporal tubes in a noisy collection of videos with minimal or even no supervision. We formulate the problem as a combination of two complementary processes: discovery and tracking (Figure 6). The first one establishes correspondences bet ween prominent regions across videos, and the second one associates similar object regions within the same video. It is empirically demonstrated that our method can handle video collections featuring multiple object classes, and substantially outperforms the state of the art in colocalization, even though it tackles a broader problem with much less supervision. This work has been published in ICCV 2015.

7.2.4. Linking Past to Present: Discovering Style in Two Centuries of Architecture

Participants: Stefan Lee, Nicolas Maisonneuve, David Crandall, Alexei A. Efros, Josef Sivic.

With vast quantities of imagery now available online, researchers have begun to explore whether visual patterns can be discovered automatically. Here we consider the particular domain of architecture, using huge collections of street-level imagery to find visual patterns that correspond to semantic-level architectural elements distinctive to particular time periods. We use this analysis both to date buildings, as well as to discover how functionally similar architectural elements (e.g. windows, doors, balconies, etc.) have changed over time due to evolving styles. We validate the methods by combining a large dataset of nearly 150,000 Google Street View images from Paris with a cadastre map to infer approximate construction date for each facade. Not only could our analysis be used for dating or geo-localizing buildings based on architectural features, but it also could give architects and historians new tools for confirming known theories or even discovering new ones. The work was published in [13] and the results are illustrated in figure 7 .

7.2.5. 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 [22], 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,

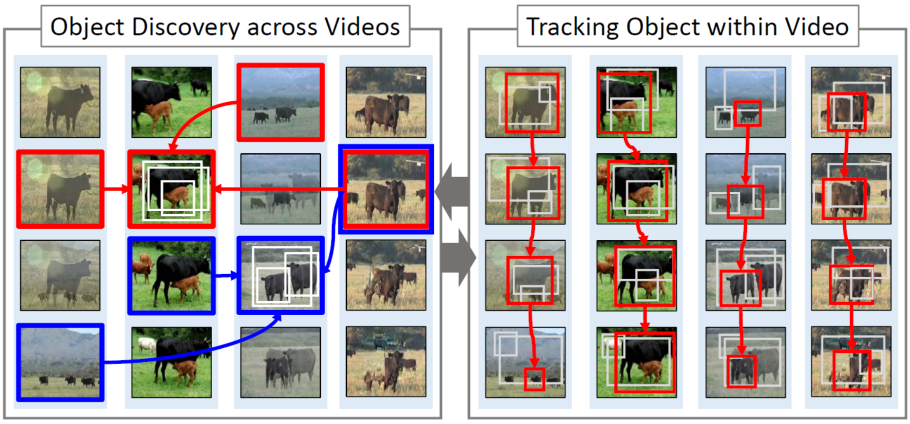


Figure 6. Dominant objects in a video collection are discovered by analyzing correspondences between prominent regions across videos (left). Within each video, object candidates, discovered by the former process, are temporally associated and a smooth spatio-temporal localization is estimated (right). These processes are alternated until convergence or up to a fixed number of iterations.

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 is under review. The proposed method and its qualitative result are illustrated in Figure 8 .

7.3. Image restoration, manipulation and enhancement

7.3.1. Learning a Convolutional Neural Network for Non-uniform Motion Blur Removal

Participants: Jian Sun, Wenfei Cao, Zongben Xu, Jean Ponce.

In this work, we address the problem of estimating and removing non-uniform motion blur from a single blurry image. We propose a deep learning approach to predicting the probabilistic distribution of motion blur at the patch level using a convolutional neural network (CNN). We further extend the candidate set of motion kernels predicted by the CNN using carefully designed image rotations. A Markov random field model is then used to infer a dense non-uniform motion blur field enforcing the motion smoothness. Finally the motion blur is removed by a non-uniform deblurring model using patch-level image prior. Experimental evaluations show that our approach can effectively estimate and remove complex non-uniform motion blur that cannot be well achieved by the previous approaches. This work has been published at CVPR 2015[15].

7.3.2. Robust Image Filtering Using Joint Static and Dynamic Guidance

Participants: Bumsub Ham, Minsu Cho, Jean Ponce.

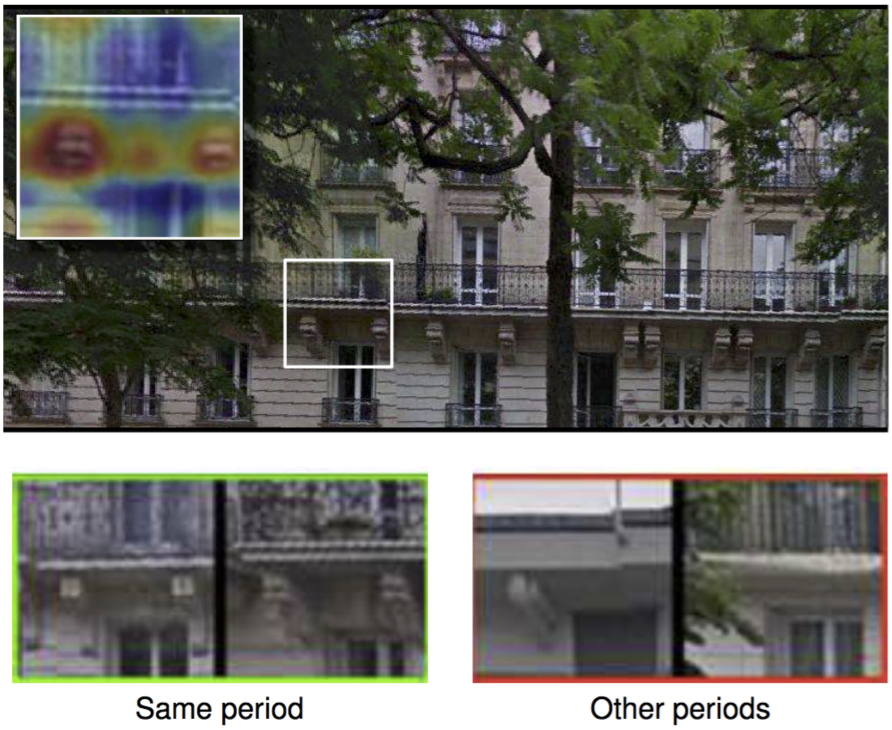


Figure 7. Using thousands of Street View images aligned to a cadastral map, we automatically find visual elements distinctive to particular architectural periods. For example, the patch in white above was found to be distinctive to the Haussmann period (late 1800's) in Paris, while the heat map (inset) reveals that the ornate balcony supports are the most distinctive features. We can also find functionally-similar elements fromthe same and different time periods (bottom).

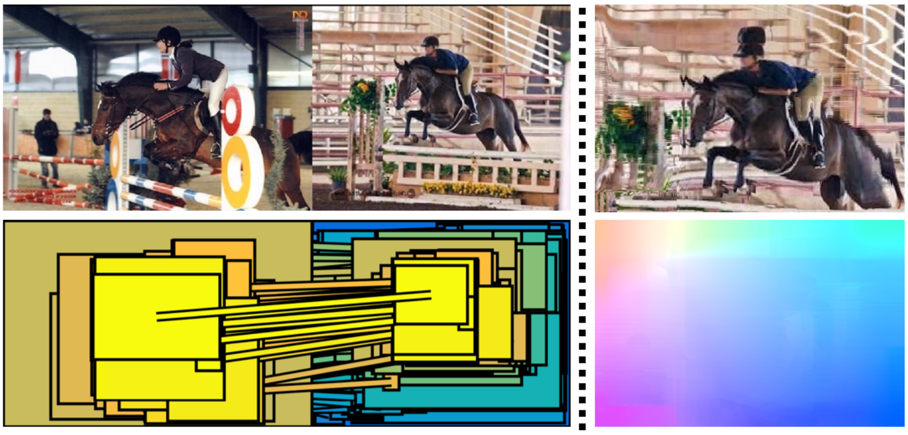


Figure 8. Proposal flow generates a reliable scene flow between similar images by establishing geometrically consistent correspondences between object proposals. (Left) Region-based scene flow by matching object proposals. (Right) Color-coded dense flow field generated from the region matches, and image warping using the flow.

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 [10]. The SD filter is illustrated in Figure 9 .

7.3.3. PCS-Net: A Deep learning approach to image restoration

Participants: Jian Sun, Jean Ponce.

This work introduces a novel framework for image restoration casting this problem as a joint classification and regression task. This is a learning-based approach, which first classifies degraded image patches into different categories, then restores these patches using category-specific models. We implement this idea by designing a novel convolutional neural network (dubbed PCS-Net), combining a CNN-based patch classification subnet with a novel patch category switched CNN architecture for category-specific restoration. The proposed PCS-Net learns different weights for different patch categories in a common network structure. Experiments on standard benchmarks show that our approach matches or improves upon the state of the art in image super-resolution and denoising. This work is under review.

7.4. Human activity capture and classification

7.4.1. P-CNN: Pose-based CNN Features for Action Recognition

Participants: Guilhem Chéron, Ivan Laptev, Cordelia Schmid.

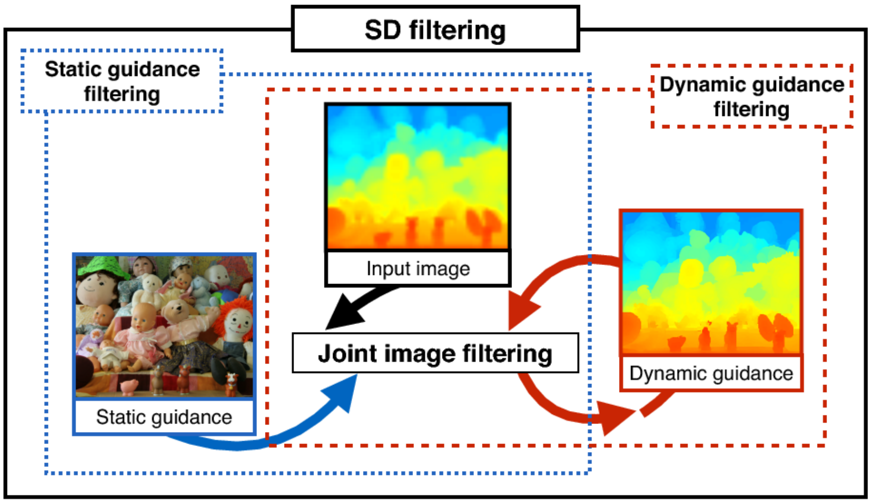


Figure 9. Sketch of joint image filtering and SD filtering: Static guidance filtering convolves an input image with a weight function computed from static guidance, as in the dotted blue box. Dynamic guidance filtering uses weight functions that are repeatedly obtained from regularized input images, 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.

This work [9] targets human action recognition in video. We argue for the importance of a representation derived from human pose. To this end we propose a new Pose-based Convolutional Neural Network descriptor (P-CNN) for action recognition. The descriptor aggregates motion and appearance information along tracks of human body parts as shown in Figure 10 . We experiment with P-CNN features obtained both for automatically estimated and manually annotated human poses. We evaluate our method on JHMDB and MPII Cooking datasets. For both datasets our method shows consistent improvement over the state of the art. This work has been published at ICCV 2015 [9], and P-CNN code (Matlab) is available online at http://www.di.ens.fr/willow/research/p-cnn/.

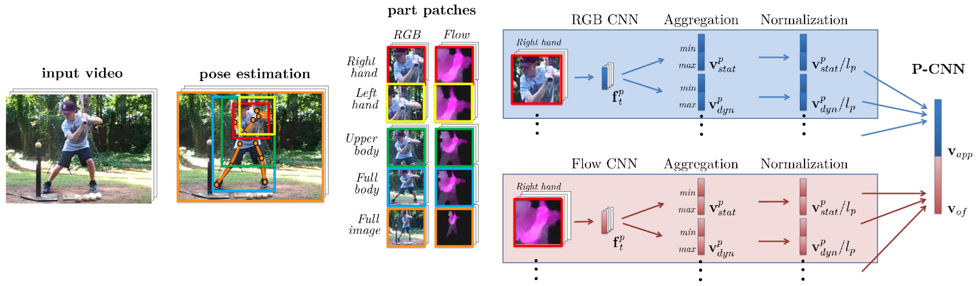


Figure 10. P-CNN features. From left to right: Input video. Human pose. Patches of appearance and optical flow for human body parts. One RGB and one flow CNN descriptor is extracted per frame and per part. Frame descriptors are aggregated over time to obtain the video descriptor. Video descriptors are normalized and concatenated into appearance features and flow features. The final P-CNN feature is the concatenation of appearance and flow.

7.4.2. Context-aware CNNs for person head detection

Participants: Tuan-Hung Vu, Anton Osokin, Ivan Laptev.

Person detection is a key problem for many computer vision tasks. While face detection has reached maturity, detecting people under a full variation of camera view-points, human poses, lighting conditions and occlusions is still a difficult challenge. In this work we focus on detecting human heads in natural scenes. Starting from the recent local R-CNN object detector, we extend it with two types of contextual cues. First, we leverage person-scene relations and propose a Global CNN model trained to predict positions and scales of heads directly from the full image. Second, we explicitly model pairwise relations among objects and train a Pairwise CNN model using a structured-output surrogate loss. The Local, Global and Pairwise models are combined into a joint CNN framework. To train and test our full model, we introduce a large dataset composed of 369,846 human heads annotated in 224,740 movie frames. We evaluate our method and demonstrate improvements of person head detection against several recent baselines in three datasets. We also show improvements of the detection speed provided by our model. This work has been published at ICCV 2015 [18]. The code and the new dataset developed in this work are available online at http://www.di.ens.fr/willow/research/headdetection/.

7.4.3. On Pairwise Costs for Network Flow Multi-Object Tracking

Participants: Visesh Chari, Simon Lacoste-Julien, Ivan Laptev, Josef Sivic.

Multi-object tracking has been recently approached with the min-cost network flow optimization techniques. Such methods simultaneously resolve multiple object tracks in a video and enable modeling of dependencies among tracks. Min-cost network flow methods also fit well within the "tracking-by-detection" paradigm where object trajectories are obtained by connecting per-frame outputs of an object detector. Object detectors, however, often fail due to occlusions and clutter in the video. To cope with such situations, we propose an approach that regularizes the tracker by adding second order costs to the min-cost network flow framework. While solving such a problem with integer variables is NP-hard, we present a convex relaxation with an efficient rounding heuristic which empirically gives certificates of small suboptimality. Results are shown on real world video sequences and demonstrate that the new constraints help selecting longer and more accurate tracks improving over the baseline tracking-by-detection method. This work has been published at CVPR 2015 [7].

7.4.4. Pose Estimation and Segmentation of Multiple People in Stereoscopic Movies

Participants: Guillaume Seguin, Karteek Alahari, Josef Sivic, Ivan Laptev.

We describe a method to obtain a pixel-wise segmentation and pose estimation of multiple people in stereoscopic videos, illustrated in Figure 11 . This task involves challenges such as dealing with unconstrained stereoscopic video, non-stationary cameras, and complex indoor and outdoor dynamic scenes with multiple people. We cast the problem as a discrete labelling task involving multiple person labels, devise a suitable cost function, and optimize it efficiently. The contributions of our work are two-fold: First, we develop a segmentation model incorporating person detections and learnt articulated pose segmentation masks, as well as colour, motion, and stereo disparity cues. The model also explicitly represents depth ordering and occlusion. Second, we introduce a stereoscopic dataset with frames extracted from feature-length movies “StreetDance 3D" and “Pina". The dataset contains 587 annotated human poses, 1158 bounding box annotations and 686 pixel-wise segmentations of people. The dataset is composed of indoor and outdoor scenes depicting multiple people with frequent occlusions. We demonstrate results on our new challenging dataset, as well as on the H2view dataset from (Sheasby et al.’s ACCV 2012). This work has been published at PAMI [4].

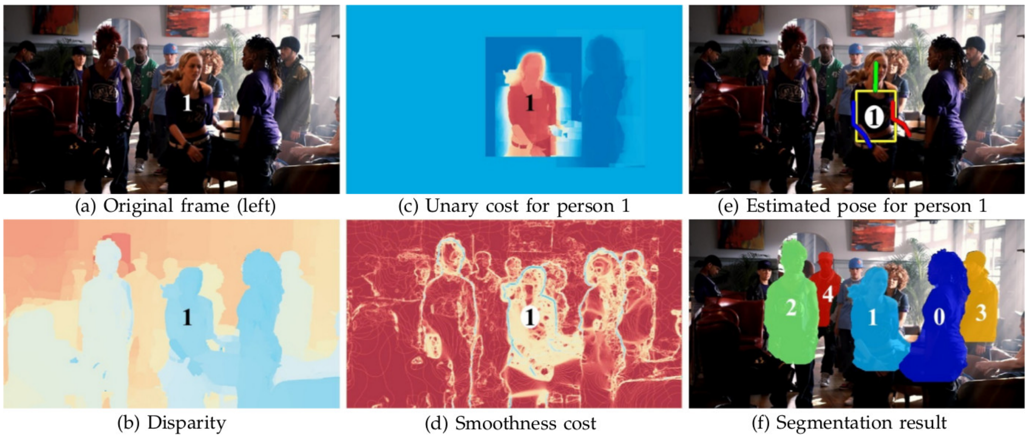


Figure 11. Starting from a stereo pair (a), we estimate disparity maps (b). Using both appearance and disparity cues, we detect persons and estimate their poses (e). We combine pose information with disparity information and occlusion reasonning to compute the unary potentials of a CRF (c) and use standard color and motion cues to compute the binary terms (d). We optimize the CRF problem to produce the final, layered segmentation (f).

7.4.5. Weakly-Supervised Alignment of Video with Text

Participants: Piotr Bojanowski, Rémi Lajugie, Edouard Grave, Francis Bach, Ivan Laptev, Jean Ponce, Cordelia Schmid.

In this work [6], we design a method for aligning natural language sentences with a video stream. Suppose that we are given a set of videos, along with natural language descriptions in the form of multiple sentences (e.g., manual annotations, movie scripts, sport summaries etc.), and that these sentences appear in the same temporal order as their visual counterparts. We propose here a method for aligning the two modalities, i.e., automatically providing a time stamp for every sentence (see Fig. 12). Given vectorial features for both video and text, we propose to cast this task as a temporal assignment problem, with an implicit linear mapping between the two feature modalities. We formulate this problem as an integer quadratic program, and solve its continuous convex relaxation using an efficient conditional gradient algorithm. Several rounding procedures are proposed to construct the final integer solution. After demonstrating significant improvements over the state of the art on the related task of aligning video with symbolic labels, we evaluate our method on a challenging dataset of videos with associated textual descriptions, using both bag-of-words and continuous representations for text. This work has been published at CVPR 2015 [6].

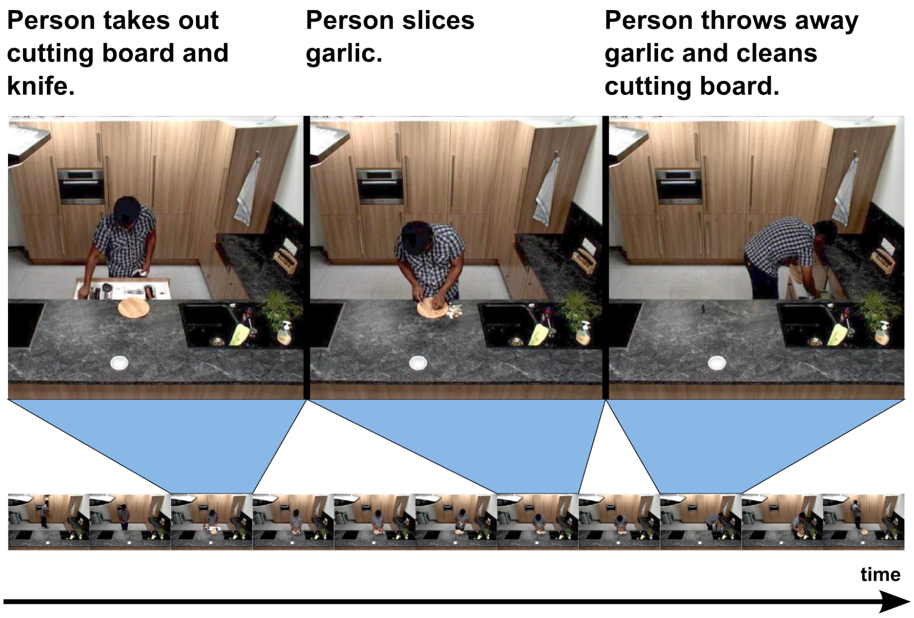


Figure 12. Illustration of the text to video alignment problem. As an output, our model provides a temporal location for every sentence.

7.4.6. Unsupervised learning from narrated instruction videos

Participants: Jean-Baptiste Alayrac, Piotr Bojanowski, Nishant Agrawal, Josef Sivic, Ivan Laptev, Simon Lacoste-Julien.

In [20], 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 is under review.

7.4.7. Long-term Temporal Convolutions for Action Recognition

Participants: Gül 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 [27], 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 13 .

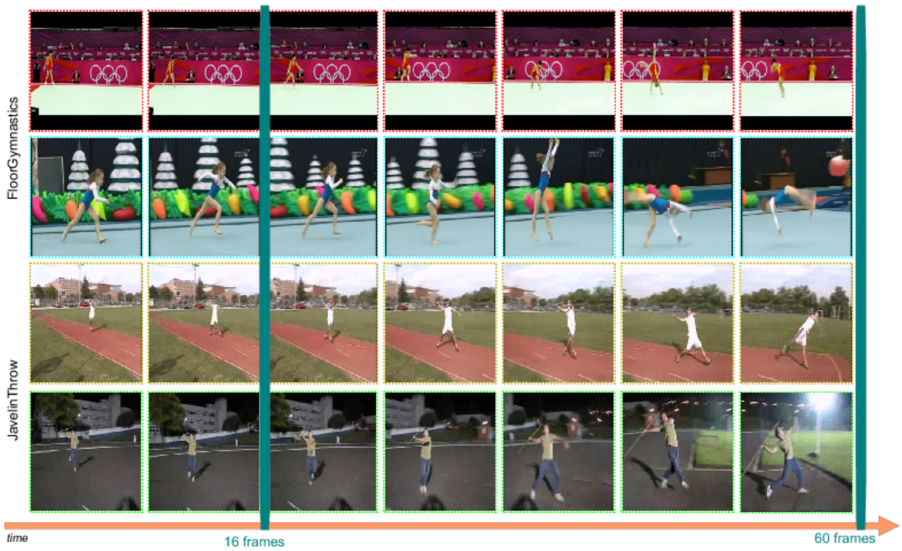


Figure 13. The highest improvement of long-term temporal convolutions in terms of class accuracy is for "JavelinThrow". For 16-frame network, it is mostly confused with "FloorGymnastics" class. We visualize sample videos with 7 frames extracted at every 8 frames. The intuitive explanation is that both 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.

7.4.8. Thin-Slicing forPose: Learning to Understand Pose without Explicit Pose Estimation

Participants: Suha Kwak, Minsu Cho, Ivan Laptev.

In [23], 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 14). 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 is under review.

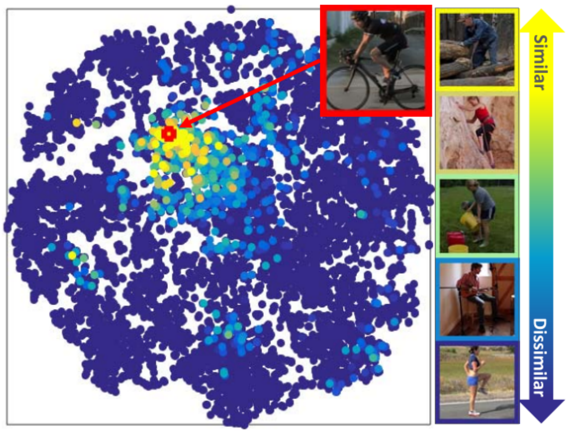


Figure 14. The manifold of our pose embedding visualized using t-SNE. Each point represents a human pose image. To better show correlation between the pose embedding and annotated pose, we color-code pose similarities in annotation between an arbitrary target image (red box) and all the other images. 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.

7.4.9. Instance-level video segmentation from object tracks

Participants: Guillaume Seguin, Piotr Bojanowski, Rémi Lajugie, Ivan Laptev.

In [26], 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 15 , 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 is under review.

Figure 15. 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).