



RESEARCH CENTER

FIELD

**Perception, Cognition and Interaction**

Activity Report 2014

# Section New Results

Edition: 2015-03-24





## DATA AND KNOWLEDGE REPRESENTATION AND PROCESSING

1. DAHU Project-Team	5
2. DREAM Project-Team	7
3. EXMO Project-Team	12
4. GRAPHIK Project-Team	15
5. LINKS Team	22
6. MAGNET Team	24
7. MAIA Project-Team	27
8. OAK Project-Team	37
9. ORPAILLEUR Project-Team	40
10. SMIS Project-Team	47
11. WIMMICS Project-Team	50
12. ZENITH Project-Team	59

## INTERACTION AND VISUALIZATION

13. ALICE Project-Team	67
14. AVIZ Project-Team	68
15. HYBRID Project-Team	76
16. IMAGINE Project-Team	89
17. IN-SITU Project-Team	98
18. MANAO Project-Team	104
19. MAVERICK Project-Team	110
20. MIMETIC Project-Team	118
21. MINT Project-Team	124
22. POTIOC Project-Team	131
23. REVES Project-Team	138
24. TITANE Project-Team	149

## LANGUAGE, SPEECH AND AUDIO

25. ALPAGE Project-Team	153
26. MULTISPEECH Team	162
27. PANAMA Project-Team	170
28. SEMAGRAMME Project-Team	182

## ROBOTICS AND SMART ENVIRONMENTS

29. E-MOTION Project-Team	186
30. FLOWERS Project-Team	204
31. HEPHAISTOS Team	215
32. LAGADIC Project-Team	219
33. RITS Team	227

## VISION, PERCEPTION AND MULTIMEDIA INTERPRETATION

34. AYIN Team	236
35. LEAR Project-Team	248
36. LINKMEDIA Project-Team	262

37. MAGRIT Project-Team .....	269
38. MORPHEO Project-Team .....	273
39. PERCEPTION Project-Team .....	277
40. PRIMA Project-Team .....	282
41. SIROCCO Project-Team .....	286
42. STARS Project-Team .....	295
43. WILLOW Project-Team .....	331

## DAHU Project-Team

# 6. New Results

## 6.1. Highlights of the Year

Victor Vianu was elected member of Academia Europaea.

BEST PAPER AWARD :

[21] **Joint 25th International Conference on Rewriting Techniques and Applications and 12th International Conference on Typed Lambda Calculi and Applications.** S. SCHMITZ.

## 6.2. Distributed data management

**Participants:** Serge Abiteboul, Émilien Antoine, Victor Vianu.

The management of Web users' personal information is increasingly distributed across a broad array of applications and systems, including online social networks and cloud-based services. While users wish to share and integrate data using these systems, it is increasingly difficult to avoid the risks of unintended disclosures or unauthorized access by applications.

In [15], we study deduction in the presence of inconsistencies. Following previous works, we capture deduction via datalog programs and inconsistencies through violations of functional dependencies (FDs). We study and compare two semantics for datalog with FDs: the first, of a logical nature, is based on inferring facts one at a time, while never violating the FDs; the second, of an operational nature, consists in a fixpoint computation in which maximal sets of facts consistent with the FDs are inferred at each stage. Both semantics are nondeterministic, yielding sets of possible worlds. We introduce a PTIME (in the size of the extensional data) algorithm, that given a datalog program, a set of FDs and an input instance, produces a c-table representation of the set of possible worlds. Then, we propose to quantify nondeterminism with probabilities, by means of a probabilistic semantics. We consider the problem of capturing possible worlds along with their probabilities via probabilistic c-tables. We then study classical computational problems in this novel context. We consider the problems of computing the probabilities of answers, of identifying most likely supports for answers, and of determining the extensional facts that are most influential for deriving a particular fact. We show that the interplay of recursion and FDs leads to novel technical challenges in the context of these problems.

Jakub Kalas (ENS Cachan) spent 4 months in the team working on Personal Information Management Systems, using primarily positioning from data mobile phone and data from search engines.

## 6.3. Query Processing for the Web

**Participants:** Johann Brault-Baron, Arnaud Durand, Nadime Francis, Luc Segoufin, Cristina Sirangelo.

In many applications the output of a query may have a huge size and enumerating all the answers may already consume too many of the allowed resources. In this case it may be appropriate to first output a small subset of the answers and then, on demand, output a subsequent small numbers of answers and so on until all possible answers have been exhausted. To make this even more attractive it is preferable to be able to minimize the time necessary to output the first answers and, from a given set of answers, also minimize the time necessary to output the next set of answers - this second time interval is known as the *delay*. We have shown that this was doable with a almost linear preprocessing time and constant enumeration delay for first-order queries over structures of low degree [18]. We also presented a survey about this work at the Intl. Symp. on Theoretical Aspects of Computer Science (STACS) [22].

We have also been interested in querying data structured as graphs, which is nowadays spreading on the Web. Examples are social networks, linked data and the semantic web, via the RDF format. We have tackled the problem of answering queries over graph databases which are available only through a given set of views. This is a common situation in many applications where access to data needs to be either controlled or optimized. In [19] we have studied when it is possible to rewrite over the views queries issued on the original data, and which query languages are needed for this purpose. We have considered views and queries expressed as Regular path queries, a very common graph query language.

## 6.4. Complexity in Counter Systems and Substructural Logics

**Participant:** Sylvain Schmitz.

The ties between propositional substructural logics (like linear logic, relevance logic, affine logic, etc.) on the one hand and extensions of vector addition systems on the other hand have long been known, as they lie for instance at the heart of undecidability proof of provability in linear logic. In a series of papers we recently revisited these connections with an eye on complexity issues. This allowed us to prove tight complexity bounds on provability in affine and contractive fragments of linear logic [20], in affine ( $!$ ,  $\oplus$ )-Horn linear logic [16], and in implicative relevance logic [21] (an open problem for more than 25 years, with consequences on type inhabitation in the  $\lambda I$ -calculus). Our work also yields a new Tower lower bound on reachability in branching vector addition systems [20], which entails the same lower bound for logics on XML trees [4], for which decidability is still open.

Although the connection with data logics might not seem obvious at first, the models of counter systems considered in these papers are tightly connected with logics for XML processing [5], [4]. Further investigations in the relationships between data logics, substructural logics, and counter systems are the main thrust behind the just accepted ANR PRODAQ project (see Section 8.1.1).

## 6.5. Incomplete Databases

**Participants:** Nadime Francis, Cristina Sirangelo.

Incomplete databases appear in several different scenarios. Intuitively, pieces of information might not be available, or can get lost due to failures in storage or transmission. Alternatively, some complex data management tasks, such as data integration or data exchange, use incomplete databases as a model for databases with missing or unspecified information. In the context of the Web, these tasks have become even more crucial, which increased the need to handle incomplete databases. Given an incomplete database, one of the first questions to answer is that of consistency: can we make sure that the incomplete database can be completed as a real database conforming to some specified schema.

Together with Claire David and Filip Murlak, we have considered this problem when incomplete instances are represented as incomplete XML documents, where labels and nodes might be missing, and we additionally assume the DOM semantics, meaning that nodes never lose their identity (otherwise, they are considered completely lost). These are further modeled as injective tree patterns using child and descendant relations. In [17], we close the question of the complexity of checking the consistency of such patterns with regards to a fixed regular tree language: it is polynomial for patterns that do not use child edges, and for patterns that use both, it is already NP-complete for patterns using at most two descendant edges per branch, the case for at most one descendant edge being already known to be polynomial.

In [12] we have studied the feasibility of query answering in the presence of incomplete information in data. In particular we have generalized conditions allowing classical query evaluation techniques to be applicable also in the presence of incompleteness. Our results show that conditions found in some of our previous work can be significantly relaxed so as to account for more complex semantics of incompleteness, originating in the fields of logic programming, programming semantics and data exchange.

## **DREAM Project-Team**

# **6. New Results**

## **6.1. Simulator-based decision support**

**Participants:** Louis Bonneau de Beaufort, Tassadit Bouadi, Marie-Odile Cordier, Thomas Guyet, Christine Largouët, Véronique Masson, René Quiniou, Sophie Robin, Laurence Rozé, Yulong Zhao.

### ***6.1.1. Model-checking an ecosystem for decision-aid***

In previous work we have proposed to use qualitative modelling to model ecosystems and we defined a set of high level query patterns to explore the system [53]. This approach has been applied on real-case ecosystems (coral-reef ecosystem in New-Caledonia, fisheries ecosystem in the English channel) and implemented in a tool called EcoMata.

In recent studies we have focussed on the formalization of the qualitative model automatically built from an abstracted ecosystem description. Ecosystems share some common features with concurrent systems represented in the model-checking field: the system complexity is due to interacting components and the system evolution is event-driven and submitted to temporal constraints. However if model-checking techniques are dedicated to finite state systems, ecosystem models are usually represented by analytical models as a set of differential equations. Some studies present how to quantize continuous-time systems in order to diagnose them as discrete-event systems. We proposed a method to build automatically a network of timed automata from various information on the system: description of interactions between components, human knowledge, simple models of population dynamics. The key point is to quantize the continuous-time sub-systems and to get a qualitative model described as network of timed automata. To reduce the size of this network, important after the automatic generation, a learning machine algorithm has been applied in order to reduce the number of "similar" locations. This work has been published in [37].

### ***6.1.2. Controller 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 achieving 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 developed for the marine ecosystem example [69].

More recently, two new methods for finding the optimal strategies have been proposed. The first one uses controller synthesis on timed automata and exploits the efficiency of well-recognized tools. The second one deals with a set of similar models and extracts a more general strategy, closer to what is expected by the stakeholders. These methods have been applied in the context of herd management on a catchment. Yulong Zhao defended his Phd this year on this research subject [5].

### ***6.1.3. 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 [35]. 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 [10], [25]. It provides the user with navigation operators along the dimensions hierarchies (i.e. specialize / generalize) while ensuring an online calculation of the associated skyline.

#### 6.1.4. Post-mining classification rules

We consider sets of classification rules with quantitative attributes inferred by supervised machine learning, as in the framework of the SACADEAU project. Our aim is to improve human understanding of such sets of rules. Often, output quantitative rules contains too many intervals that are difficult to interpret. It is thus important to merge some of these intervals in order to get more understandable rules. However, blindly merging rules may decrease rule quality. To counter that, we proposed two algorithms for merging intervals via clustering techniques that take into account the final rule quality. The approach automatically detects the most adapted number of clusters required to merge intervals while maintaining rule quality.

### 6.2. Data Mining

**Participants:** Sid Ahmed Benabderrahmane, Marie-Odile Cordier, Serge Vladimir Emteu Tchagou, Thomas Guyet, Yves Moinard, René Quiniou, Alexandre Termier.

#### 6.2.1. Application of sequential pattern mining with intervals

Our theoretical work on sequential pattern mining with intervals [47] has been applied to two real issues: the customer relationship management and analysis of care pathways.

Customer Relationship Management (CRM) comprises a set of tools for managing the interactions between a company and its customers. The main objective of the data analysts is to propose the correct service to a customer at the correct moment by applying decision rules. If rules or sequential patterns can predict the interaction that can follow a sequence of actions or events, they can not predict at what time such actions have the highest probability to occur. The objective of temporal pattern mining is to refine the prediction by extracting patterns with information about the duration and delay between the events. This year we have experimented two algorithms on a CRM databases, QTIPrefixSpan [47] and TGSP [68], to extract sequential patterns with quantitative temporal information. We have integrated the TGSP algorithm into an interface to visualize and to browse the extracted patterns. A paper describing this contribution have been recently accepted in a workshop [41].

The QTIPrefixSpan algorithm has also been applied to the analysis of care pathways. The pharmaco-epidemiology platform of the Rennes hospital was interested in characterizing the care pathways preceding the epileptic seizures of stable epileptic patients. A care pathway consist of the sequence of drug exposures (temporal intervals). The objective is to study the ability of QTIPrefixSpan to identify drug switches between original and generic anti-epileptic drugs. This work is still in progress and will be extended in the PEPS project (see section 8.1.1).

#### 6.2.2. Multiscale segmentation of satellite image time series

Satellite images allow the acquisition of large-scale ground vegetation. Images are available along several years with a high acquisition frequency (1 image every two weeks). Such data are called satellite image time series (SITS). In [45], 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. This work is now under submission to the journal on Remote Sensing in Environment. The main issue of this approach was the required computation resources (time and memory). Last year, we applied 1D-SAX to reduce data dimensionality [21]. This approach on the supervised classification of large SITS of Senegal and we showed that 1D-SAX approaches the classification results of time series while significantly reducing the required memory storage of the images.

This year, we first continued to explore the supervised classification of SITS using classification trees for time-series [40] by implementing a parallelized version of this algorithm. Secondly, we explored the adaption of the object-oriented segmentation to SITS. The object-oriented segmentation [34] is able to segment images based on the segment uniformity. We proposed a measure for time-series uniformity to adapt the segmentation algorithm and applied it on large multivariate SITS of Senegal. This work have been presented to the conference on spatial analysis and geography [16]. A collaboration with A. Fall (Université Paris-13) have been initiated to compare our results on the Senegal with ground observations. Moreover, we planned to apply our algorithm to analyse the land use in Peru (collaboration with A. Marshall, Université Paris 13/PRODIG).

### 6.2.3. 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 agronomist to test all their hypothesis. 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 is made 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.), of a specific geographic area.

In past years, we worked on the exploration of graph mining techniques, such as gSPAN [67], 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 will reproduce the same patterns.

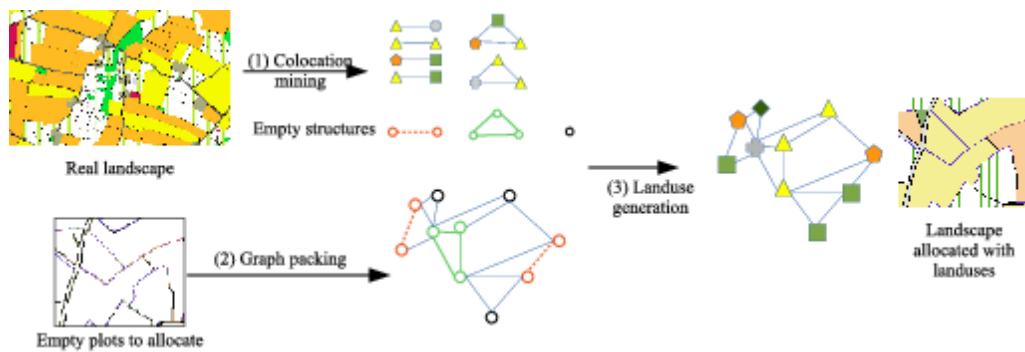


Figure 1. Simulation process in three steps : 1) characteristic graph-patterns mining, 2) graph packing of the cadastral landscape and 3) land use assignment.

This year, we formalized the simulation process by a formal problem of graph packing [51]. The process is illustrated by Figure 1 . Solving instances of the general graph packing problem has a high combinatorics and there does not exist any efficient algorithm to 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 the expert constraints on the land use. Constraints about the minimum surface of crops or about the impossibility of some crops colocation can be easily defined. This work has been presented at the conference RFIA [19] and we have been invited to provide an extended version to the Revue d'Intelligence Artificielle (RIA). The application results have been presented to the national colloquium on landscape modelling (<http://www.reseau-payote.fr/?q=colloque2014>).

In addition to the landscape simulation, the challenging tasks of solving the general graph packing with ASP raises interests in more general problem (such as graph compression). We initiated a collaboration with J. Nicolas (Inria/Dyliss) to improve the efficiency of our first programs.

### 6.2.4. 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 [44] 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. A first attempt have been proposed by Jarvisälo for simple settings [49].

This year, we worked on several classical pattern mining tasks: episodes, sequences and closed/maximal itemsets. In [20], we explore the use of ASP to extract frequent episodes (without parallel events) in a unique long sequence of itemsets. We especially evaluate the incremental resolution to improve the efficiency of our program. We next worked on sequence mining to extract pattern from the sequence of TV programs (V. Claveau, CNRS/LinkMedia). This tasks was simpler, but the computation time was significantly higher than dedicated algorithms. Nonetheless, our recent programs extracting closed or maximal patterns have better results.

### **6.2.5. Monitoring cattle**

Following the lines of a previous work [62], 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 analysis to devising multivariate decision rules for improving the specificity of detectors [15].

### **6.2.6. Subdimensional clustering for fast similarity search over time series data. Application to Information retrieval tasks**

Information retrieval and similarity search tasks in time series databases remains a challenge that requires to discover relevant pattern-sequences that are recurrent over the overall time series sequences, and to find temporal associations among these frequently occurring patterns. However, proposed methods suffer from a lack of flexibility of the used similarity measures, a lack of scalability of the representation model, and a penalizing runtime to retrieve the information. Motivated by these observations, we have designed a framework tackling the query by content problem on time series data, ensuring (i) fast response time, (ii) multi-level information representation, and (iii) representing temporal associations between extracted patterns. This year we have compared several distance measures on time series with different criteria and proposed a hybrid retrieval method based on pattern extraction and clustering [8].

### **6.2.7. Knowledge Extraction from Heterogeneous Data**

Recently, mining microarrays data has become a big challenge due to the growing sources of available data. We are using machine learning methods such as clustering, dimensionality reduction, association rules discovery on transcriptomic data, by combining a domain ontology as source of knowledge, in order to supervise the KDD process. Our objectives concern the identification of genes that could participate in the development of tumors. This year, we have introduced a new method for extracting enriched biological functions from transcriptomic databases using an integrative bi-classification approach based on formal concept analysis [7].

### **6.2.8. 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, and as been accepted in the DATE'15 conference [14].

## **6.3. Causal reasoning and argumentation**

**Participants:** Philippe Besnard, Marie-Odile Cordier, Yves Moinard.

We have continued our work on reasoning (precisely search for explanations) from causal relations and ontology [9]. Mainly, we have enforced the use of argumentation [32] in order to help choosing the best explanations among the (rather big) set of explanations given by our previous formalism. Then, we hope to be able to use the last versions of clingo in order to get an efficient tool to deal with complex situations (our example is Xynthia storm, february 2012 in western France for which there exists a huge amount of data from various official reports) by using clingo. For now we have a preliminary program which provides (besides the applications already mentioned: mining and landscape simulation) another application of the recent versions of ASP. One interest is that the nature of the rules in ASP should allow to translate rather directly (and hopefully efficiently) our previous formalism together with the improved argumentation part.

## EXMO Project-Team

# 6. New Results

## 6.1. Highlights of the Year

- Our work on link key extraction and evaluation (§6.3.4) has been published at ECAI 2014.
- Jérôme Euzenat has been elected fellow of the European Coordination Committee for Artificial Intelligence (ECCAI).

## 6.2. Ontology matching and alignments

We pursue our work on ontology matching and alignment support [8] [10] with contributions to evaluation and alignment semantics.

### 6.2.1. Evaluation

**Participant:** Jérôme Euzenat.

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 [2].

This year, we ran the OAEI 2014 evaluation campaign [15]. We used again our generator for generating new version of benchmarks. The Alignment API was used for manipulating alignments and evaluating results.

A novelty of this year was that data interlinking evaluation was using the SEALS platform and a new query-based evaluation track was created.

The participating systems and evaluation results were presented in the 9th Ontology Matching workshop, held in Riva de Garda, Italy. More information on OAEI can be found at <http://oaei.ontologymatching.org/>.

### 6.2.2. Algebras of alignment relations

**Participants:** Armen Inants [Correspondent], Jérôme Euzenat.

Qualitative calculus is the central concept 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. We tackle the problem of combining two or more calculi over disjoint universes into a single calculus. The problem is important, because in the ontology matching domain we deal with various kinds of ontological entities: concepts, individuals, properties. We define a new formalism called a heterogeneous qualitative calculus, based on an algebraic construct called Schröder category. A Schröder category is to binary relations over heterogeneous universes what a relation algebra is to homogeneous ones. We establish the connection between homogeneous and heterogeneous qualitative calculi by defining two mutually inverse transition operators. We provide an algorithm for combining two homogeneous calculi with different universes into a single calculus.

This work has vocation to support developments of the Alignment API towards relation algebras. It is part of the PhD of Armen Inants.

## 6.3. 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.

### 6.3.1. *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 is based taking advantage of multilingual references, such as BabelNet. We evaluated variation of these two settings on English (DBpedia) and Chinese (XLORE) datasets. Both approaches demonstrated promising results [20]. We will conduct more experiments including other language pairs and larger corpus.

This work is part of the PhD of Tatiana Lesnikova developed in the LINDICLE project (§7.1.2).

### 6.3.2. *Interactive learning of interlinking patterns*

**Participants:** Zhengjie Fan [Correspondent], Jérôme Euzenat.

We proposed an interlinking method which, from class correspondences between data source ontologies, uses  $k$ -means or  $k$ -medoids clustering to produce property correspondences. It then generates a first interlinking pattern which is a combination of a link key and similarity measures. Such patterns can be transformed into a SILK script for generating an initial link set. A sample of these links are assessed by users as either correct or incorrect. These are taken as positive and negative example by an extension of the disjunctive version space method to find an interlinking pattern, that can justify correct links and incorrect links. Experiments show that, with only 1% of sample links, this method reaches a F-measure over 96%. The F-measure quickly converges, being improved by nearly 10% than other comparable approaches [19].

This work is part of the PhD of Zhengjie Fan [4], co-supervised with François Scharffe (LIRMM), and developed in the DATALIFT project (§7.1.1).

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

**Participants:** Manuel Atencia Arcas [Correspondent], Mustafa Al-Bakri, Steffen Lalande, Marie-Christine Rousset.

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. It only imports data as specific as possible for inferring or contradicting target 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.

This work is part of the PhD thesis of Mustafa Al-Bakri, co-supervised by Manuel Atencia and Marie-Christine Rousset, developed in the QUALINCA project.

### 6.3.4. *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. Keys can be extracted from data by assuming that keys holding for specific data sets, may hold universally. We have extended such a classical key extraction technique for extracting weak link keys. We designed an algorithm to generate first a small set of candidate link keys and described this approach in the framework of formal context analysis [13]. Depending on whether some of the, valid or invalid, links are known, we defined supervised and non supervised measures for selecting the appropriate link keys. The supervised measures approximate precision and recall on a sample, while the non supervised measures are the ratio of pairs of entities a link key covers (coverage), and the ratio of entities from the same data set it identifies (discrimination). We have experimented these techniques, showing the accuracy and robustness of both approaches [12].

This work has been developed partly in the LINDICLE project (§7.1.2).

## 6.4. 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, ending up as a mesh of reticulated knowledge.

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.

### 6.4.1. Evolution of ontology networks and linked data

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

As link keys are obtained by statistical analysis of datasets (§6.3.4), they are both data-dependent and computation-intensive. Therefore, their recalculation should be avoided if possible. We are developing methods to analyse if changes performed in the data, necessarily require link key recomputation.

To reach this goal, we are developing an approach considering datasets as logical theories. In this context, changes that affect a link key are meta-logical operations. We adopt the framework of belief revision to define postulates that evolution operators must satisfy.

This work is part of the PhD thesis of Adam Sanchez Ayte developed in the LINDICLE project (§7.1.2).

### 6.4.2. Cultural alignment repair

**Participant:** Jérôme Euzenat [Correspondent].

Alignments between ontologies may be established through agents holding such ontologies attempting at communicating and taking appropriate action when communication fails. This approach, that we call cultural repair, has the advantage of not assuming that everything should be set correctly before trying to communicate and of being able to overcome failures. We tested this approach on alignment repair, i.e., the improvement of incorrect alignments. For that purpose, we performed a series of experiments in which agents react to mistakes in alignments. Agents only know about their ontologies and alignments with others and they act in a fully decentralised way. We showed that cultural repair is able to converge towards successful communication through improving the objective correctness of alignments. The obtained results are on par with a baseline of state-of-the-art alignment repair algorithms [7] [17]

The benchmarks, results and software are available at <http://lazylav.gforge.inria.fr>.

## GRAPHIK Project-Team

## 6. New Results

### 6.1. Highlights of the Year

- *Michael Thomazo was awarded the AFIA Prize 2014 (French Association for Artificial Intelligence) for his PhD entitled "Conjunctive Query Answering Under Existential Rules - Decidability, Complexity, and Algorithms" defended in October 2013. He was also awarded the first accessit of Gilles Kahn Prize 2014 by the SIF (French Society for Computer Science) [14].*
- *Madalina Croitoru and Alain Gutierrez were awarded the Best Technical Paper of SGAI-2014 for "On Ontological Expressivity and Modelling Argumentation Schemes using COGUI", in collaboration with Wael Hamdan, Rady Khazem and Ghaida Rebdawi .*
- *Abdallah Arioua was awarded the Best Student Paper Award of SGAI-2014 for "Query Failure Explanation in Inconsistent Knowledge Bases: A Dialogical Approach" in collaboration with Nouredine Tamani, Madalina Croitoru and Patrice Buche .*

BEST PAPERS AWARDS :

[36] **AI'2014: Thirty-fourth SGAI International Conference on Artificial Intelligence.** W. HAMDAN, R. KHAZEM, G. REBDAMI, M. CROITORU, A. GUTIERREZ, P. BUCHE.

[28] **AI'2014: 34th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence.** A. ARIOUA, N. TAMANI, M. CROITORU, P. BUCHE.

### 6.2. Ontology-Based Query Answering with Existential Rules

**Participants:** Jean-François Baget, Fabien Garreau, Mélanie König, Michel Leclère, Marie-Laure Mugnier, Swan Rocher, Federico Ulliana.

Ontology-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 [59], [58]. 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 5.2).

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 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.

### 6.2.1. Improvement of Query Rewriting Algorithms

These last two years, we designed and implemented a query rewriting algorithm that takes as input a set of existential rules and a UCQ  $q$  and outputs a UCQ, which is a sound and complete rewriting of  $q$ , whenever such a rewriting exists [60], [61], [62]. This year's main improvement to this algorithm is the implementation of a unifier able to process rules without decomposing their head into single atoms. This improvement appeared to have a very high impact on the efficiency of query rewriting (up to 274 quicker on an ontology where 32% of the rules have a head composed of two atoms instead of a single one). Beside, much effort has been devoted to experiments: to find appropriate benchmarks, to build a translator from the Semantic Web format OWL/OWL2 to our existential rule format dlgp (since most existing ontologies are available in OWL/OWL2 format), to select existing tools to compare with, run them, finally compare tools on several criteria.

- *Results partially published in the Semantic Web Journal [22].*

Query rewriting techniques have the interest of being independent from the data. However, a main bottleneck is that the size of the rewritten query can be exponential in the size of the original query, hence the produced rewriting maybe not usable in practice. A well-known source of combinatorial explosion are some very simple rules, which form the core of any ontology, typically expressing concept and relation hierarchies, concept properties and relation signatures. We have proposed a rewriting technique, which consists in compiling these rules into a preorder on atoms and embedding this preorder into the rewriting process. This allows to compute compact rewritings 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. The provided algorithm computes a sound, complete and minimal UCQ rewriting, if one exists. Experiments show that this technique leads to substantial gains in the query rewriting process, in terms of size and runtime, and scales on very large ontologies (several ten thousands of rules).

- *Results not published yet. Reported in Mélanie König's PhD thesis [17].*

### 6.2.2. A Better Approximation of Chase Termination for Existential Rules and their Extension to Non-monotonic Negation

Forward chaining with existential rules is known as the *chase* in databases. Various acyclicity notions ensuring chase termination have been proposed in the knowledge representation and databases. Acyclicity conditions found in the literature can be classified into two main families: the first one constrains the way existential variables are propagated during the chase and the second one constrains dependencies between rules *i.e.*, the fact that a rule may lead to trigger another rule. These conditions are based on different graphs, but all of them can be seen as forbidding "dangerous" cycles in the considered graph. We defined a new family of graphs that allows to unify and strictly generalize these acyclicity notions without increasing worst-case complexity.

Second, we considered the extension to existential rules with nonmonotonic negation under stable model semantics and further extended acyclicity results obtained in the positive case by exploiting negative information.

- *This work is part of Fabien Garreau and Swan Rocher's PhD theses. Results published at the European Conference on Artificial Intelligence (ECAI 2014)[30](long version as an arXiv report) and at the Workshop on Non-monotonic Reasoning (NMR 2014) [31].*

### 6.2.3. Detailed Results and Complements on Query Answering under Greedy Bounded-Treewidth Sets of Existential Rules

The family of greedy bounded-treewidth sets of existential rules (gbts) is an expressive class of rules for which entailment is decidable. This decidability property relies on a structural property of the saturation by the chase (*i.e.*, the set of inferred facts): for any initial set of facts, the saturation of these facts has a bounded treewidth (where the treewidth is computed on a graph associated with a set of atoms). Moreover, a tree decomposition of bounded width can be incrementally built during the chase. This family generalizes the important family of guarded existential rules, which itself generalizes Horn description logics used in OBDA.



In papers published at IJCAI 2011 and KR 2012, we studied the complexity of entailment under gbts rules as well as under known subclasses of gbts (with respect to data, combined and query complexity) and provided a generic algorithm with optimal worst-case complexity. This year, we finally completed a long report (75 pages) containing the detailed proofs of the results, some of them being very technical; in this report, we also clarified and reformulated the description of the generic algorithm, according to Michael Thomazo's PhD thesis (defended in October 2013); finally, we complemented the landscape of gbts classes by studying the complexity of all subclasses obtained by combining the syntactic criteria which define already known classes.

- *Results available as an arXiv report [56]. Submitted to a major journal in Artificial Intelligence. In collaboration with Sebastian Rudolph (TU Dresden) and Michael Thomazo (now postdoctoral student in Sebastian Rudolph's group).*

#### 6.2.4. Extracting Bounded-level Modules from Deductive RDF Triplestores

The Semantic Web is consolidating a legacy of well-established knowledge bases spanning from life sciences, to geographic data and encyclopedical repositories. Today, reusing knowledge and data available online is vital to ensure a coherent development of the Semantic Web, thereby capitalizing on the efforts made in the last years by many institutions and domain experts to publish quality information.

In this paper we studied how to extract modules from RDF knowledge bases equipped with Datalog inference rules, we called Deductive RDF Triplestores. A module is a Deductive RDF Triplestore entailed from the reference system, which is defined upon a restricted vocabulary (or signature). We proposed a new semantics for bounded-level modules allowing to control their size, and then presented extraction algorithms compliant with the novel semantics. This feature is helpful since many ontologies are extremely large, while users often need to reuse only a small part of resources in their work.

This work was partially carried out before the arrival of Federico Ulliana at GraphIK. For the future, we plan to study module extraction for knowledge bases equipped with existential rules, which extend the rules considered here.

- *Results published at the Twenty-Ninth AAI Conference on Artificial Intelligence (AAAI 15) [44]. In collaboration with Marie-Christine Rousset from LIG (University of Grenoble).*

#### 6.2.5. Axiomatisation of Consistent Query Answering via Belief Revision

This work takes place in the OBQA setting where a query is being asked over a set of knowledge bases defined over a common ontology. When the union of knowledge bases along with the ontology is inconsistent, several semantics have been defined which are tolerant to inconsistency. These semantics all rely on computing repairs, *i.e.*, maximal (in terms of set inclusion) consistent subsets of the data set. They have been studied from a productivity point of view and a complexity point of view. We take a new point of view to define axiomatic characterisations of two such semantics, namely IAR (Intersection of All Repairs) and ICR ((Intersection of Closed Repairs). We argue that such characterisation can provide an alternative way of comparing the semantics and new insights into their properties. Furthermore such axiomatisation can be used when proposing a generalisation of inconsistency tolerant semantics. In order to provide the axiomatic characterisations we define belief revision operators that correspond to IAR and ICR.

- *Work published at [43]. In collaboration with Ricardo Rodriguez from University of Buenos Aires.*

### 6.3. Reasoning with Imperfect Information and Priorities

**Participants:** Abdallah Arioua, Patrice Buche, Madalina Croitoru, Jérôme Fortin, Souhila Kaci, Namrata Patel, Tjitze Rienstra, Nouredine Tamani, Rallou Thomopoulos.

This work focuses on two main notions, namely *argumentation systems*, which allow to represent and deal with conflicting information, and formalisms to represent *preferences*, which allow to compare possible outcomes in decision making and recommender systems.



### 6.3.1. Fundamental Aspects of Argumentation

A Dung-style argumentation framework aims at representing conflicts among elements called arguments. The basic ingredients of this framework is a set of arguments and a Boolean abstract (*i.e.*, its origin is not known) binary defeat relation on these arguments. This abstract framework can be instantiated in different ways, by representing arguments in a given knowledge representation formalism, which allows to take the semantics of arguments into account in the computation of the defeat relation.

Preference-based argumentation frameworks are instantiations of Dung's framework in which the defeat relation is derived from an attack relation and a preference relation over the arguments. Recently, Dung's framework has been extended in order to consider the strength of the defeat relation *i.e.*, to quantify the degree to which an argument defeats another argument. In this work, we instantiated this extended framework by a preference-based argumentation framework with a valued preference relation. As particular cases, the latter can be derived from a weight function over the arguments or a Boolean preference relation. We showed under some reasonable conditions that there are "less situations" in which a defense between arguments holds with a valued preference relation compared to a Boolean preference relation. Finally, we provided some conditions that the valued preference relation shall satisfy when it is derived from a weight function.

- *This is a joint work with Christophe Labreuche from Thales and published in [20]*

We also considered an extension to argumentative frameworks based on fuzzy set theory. The knowledge base is fuzzified to allow agents expressing their expertise (facts and rules) attached with grades of importance in the unit interval. Arguments are then attached with a strength score aggregating the importance expressed on their facts and rules. Extensions, corresponding to subsets of consistent arguments, are also attached with forces computed based on their strong arguments. The forces are used then to rank extensions from the strongest to the weakest one, upon which decisions can be made. We have also shown that the strength preference relation defined over arguments is reasonable according to classical rationality postulates and our fuzzy logical argumentation system can be seen as a computationally efficient instantiation of the generic model of structured argumentation framework. We furthered our theoretical research and demonstrate the added value of our approach in the practical setting of the European project EcoBioCap (see Sect.8.2 ).

- *Work published in IPMU 2014 [47] and Fuzz. IEEE 2014 [46].*

One instantiation, among many others, of Dung's framework consists in constructing the arguments from a set of propositional logic formulas. Thus an argument is seen as a reason for or against the truth of a particular statement. Despite its advantages, the argumentation approach for inconsistency handling also has important shortcomings. More precisely, in some applications what one is interested in are not so much only the conclusions supported by the arguments but also the precise explanations of such conclusions. We showed that argumentation framework applied to classical logic formulas is not suitable to deal with this problem. On the other hand, intuitionistic logic appears to be a natural alternative candidate logic (instead of classical logic) to instantiate Dung's framework. We developed *constructive argumentation framework*. We showed that intuitionistic logic offers nice and desirable properties of the arguments. We also provided a characterization of the arguments in this setting in terms of minimal inconsistent subsets when intuitionistic logic is embedded in the modal logic S4.

- *This is a joint work with Yakoub Salhi from CRIL and published in [39]*

Lastly, we developed a model of abduction in abstract argumentation, where changes to an argumentation framework act as hypotheses to explain the support of an observation. We presented dialogical proof theories for the main decision problems (*i.e.*, finding hypotheses that explain skeptical/credulous support) and we showed that our model can be instantiated on the basis of abductive logic programs.

- *This work has been done in Tjitze Rienstra's thesis and published in [32].*

### 6.3.2. Use of Argumentation in Explanation, Querying and Decision Making

Besides work on the foundations of argumentation frameworks, we have studied the use of argumentation techniques in various tasks: explanation of query failure, reverse engineering, and decision making. These studies are mainly motivated by agri-food scenarios: bread conception, packaging conception, and durum wheat conception.

We have proposed an argumentation-based explanation for *query failure explanation* under the inconsistency tolerant semantics ICR in an Ontology-Based Data Access setting with existential rules. We used a rule-based language and we base our work on the equivalence between ICR-based query answering in inconsistent knowledge bases and sceptical acceptance of arguments. We proposed a multilevel explanation that exploits both the inference power of the logical language as well as arguments of dialectical nature. We also investigated an interactive argumentative approach where the process of explanation takes the form of a dialogue between the user and the reasoner.

- *Work published in COMMA 2014 [27] and SGAI 2014 [28] where it received the best student paper award.*

Within the framework of the European project EcoBioCap <http://www.ecobiocap.eu> about the design of next generation packagings using advanced composite structures based on constituents derived from the food industry, we have been developing a Decision Support System (DSS) for packaging material selection. [40], [49]. The DSS consists of two steps: (1) aggregating possibly conflicting needs expressed by several parties involved in the considered field and (2) querying a database of packagings with the resulting aggregation obtained at point (1). We instantiate for each need, called viewpoint or aspect, an argumentation system to reason about arguments solely expressed on it [45]. This will then be used to generate the query on the packaging database. To this aim we show how to instantiate ASPIC with the DLR-Lite logic modeling expert ontologies in this real world scenario [47].

- *Work published in AAMAS 2014 [45], IPMU 2014 [47], ICCS 2014 [40], and COMMA 2014 [49].*

Evaluating food quality is a complex process since it relies on numerous criteria historically grouped into four main types: nutritional, sensorial, practical and hygienic qualities. They may be completed by other emerging preoccupations such as the environmental impact, economic phenomena, etc. However, all these aspects of quality and their various components are not always compatible and their simultaneous improvement is a problem that sometimes has no obvious solution, which corresponds to a real issue for decision making. We propose a decision support method guided by the objectives defined for the end products of an agrifood chain. It is materialized by a backward chaining approach based on argumentation [47]. An extended version of this paper reporting on experimental results and expert evaluation has been published in Ecological Informatics [24].

- *Work published in IPMU 2014 [47], and Ecological Informatics 2014 [24].*

Knowledge elicitation, representation and reasoning explanation by / to non-computing experts has always been considered as a crafty task due to difficulty of expressing logical statements by non-logicians. We use the COGUI editor in order to elicit and represent argumentation schemes expressed using existential rules within an inconsistent knowledge base. COGUI is a visual, graph based knowledge representation editor compatible with main Semantic Web languages (see Section 5.1). COGUI allows for default reasoning on top of ontologies. We investigate its use for modelling and reasoning using argumentation schemes and discuss the advantages of such representation. We show how this approach can be useful in the practical setting of EcoBioCap where the different argumentation schemes can be used to lead reasoning.

- *Work published in SGAI 2014 [36] where it received the best technical paper award. In collaboration with Wael Hamdan, Rady Khazem and Ghaisa Rebdawi from the Higher Institute of Applied Science and Technology (HIAST), Syria.*

### 6.3.3. Preferences

Qualitative and comparative preference statements of the form “prefer  $\alpha$  to  $\beta$ ” are useful components of many applications. This statement leads to the comparison of two sets of alternatives: the set of alternatives in which  $\alpha$  is true and the set of alternatives in which  $\beta$  is true. Different ways are possible to compare two sets of objects leading to what is commonly known as preference semantics. The choice of the semantics to employ is important as they differently rank-order alternatives. Existing semantics are based on philosophical and non-monotonic reasoning grounds. In the meanwhile, they have been widely and mainly investigated by AI researchers from algorithmic point of view. We came to this problem from a new angle and completed existing

theoretical investigations of the semantics. In particular, we provided a comparison of the semantics on the basis of their psychological plausibility by evaluating their closeness to human behavior.

- *This is a joint work with Eric Raufaste from CLLE and published in [38]*

There has been a growing interest in the study of preferences for their utility in solving problems related with decision making. Most of the preference representation languages developed in the literature are based on comparative preference statements since they offer a simple and intuitive way for expressing preferences. They can be further interpreted following different semantics, imparting a greater flexibility on how outcomes can be compared. So far the main objective has been to rank-order the set of outcomes given a set of comparative preference statements and one or several semantics. Tackling this problem from a different angle, we looked into the behavioral aspects of the preference semantics and statements by attempting to formalise the intuition behind them using postulates studied in preference logics and non-monotonic reasoning. We selected the postulates w.r.t. three criteria: coherence, syntax independence and inference. Thus, our analysis provided a means to determine those properties that are satisfied for a given preference semantics.

- *This work has been done in Namrata Patel's thesis and published in [21]*

Intelligent 'services' are increasingly used on e-commerce platforms to provide assistance to customers. Numerous preference elicitation methods developed in the literature are now employed for this purpose. However, it is commonly known that there is a real bottleneck in preference handling as concerns the elicitation of preferences because it does not cater to the wide range of preference representation languages available. Thus, as a first step in developing a decision-support tool using an AI based on such languages, this paper describes a preliminary study conducted to address this issue. We proposed a method of eliciting real-time user preferences expressed in natural language (NL) which can be formally represented using comparative preference statements complying with different semantics, and provided a proof of concept to demonstrate its feasibility. Since we developed NL resources to detect preference semantics, we also made a comparative study with existing resources to underline the peculiarities of our model.

- *This work has been done in Namrata Patel's thesis and published in [37]*

## 6.4. Semantic Data Integration

**Participants:** Michel Chein, Madalina Croitoru, Léa Guizol, Michel Leclère, Rallou Thomopoulos.

It often happens that different references (*i.e.*, data descriptions), possibly coming from heterogeneous data sources, concern the same real world entity. In such cases, it is necessary: (i) to detect whether different data descriptions really refer to the same real world entity and (ii) to fuse them into a unique representation. This issue has been studied under various names: "record linking", "entity resolution", "reference resolution", "de-duplication", "object identification", "data reconciliation", etc., mostly in databases. It has become one of the major challenges in the Web of Data, where the objective is to link data published on the web and to process them as a single distributed database.

We investigate this problem in the specific context of bibliographic databases. Indeed, people working in bibliographical information systems have a lasting tradition of using norms and have integrated, along collections of documents notices (e.g. bibliographic records), collections of authority notices that categorize the different named entities used to describe documents (people, organizations, places, ...). In current databases, documents notices do not use directly the names of named entities to fill a particular field (author, editor, ...), but the unique identifier of the authority notice representing that named entity.

A few years ago, we began a collaboration with ABES (National Bibliographic Agency for Universities) to develop a method and a prototype to perform entity resolution between on one hand the authors of a new bibliographic record, and, on the other the authority references of an authority catalog (and namely the Sudoc catalog from the ABES agency). A problem with this approach is that it relies upon pre-established links between bibliographic records and authority notices. However, our experimentation and evaluation have shown that many existing links were erroneous, and thus led to the propagation of new linkage errors. We have thus began to work on methods and tools to repair linkage errors in bibliographical databases. The first step of our approach was to build a knowledge-base over an ontology (based on the international standards FRBR and CIDOC-CRM) aiming at representing bibliographic data (an RDFS base) as well as librarian knowledge.

From that, we developed a methodological framework allowing to design rules concluding on the coreference or the difference between entities of the bibliographic knowledge base. This framework was implemented in Cogui.

#### 6.4.1. An Original Methodology to Compute Coreference and Difference Links

Our methodology can be briefly summarized as follows. The first step consists in computing “sure” links. In the second step, authority notices are enriched by information that comes from bibliographic notices to which they are linked by sure links. In the third step, Datalog rules that conclude on coreference or difference are triggered. The results are used to compute new sure links. These steps are iterated until stability *i.e.*, no new sure link is discovered. More specifically, the Datalog rules are the following form. The body of a rule is a conjunction of similarity criteria on attributes and its head states the coreference or the difference of two individual entities with a specific confidence level (represented as a symbolic value). We are currently instantiating this methodology for the Sudoc catalog, jointly with the ABES librarians, which will allow them to evaluate it.

#### 6.4.2. Partitioning Semantics for Link Discovery in Bibliographic Knowledge Bases

With the aim of evaluating and improving the quality of links in bibliographical knowledge bases, we have developed a decision support system based on partitioning semantics. The novelty of our approach consists in using symbolic values criteria for partitioning and suitable partitioning semantics. We have investigated the limits of those partitioning semantics: how the characteristics of the input (objects and criteria) influences characteristics of the result, namely correctness of the result and execution time. We have also evaluated and compared the above mentioned semantics on a real qualitative sample. This sample is issued from the catalogue of French university libraries (SUDOC) maintained by ABES.

- *This work is part of Lea Guizol's PhD thesis [16]. Work published in Fuzz IEEE 2014 [46].*

#### 6.4.3. Key Discovery on the Semantic Web

Many techniques were recently proposed to automate the linkage of RDF datasets. Predicate selection is the step of the linkage process that consists in selecting the smallest set of relevant predicates needed to enable instance comparison. We call keys this set of predicates that is analogous to the notion of keys in relational databases. We have formally explained the different assumptions behind two existing key semantics (IC), and have evaluated experimentally these keys semantics by studying how discovered keys could help dataset interlinking or cleaning.

- *Work published in IC 2014 [50] and ICCS 2014 [29] in collaboration with Manuel Atencia and Jerome David from LIG, and Nathalie Pernelle, Fatiha Sais and Danai Symeonidou from LRI. See also the reconciliation-based approach in[23].*

#### 6.4.4. Fusion of Linked Data

The problem of data fusion starts from reconciled datasets, whose objects are linked with semantic sameAs relations, as described above. We attempt to merge the often conflicting information of these reconciled objects in order to obtain unified representations that only contain the best quality information. We are studying an approach to determine the most appropriate value(s). Our method combines different quality criteria based on the value and its data source, and exploits, whenever possible, the ontology semantics, constraints and relations. Moreover we create a mechanism to provide explanations about the quality of each value, as estimated by our system. To achieve this, we generate annotations used for traceability and explanation purposes.

- *Work described in the Qualinca deliverable 4.2 research report, and accepted for publication in EGC'2015 : "Linked Data Annotation and Fusion driven by Data Quality Evaluation" (authors: Ioanna Giannopoulou, Fatiha Sais from LRI, and Rallou Thomopoulos)*

## LINKS Team

# 5. New Results

## 5.1. Highlights of the Year

In the objective Querying Heterogeneous Linked Data, Slawomir Staworko and Iovka Boneva have developed new ways to define schema for Graph Database and RDF [19]. This work has been influencing a group work of W3C on defining a schema for the DF format. This work is a continuation of [3] (by Iovka Boneva, Radu Ciucanu and Slawomir Staworko) developing a new schema for unordered trees over XML. Due to these works, Boneva is now a member of the Data Shapes Working Group which mission is to produce a language for defining structural constraints on RDF graphs. <http://www.w3.org/2014/data-shapes/charter>

In the objective Managing Dynamic Linked Data, the main break through is the development of QuixPath that now covers 100 per cent of the XPathMark, a W3C benchmark for the language Xpath (querying XML trees). In particular, it includes aggregation operators, joins and arithmetics operations. The core of QuixPath is based on techniques presented in [6] (by Tom Sebastian, Denis Debardieux and Joachim Niehren).

In the objective Linking Data Graphs, different methods have been developed to learn queries over graph. More precisely, the queries learned are conjunctive queries with joins. These techniques have been presented in [13] and demonstrated in [4] at the conference VLDB.

## 5.2. Querying Heterogeneous Linked Data

Angela Bonifati, Gianvito Summa, Esther Pacitt (U Montpellier 2)i and Fady Draidí (U Montpellier 2) [5] consider peer-to-peer data management systems (PDMS), where each peer maintains mappings between its schema and some acquaintances, along with social links with peer friends. In this context, the goal is reformulating conjunctive queries from a peer's schema into other peer's schemas. Precisely, queries against a peer node are rewritten into queries against other nodes using schema mappings thus obtaining query rewritings. They propose a new notion of 'relevance' of a query with respect to a mapping that encompasses both a local relevance (the relevance of the query w.r.t. the mapping) and a global relevance (the relevance of the query w.r.t. the entire network). Based on this notion, they conceived a new query reformulation approach for social PDMS which achieves great accuracy and flexibility. This has been implemented and experimented in a prototype.

Pierre Bourhis, Andreas Morak and Andreas Pieris [14] investigated classes of queries for which the problem of open query answering of disjunctive guarded TGDs a decent complexity (e, g exptime). The complete picture of the complexity of answering (unions of) conjunctive queries under the main guarded-based classes of disjunctive existential rules has been recently settled. It has been shown that the problem is very hard, namely 2ExpTime-complete, even for fixed sets of rules expressed in lightweight formalisms. The central objective of the present paper is to understand whether simpler query languages (bounded tree width and acyclic queries) have a positive impact on the complexity of query answering under the main guarded-based classes of disjunctive existential rules.

In [3], a new formalism for schema for unordered trees have been developed. It is based on a notion of regular expressions of multisets of labels. Different problems of static analysis like emptiness and containment are studied and their complexity. Different simpler schema are studied leading to interesting complexity for the different studied problems. Finally, they study the expressive power of the proposed schema languages and compare them with yardstick languages of unordered trees (FO, MSO, and Presburger constraints) and DTDs under commutative closure. The results show that the proposed schema languages are capable of expressing many practical languages of unordered trees and enjoy desirable computational properties.

In [7], Adrian Boiret, Vincent Hugot and Joachim Niehren and Ralf Treinen (University Paris 7) proposes a notion deterministic tree automata for unordered trees. While the existing notions are well-investigated concerning expressiveness, they all lack a proper notion of determinism, which makes it difficult to distinguish subclasses of automata for which problems such as inclusion, equivalence, and minimization can be solved efficiently. In this paper, the authors propose and investigate different notions of "horizontal determinism", starting from automata for unranked trees in which the horizontal evaluation is performed by finite state automata.

### **5.3. Managing Dynamic Linked Data**

Tom Sebastian, Denis Debarbieux, Olivier Gauwin (U Bordeaux), Joachim Niehren, Mohamed Zergaoui (Innovimax) [6] present new techniques to evaluate XPath queries on trees received in a streaming way. It introduces early nested word automata in order to approximate earliest query answering algorithms for nested word automata. The notion early query answering algorithm is based on stack-and-state sharing for running early nested word automata on all answer candidates with on-the-fly determinization. These techniques allow to implement a more important part of XPath and outperform all the previous tools in coverage of XPathMark benchmark.

### **5.4. Linking Data Graphs**

Angela Bonifati, Radu Ciucanu, Slawomir Staworko developed techniques to learn conjunctive queries from example given by a user. The main part is to infer joins between relations from the positive and negative tuples. Different techniques to deduce informative examples are presented and interestingly they can be done in polynomial time. The techniques are published in [13] and demonstrated in [4].

Grégoire Laurence, Aurélien Lemay, Joachim Niehren, Slawek Staworko, Marc Tommasi [16] explain how to learn sequential top-down tree-to-word transducers (STWs). First, they present a Myhill-Nerode characterization of the corresponding class of sequential tree-to-word transformations (STW). Next, they investigate what learning of stws means, identify fundamental obstacles, and propose a learning model with abstain. Finally, they present a polynomial learning algorithm



## MAGNET Team

# 6. New Results

## 6.1. Highlights of the Year

We developed a new framework for high order learning [4].

We have illustrated the usefulness of automatically annotated examples in complex learning supervised by few training examples [2], [1].

We propose a new algorithm for semi-supervised spectral clustering and apply it to the NLP task of noun phrase coreference resolution [6].

## 6.2. Higher-order Learning with Graphs

Along the thesis of THOMAS RICATTE, in [4] and [8], we propose methods for learning from interactions between groups in networks. We propose a proper extension of graphs, called hypernode graphs as a formal tool able to model group interactions. A hypernode graph is a collection of weighted relations between two groups of distinct nodes. Weights quantify the individual participation of nodes to a given relation. We define Laplacians and kernels for hypernode graphs and prove that they strictly generalize over graph kernels and hypergraph kernels. We prove that hypernode graphs correspond to signed graphs such that the matrix  $D - W$  is positive semidefinite. As a consequence, homophilic relations between groups may lead to non homophilic relations between individuals. We define the notion of connected hypernode graphs and a resistance distance for connected hypernode graphs. We propose spectral learning algorithms on hypernode graphs allowing to infer node ratings or node labelings. As a proof of concept, we model multiple players games with hypernode graphs and we define skill rating algorithms competitive with specialized algorithms.

## 6.3. Natural Language Processing

In [6] (presented by DAVID CHATEL at the ECML-PKDD and CAP'2014 conferences) we propose a new algorithm for semi-supervised spectral clustering and apply it to the task of noun phrase coreference resolution. The main insight is in the inclusion of pairwise constraints into spectral clustering: our algorithm learns a new representation space for the data together with a distance in this new space. The representation space is obtained through a constraint-driven linear transformation of a spectral embedding of the data, and constraints are expressed with a Gaussian function that locally reweights the similarities in the projected space. A global, non-convex optimization objective is then derived and the model is learned via gradient descent techniques. Our algorithm is evaluated on the CoNLL-2012 coreference resolution shared task dataset, and shows some encouraging results.

In [2] and [1], we develop a new approach for the automatic identification of so-called implicit discourse relations. Specifically, our system combines hand labeled examples and automatically annotated examples based on explicit relations using several simple methods inspired by work in domain adaptation. Our system is evaluated empirically on the Annodis corpus, a French corpus annotated with discourse structures. Our system yields significant performance gains compared to only using hand-labeled data or using only automatically annotated data.

## 6.4. Ongoing work

### 6.4.1. Adaptive Graph Construction

We worked on developing a new algorithm in order to construct a graph in a adaptive way for a specific task. More precisely, we looked for a metric learning algorithm that could depend on the target task. Previous works on metric learning ([12]) aim at learning a relevant metric using a linear approach, which cannot capture the non-linearity of the data. Our approach, instead, aims at learning the most appropriate non-linear data projection for the target task. For this purpose, we train a neural network with relative constraints depending on the target task and a target classic metric (e.g. euclidean distance, cosine similarity, ...), in order to make the metric meaningful for the new data representation and our target task.

### 6.4.2. Correlation Clustering and Similarity/Dissimilarity Links

From a mathematical point of view, signed networks are graphs whose edges carry a sign representing the positive or negative nature of the relationship between the incident nodes. These structures are extremely useful for modeling, at the same time, similarity and dissimilarity object relationships. Given an undirected signed graphs, in the Correlation Clustering problem the goal is to find a node partition into clusters minimizing the number of negative (dissimilarity) edges linking two nodes within the same cluster and the number of positive (similarity) edges between different clusters.

We focused on devising an algorithm able to solve the Correlation Clustering problem for general input signed graphs (if the input is a complete signed graph the problem is proven to be much easier). One of the main objective of this work is the use of the proposed algorithm for creating a learner able to predict the unknown edges signs of a given signed graph. This prediction task is known as Link Classification in signed graphs. In fact, given an undirected signed graph whose edge set is split into training and test set, we could use the Correlation Clustering solution working for general input graphs for partitioning the training set and using the node partition generated for predicting the test edge signs. Moreover, one could exploit such an algorithm for developing new strategies for the Link Classification problem operating within the online and active Machine Learning setting.

Since the node set partitioning turns out to be strictly related to the Link Classification problem, we also focused on the very challenging goal of obtaining a deep understanding of the complex interplay between Link and Node Classification. More precisely, we investigated the relationships between the Vapnik Chervonenkis dimension of any given set of hypothesis space of node and edge similarity functions operating within this framework.

### 6.4.3. Ranking from Pairwise Sets of User Preferences

Given a set of objects (vertices of a graph) and a set of pairwise preference labels between objects (directed edges connecting vertices) which may be non-transitive due to irrationality or arbitrary noise, what is a correct way to sample preference labels for ordering the set of objects? This long standing open problem is, as far as we know, unsolved when each pairwise preference labels refers to two (disjoint) sets of objects (vertices). This framework can be easily motivated considering that quite often, in many real world contexts, users express their preferences between sets of items rather than single items, and turns out to be strictly connected with our recent model of hypergraphs with bipartite hyperedges [4]. We are working on devising a new algorithm able to rank a given set of items (graph node set) when only comparisons between sets containing at least 2 items are allowed. This challenging and interested problem is, as far as we know, quite novel and can be studied within different Machine Learning setting (online, batch, active, ...). The preliminaries results we are obtaining, when setting the cardinality of the item sets equal to 2, are encouraging and indicate that it could be possible to extend our strategies in order to deal with larger item sets.

## 6.5. Other results

In this section we provide the results we obtained that are not related with our main research directions.



In [3] we study the problem of learning sequential top-down tree-to-word transducers (STWs). First, we present a Myhill-Nerode characterization of the corresponding class of sequential tree-to-word transformations (STW). Next, we investigate what learning of stws means, identify fundamental obstacles, and propose a learning model with abstain. Finally, we present a polynomial learning algorithm.

## MAIA Project-Team

## 6. New Results

### 6.1. Highlights of the Year

- Two Research Fellow have been recruited with a focus on Service Robotics: Serena Ivaldi (CR2) and Francis Colas (CR1).
- The paper entitled : Exploiting Separability in Multiagent Planning with Continuous-State MDPs Jilles Dibangoye, Christopher Amato, Olivier Buffet, François Charpillat won the best paper award at AAMAS'2014, the international conference on autonomous agents and multi-agents.
- Jilles Dibangoye got an Assistant Professor position at INSA Lyon.

BEST PAPER AWARD :

[12] **13th International Conference on Autonomous Agents and Multiagent Systems**. J. S. DIBANGOYE, C. AMATO, O. BUFFET, F. CHARPILLET.

### 6.2. Decision Making

#### 6.2.1. Complexity Analysis of Exact Dynamic Programming Algorithms for MDPs

**Participant:** Bruno Scherrer.

*Eugene Feinberg and Jefferson Huang are external collaborators from Stony Brooks University.*

Following last year's work on the strong polynomiality of Policy Iteration, we show that the number of arithmetic operations required by any member of a broad class of optimistic policy iteration algorithms to solve a deterministic discounted dynamic programming problem with three states and four actions may grow arbitrarily. Therefore any such algorithm is not strongly polynomial. In particular, the modified policy iteration and  $\lambda$ -policy iteration algorithms are not strongly polynomial. This work was published in the *Operations Research Letters* [4].

#### 6.2.2. Analysis of Approximate Dynamic Programming Algorithms for MDPs

**Participants:** Bruno Scherrer, Manel Tagorti.

*Mathieu Geist is an external collaborator from Supélec.*

In [40], we consider LSTD( $\lambda$ ), the least-squares temporal-difference algorithm with eligibility traces algorithm proposed by Boyan (2002). It computes a linear approximation of the value function of a fixed policy in a large Markov Decision Process. Under a  $\beta$ -mixing assumption, we derive, for any value of  $\lambda \in (0, 1)$ , a high-probability estimate of the rate of convergence of this algorithm to its limit. We deduce a high-probability bound on the error of this algorithm, that extends (and slightly improves) that derived by Lazaric et al. (2012) in the specific case where  $\lambda = 0$ . In particular, our analysis sheds some light on the choice of  $\lambda$  with respect to the quality of the chosen linear space and the number of samples, that complies with simulations. This work was presented at the National JFPDA conference [34].

In the context of infinite-horizon discounted optimal control problem formalized by Markov Decision Processes, we focus on several approximate variations of the Policy Iteration algorithm: Approximate Policy Iteration (API), Conservative Policy Iteration (CPI), a natural adaptation of the Policy Search by Dynamic Programming algorithm to the infinite-horizon case (PSDP), and the recently proposed Non-Stationary Policy Iteration (NSPI). For all algorithms, we describe performance bounds with respect to the per-iteration error  $\epsilon$ , and make a comparison by paying a particular attention to the concentrability constants involved, the number of iterations and the memory required. Our analysis highlights the following points: 1) The performance guarantee of CPI can be arbitrarily better than that of API, but this comes at the cost of a relative—exponential in  $\frac{1}{\epsilon}$ —increase of the number of iterations. 2) PSDP $_{\infty}$  enjoys the best of both worlds: its performance guarantee is similar to that of CPI, but within a number of iterations similar to that of API. 3) Contrary to API that requires a constant memory, the memory needed by CPI and PSDP is proportional to their number of iterations, which may be problematic when the discount factor  $\gamma$  is close to 1 or the approximation error  $\epsilon$  is close to 0; we show that the NSPI algorithm allows to make an overall trade-off between memory and performance. Simulations with these schemes confirm our analysis. This work was presented at this year’s international conference on Machine Learning (ICML) [28].

Finally, we consider Local Policy Search, that is a popular reinforcement learning approach for handling large state spaces. Formally, it searches locally in a parameterized policy space in order to maximize the associated value function averaged over some predefined distribution. The best one can hope in general from such an approach is to get a local optimum of this criterion. The first contribution of this article is the following surprising result: if the policy space is convex, *any* (approximate) *local optimum* enjoys a *global performance guarantee*. Unfortunately, the *convexity* assumption is strong: it is not satisfied by commonly used parameterizations and designing a parameterization that induces this property seems hard. A natural solution to alleviate this issue consists in deriving an algorithm that solves the local policy search problem using a boosting approach (constrained to the convex hull of the policy space). The resulting algorithm turns out to be a slight generalization of conservative policy iteration; thus, our second contribution is to highlight an original connection between local policy search and approximate dynamic programming. This work was presented at this year’s European conference on Machine Learning (ECML) [27].

### 6.2.3. Adaptive Management with POMDPs

**Participants:** Olivier Buffet, Jilles Dibangoye.

*Samuel Nicol and Iadine Chadès (CSIRO) are external collaborators.*

In the field of conservation biology, adaptive management is about managing a system, e.g., performing actions so as to protect some endangered species, while learning how it behaves. This is a typical reinforcement learning task that could for example be addressed through Bayesian Reinforcement Learning.

During Samuel Nicol’s visit, the main problem we have studied is how to manage company inspections to deter these companies from adopting dangerous behaviors. This was modeled as a particular Stackelberg game, where  $N$  companies benefit from acting badly as long as they are not caught by inspections, and where 1 government agency has to decide which companies to inspect given a limited budget. The expected result is a stochastic strategy (randomly deciding which companies to inspect, with probabilities that depend on the benefits/losses of both types of players). We are working on exploiting particular features of this computationally complex problem to make it more tractable.

### 6.2.4. Solving decentralized stochastic control problems as continuous-state MDPs

**Participants:** Jilles Dibangoye, Olivier Buffet, François Charpillet.

*External collaborators: Christopher Amato (MIT).*

Decentralized partially observable Markov decision processes (DEC-POMDPs) are rich models for cooperative decision-making under uncertainty, but are often intractable to solve optimally (NEXP-complete), even using efficient heuristic search algorithms.

State-of-the-art approaches relied on turning a Dec-POMDP into an equivalent deterministic MDP—whose actions at time  $t$  correspond to a vector containing one decision rules (instantaneous policy) per agent—typically solved using a heuristic search algorithm inspired by A\*. In recent work (IJCAI’13), we have identified a sufficient statistic of this MDP—an *occupancy state*, i.e., a probability distribution over possible states and joint histories of the agents—and demonstrated that the value function was piecewise-linear and convex with respect to this statistic. This brings us in the same situation as POMDPs, allowing to generalize the value function from one occupancy state to another and to propose much faster algorithms (also using efficient compression methods).

This year, we have further progressed on this line of research.

- A journal paper has been submitted that presents the “occupancy MDP” approach in details.
- In the case of Network-Distributed POMDPs, a particular setting where the relations between agents follow a fixed network topology, we have shown that the value function could be decomposed additively with one value function per neighborhood. This work has been presented at AAMAS’2014 [12], receiving the conference’s best paper award.
- To further scale up the resolution of Dec-POMDPs, we have proposed multiple approximations techniques that can be combined and allow controlling error bounds. This work has been presented at ECML’2014 [13].

### 6.2.5. Learning Bad Actions

**Participant:** Olivier Buffet.

*Jörg Hoffmann, former member of MAIA, Michal Krajňanský (Saarland University), and Alan Fern (Oregon State University) are external collaborators.*

In classical planning, a key problem is to exploit heuristic knowledge to efficiently guide the search for a sequence of actions leading to a goal state.

In some settings, one may have the opportunity to solve multiple small instances of a problem before solving larger instances, e.g., trying to handle a logistics problem with small numbers of trucks, depots and items before moving to (much) larger numbers. Then, the small instances may allow to extract knowledge that could be reused when facing larger instances. Previous work shows that it is difficult to directly learn rules specifying which action to pick in a given situation. Instead, we look for rules telling which actions should not be considered, so as to reduce the search space. But this approach requires considering multiple questions: What are examples of bad (or non-bad) actions? How to obtain them? Which learning algorithm to use?

A first algorithm (with variants) has been proposed that learns rules for detecting (supposedly) bad actions. It has been empirically evaluated, providing encouraging results, but also showing that different variants will perform best in different settings. This algorithm has been presented at ECAI’2014 [24], and has participated in the learning track of the international planning competition in 2014 (<http://ipc.icaps-conference.org/>).

## 6.3. Ambient Intelligence And Robotic Systems

### 6.3.1. Adaptation of autonomous vehicle traffic to perturbations

**Participants:** Mohamed Tlig, Olivier Buffet.

*Olivier Simonin, a former member of the MAIA team, is an external collaborator from INSA-Lyon.*

The aim of the European project InTraDE is to propose more efficient ways to handle containers in seaports through the use of IAVs (Intelligent Autonomous Vehicles).

In his PhD thesis, Mohamed Tlig considers the displacements of numerous such IAVs whose routes are a priori planned by a supervisor. However, in such a large and complex system, different unexpected events can arise and degrade the traffic: failure of a vehicle, human mistake while driving, obstacle on roads, local re-planning, and so on.

In 2013, we have started looking at improving vehicle flows in complete road networks. In particular, we have proposed an approach that allows multiple flows of vehicles to cross an intersection without stopping, allowing to reduce delays as well as energy consumption. This has led to a publication in ICALT-14 [30], with more details in a research report [41].

This year, we have made a further step by coordinating the controller agents located in each of the network's intersections. More precisely, they are constrained to let the vehicles alternate at the same frequency —at the expense of potentially reducing the maximum flow of some roads— and a distributed algorithm offsets these “signals” so as to optimize either the energy consumption, or the time spent in the network. This tends to induce “green waves” wherever possible, i.e., to prevent vehicles from having to slow down before a traffic light. This work has been presented at ECAI-14 [31].

### 6.3.2. *Platooning: safe and precise virtual hooking mechanism or automated vehicles*

**Participants:** Jano Yazbeck, Alexis Scheuer, François Charpillat.

Among the several goals that we were trying to achieve in InTraDE, we were interested in platooning too. In her PhD thesis, Jano Yazbeck considers Platooning as a technique that aims at steering, safely and precisely, a train of vehicles along a path generated by a leader which can be driven by a human. Thus the trajectory is unknown to the followers. Platooning is considered in this project in order to move containers efficiently from the discharge zones of ships to the storage areas.

To obtain a safe and precise platooning, we aim at controlling the longitudinal and lateral behaviors of each vehicle of the platooning. On the one hand, the longitudinal controller computes a longitudinal velocity (or acceleration) which avoids collisions between vehicles by maintaining a safe inter-distance between each couple of successive vehicles. On the other hand, the lateral controller computes an angular velocity or a steering angle so that the vehicle follows precisely the leader's path. These two controllers can be decoupled and computed separately when the convoy moves at a low velocity.

This year, we proposed a platooning algorithm based on a near-to-near decentralized approach which has been published at ICRA 2014 [32]. In this approach, each vehicle estimates and memorizes on-line the path of its predecessor as a set of points. After choosing a suitable position to aim for, the follower estimates on-line the predecessor's path curvature around the selected target. Then, based on a heuristic search, it computes an angular velocity using the estimated curvature. The optimization criteria used in this work allows the robot to follow its predecessor's path without oscillation while reducing the lateral and angular errors.

In October, Jano Yazbeck defended her PhD Thesis [2].

### 6.3.3. *Map Matching*

**Participant:** François Charpillat.

This work [8] has been realized during the Intrade Projet with Maan Badaoui from Lille University. It addresses an important issue for intelligent transportation system, namely the ability of vehicles to safely and reliably localize themselves within an a priori known road map network. For this purpose, we have proposed an approach based on hybrid dynamic bayesian networks enabling to implement in a unified framework two of the most successful families of probabilistic model commonly used for localization: linear Kalman filters and Hidden Markov Models. The combination of these two models enables to manage and manipulate multi-hypotheses and multi-modality of observations characterizing Map Matching problems and it improves integrity approach. Another contribution of the paper is a chained-form state space representation of vehicle evolution which permits to deal with non-linearity of the used odometry model. Experimental results, using data from encoders' sensors, a DGPS receiver and an accurate digital roadmap, illustrate the performance of this approach, especially in ambiguous situations.

### 6.3.4. *Multi-Camera Tracking in Partially Observable Environment*

**Participants:** Arsène Fansi Tchango, Olivier Buffet, Vincent Thomas, Alain Dutech.

*Fabien Flacher (Thales ThereSIS) is an external collaborator.*

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 problem is difficult 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 (STAR) as presented in current state-of-the-art approaches.

A first novelty lies in the use of a complex behavioral simulator. In a single-target setting, we demonstrated that it allows inferring the behavior of a complex individual, even in case of long periods of occlusions (when cameras do not cover the trajectory of the target). This idea led to publications in AAMAS-14 [16], STAIRS-14 [18], and ECAI-14 [17].

A remaining issue is to find tractable algorithms for efficiently tracking multiple targets simultaneously, which requires using a factored particle filter (with one distribution per target). To that end, we use a Joint Probabilistic Data Association Filter with two key ingredients. The first ingredient is a particular model of dynamics that largely decouples the evolution of several targets, and turns out to be very natural to apply (which has led already to a publication in Fusion-14 [19]). Then, the factorization *a priori* implies, for a given target, simulating each of its particles with each particle of each other target (which leads to a huge number of simulations). The second proposed ingredient is to simulate each particle of a given target only with a small number of “representatives” of each other target (and then, because more particles are produced than needed, a selection/resampling step is required).

### 6.3.5. Emergence et Developmental Learning

**Participants:** Alain Dutech, Matthieu Zimmer.

*Yann Boniface (CORTEX, Loria) is an external collaborator*

Following our ongoing work on using reinforcement learning for the control of redundant continuous robotic systems, we explore how learning such complex tasks can benefit from a developmental approach, following some line of work already tested in robotics [50].

“Emergence”, one of the key concepts grounding this work, has been presented – from an artificial intelligence perspective – and discussed with researchers from other fields. This led to fruitful exchanges and a chapter in a book dedicated to the dual aspects of (human gestures) : appearance and emergence [36]. “Developmental Learning” was also the main subject of a seminar in Lyon in which Alain Dutech has been invited [47].

More concretely, we have developed several algorithms which mix artificial neural networks (like Dynamic Self-Organizing Maps or Reservoir Computing Network) with reinforcement learning mechanisms in order to build simple artificial systems that are *autonomous* and that learn without any *exogeneous* intervention from an external being. This work, initiated through two master thesis, is now the central topic of the PhD of Matthieu Zimmer, started in October 2014.

### 6.3.6. Online Evolutionary Learning

**Participants:** Amine Boumaza, François Charpillet, Iñaki Fernández.

Evolutionary Robotics (ER) deals with the design of agent behaviors using artificial evolution. Within this framework, the problem of learning optimal decision functions (or controllers) is treated as a policy search problem in the parameterized space of candidate policies. In this work we are interested in learning optimal behaviors for a swarm of mobile agents online (while solving the task). We adopt an online onboard distributed view [56], [48] and consider the learning process as executed at the agents’ level in a decentralized way. This kind of algorithms raises several questions concerning the usefulness of selection pressure (partial views of population, noisy fitness values, etc.).



We studied the impact of task-driven selection pressures in on-line distributed ER for swarm behavior learning. We proposed a variant of the mEDEA [45] algorithm in which we added a selection operator, in a task-driven scenario. We evaluated four selection methods that induce different intensity of selection pressure in a multi-robot navigation with obstacle avoidance task and a collective foraging task.

Experiments showed that a small intensity of selection pressure is sufficient to rapidly obtain good performances on the tasks at hand. We introduced different measures to compare the selection methods, and show that the higher the selection pressure, the better the performances obtained, especially for the more challenging food foraging task. This research was presented at the 13th International Conference on the Synthesis and Simulation of Living Systems [21].

### 6.3.7. *Frailty evaluation and Fall detection*

**Participants:** Amandine Dubois, François Charpillat, Thomas Moinel, Maxime Rio.

This work is related to the IPL PAL and Satelor project and is related to Personal Assistant Living (PAL) for elderly people with loss of autonomy.

- Clinical evaluation of frailty in the elderly is the first step to decide the degree of assistance that elderly people require. No standard tests exist to detect the level of frailty, each clinician chooses his protocol among existing tools. There are clinical tests as *Tinetti test*, *Timed Up and Go* test for evaluating the degree of dependance and the frailty of elderly people. These tests consist in asking a person to realize exercises simulating movements of daily life. The physician evaluates the quality of gait and the balance of the patient. These tests are often used but, the disadvantage is that the final verdict relies primarily on a subjective opinion. The aim of our work is to provide new objective criteria to refine the elderly frailty quantification. We base ourselves on the frailty definition of Fried *et al* as being a clinical syndrome in which three or more of the following criteria are present: unintentional weight loss, self reported exhaustion, weakness (with regards to grip strength), slow walking speed and low physical activity. From this definition, we have defined two axis of development to evaluate the frailty of a person: Sensor based Activity recognition with the aim to follow and report daily life activities in order to detect evolution that could reveal increased frailty [1], gait analysis in order to assess gait pattern and their evolution over time [14].
- An other PAL research domain, which is related to activity recognition, has attracted our attention: fall detection. Falls in the elderly is a major public health problem because of their frequency and their medical and social consequences. One of our objectives is to design an automatic system to detect fall at home, which in its final version will be made up of a network of RGB-D sensors, some of them being mobile embedded a wheel mobile robot.

The main contribution of this work has been to design a simple but robust method based on the identification and tracking of the center of mass of people evolving in an indoor environment through a RGB-D camera. Using a simple Hidden Markov Model whose observations are the position of the center of mass, its velocity and the general shape of the body, we have shown that we can surprisingly monitor the activity of a person with high accuracy, detect falls with very good accuracy without false positives and also measure some interesting parameter such as speed of gait, length of steps, etc. An experimental study, that is reported in [46], has been driven in our smart apartment lab. 26 subjects were asked to perform a predefined scenario in which they realized a set of eight postures. 2 hours of video (216 000 frames) were recorded for the evaluation, half of it being used for the training of the model. The system detected the falls without false positives. This result encourages us to use this system in real situation for a better study of its efficiency. Therefore, we started this year an experimentation in a room of a follow-up care and rehabilitation facility (OHS) in Nancy. "Office d'Hygiène Sociale" (OHS) is an association under the law of 1901. It supports nearly 800 people over 60 years and nearly 1,000 children and adults with disabilities. The association manages 26 facilities (40% health field, 40% medical-social field and 20% social field) and employs more than 1,500 professionals.

### 6.3.8. *Posture recognition with a Depth camera*

**Participants:** Abdallah Dib, François Charpillat, Xuan Nguyen, Alain Filbois [SED].

In this research line, we focus our contribution on improving model-based approaches that use a population-based stochastic framework for full human body tracking using monocular depth camera. One of the major challenges in human tracking is the high-dimensional state spaces. To address this problem, we propose a tracking algorithm based on APF and CMA-ES. While APF has been widely applied for human tracking in RGB and depth images, the application of CMA-ES to human tracking is still limited. Yet, CMA-ES shares many similar ideas with APF and can be exploited to improve the performance of APF. Our key idea is to update the covariance matrix for sampling particles at each layer of APF, using a subset of best particles, an idea inspired from CMA-ES. The resulting algorithm is shown to greatly reduce the number of particles required for successful tracking. In the absence of image features such as texture or color, existing likelihood models for human tracking in depth images are often built by computing distances between data points and model points sampled on the surface of the human body model. When human body parts are close or when severe self-occlusions are present, these models fail to capture good pose hypotheses. As a result, existing approaches are unable to track a broad range of human motions. To deal with this issue, we propose a likelihood model which is based on comparing observed depth images and rendered depth images obtained by classic rendering techniques. Combining with our tracking algorithm, the proposed likelihood model has been shown to be effective when tracking under severe self-occlusions. To the best of our knowledge, our approach is the first model-based one that uses a population-based stochastic framework able to track full human body with non-frontal and unusual poses, using monocular depth camera.

### 6.3.9. Pressure sensing floor

**Participants:** Mihai Andries, François Charpillet, Olivier Simonin.

The use of floor-sensors in ambient intelligence contexts began in the late 1990's, with projects like ORL active floor, the Magic carpet by Paradiso *et al.*, and the smart floor by Orr *et al.* These floors were, later on, integrated in smart environments, aimed at delivering assistance services like continuous diagnosis of users' health. According to the literature there are currently at least 6 main types of floor pressure sensing technologies: binary switches, piezoelectric, load cells, capacitive, polymer thick film (PTF), and photo interrupter sensors. Most of presented solutions extract a set of features for their tracking and identification task. Recently, sensing floors products like the SensFloor (a floor network of capacitive proximity sensors), Capfloor (a network of capacitive sensors), Elsi® smart floor (<http://www.elsitechnologies.com>) and FloorInMotion (Tarkett France) started being commercialized by companies, mainly for the senior care industry.

We have ourselves developed a sensing floor. This load-sensing floor 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. Each tile has 16 light-emitting diodes which provide visual feedback. The processing units were manufactured by Hikob<sup>0</sup>. This prototype was originally designed as a medium of interaction for robots with distributed control, in an ant-like fashion. The computing unit available on each tile can register a virtual pheromone trace, that can then be transmitted to other robots, using either wired or wireless communication. In a different perspective, the sensing-floor acts merely as a sensor for an ambient intelligence. Using the magnetometer embedded on the processing unit of the tile, each tile can detect disturbances in its surrounding magnetic field, that can be caused by the presence of robots. Each tile also has an embedded accelerometer, that allows it to detect shocks that can be caused by objects or humans falling on the ground.

Several functionalities have been implemented this year on this prototype floor, including weight measurement, fall detection, footstep tracking and activity recognition. We also implemented heuristic real-time multi-user localisation (without user identification) in an indoor setting using this prototype floor.

### 6.3.10. Living assistant Robot

**Participants:** François Charpillet, Nicolas Beaufort, Abdallah Dib.

---

<sup>0</sup><http://www.hikob.com/>



With LAR (**living Assistant Robot**), a PIA projet which started in March, Abdallah Dib joined our team for a PhD. His work is about the development of a low cost navigation system for a robot evolving in an indoor environment. The main issue of his work is to design a Simultaneous Localisation and Mapping algorithm working in a dynamic environment in which people are moving. This is very challenging if we restrict the sensing capabilities of the robot with low cost sensors such as RGB-D camera. An important service we expect the robot to achieve, is realizing similar services as the one we described below: fall detection, activity recognition. This year first result have been published [11]. A feature based visual SLAM method that uses chamfer distance to estimate the camera motion from RGB-D images has been presented. The method does not require any matching which is an expensive operation and always generates false matching that affects the estimated camera motion. Our approach registers the input image iteratively by minimizing the distance between the feature points and the occupancy grid using a distance map. We demonstrate with real experiments the capability of the method to build accurate 3D map of the environment with a hand-held camera. While the system was mainly developed to work with RGB-D camera, occupancy grid representation gives the method the ability to work with various types of sensors, we show the capacity of the system to construct accurate 2D maps using telemeter data. We also discuss the similarities between the proposed approach and the traditional ICP algorithm.

### 6.3.11. Exploring an unknown environment with a team of mobile robots

**Participants:** François Charpillat, Olivier Simonin, Nassim Kaldé.

This work is the continuation of the work realized during the ANR Cart-O-matic (2010 to 2013). We address, here, the problem of efficient allocation of the navigational goals in the multi-robot exploration of unknown environment. 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, this year, we have proposed an evaluation framework to study exploration strategies independently on the available computational resources. A comparison of the selected task-allocation algorithms deployed in multi-robot exploration has been done and published with Jan Faigl from Czech Technical University in Prague in the framework of the PHC project MACOREX.

An other point that is addressed by Nassim Kaldé is to consider the same problem but with dynamical environment in particular populated with human beings. First results of Nassim Kalde have been published in JFSMA'14 [33]. He published too the work done during his Master thesis [23].

## 6.4. Understanding and mastering complex systems

### 6.4.1. Adaptive control of a complex system based on its multi-agent model

**Participant:** Vincent Chevrier.

*Laurent Ciarletta (Madyne team, LORIA) is an external collaborator.*

Complex systems are present everywhere in our environment: internet, electricity distribution networks, transport networks. These systems have as characteristics: a large number of autonomous entities, dynamic structures, different time and space scales and emergent phenomena. This thesis work is centered on the problem of control of such systems. The problem is defined as the need to determine, based on a partial perception of the system state, which actions to execute in order to avoid or favor certain global states of the system. This problem comprises several difficult questions: how to evaluate the impact at the global level of actions applied at a global level, how to model the dynamics of an heterogeneous system (different behaviors issue of different levels of interactions), how to evaluate the quality of the estimations issue of the modeling of the system dynamics.

We propose a control architecture based on an “equation-free” approach. We use a multi-agent model to evaluate the global impact of local control actions before applying the most pertinent set of actions.

Associated to our architecture, an experimental platform has been developed to confront the basic ideas or the architecture within the context of simulated “free-riding” phenomenon in peer to peer file exchange networks. We have demonstrated that our approach allows to drive the system to a state where most peers share files, despite given initial conditions that are supposed to drive the system to a state where no peer shares. We have also executed experiments with different configurations of the architecture to identify the different means to improve the performance of the architecture.

This work helped us to better identify [26] the key questions that rise when using the multi-agent paradigm in the context of control of complex systems, concerning the relationship between the model entities and the target system entities.

#### 6.4.2. *Multi Modeling and multi-simulation*

**Participants:** Vincent Chevrier, Christine Bourjot, Benjamin Camus, Julien Vaubourg.

*Laurent Ciarletta and Yannick Presse (Madynes team, LORIA) are external collaborators.*

*Laurent Ciarletta is the co-advisor of the thesis of Julien Vaubourg.*

Complex systems generally require to use different points of view (abstraction levels) at the same time on the system in order to capture and to understand all the dynamics and the complexity. Being made of different interacting parts, a model of a complex system also requires simultaneously modeling and simulation (M&S) tools from different scientific fields.

We proposed the AA4MM meta-model [54] that solves the core challenges of multimodelling and simulation coupling in an homogeneous perspective. In AA4MM, we chose a multi-agent point of view: a multi-model is a society of models; each model corresponds to an agent and coupling relationships correspond to interaction between agents.

This year we progress in the definition of multi-level modeling [42]. We identified several facets of multi-level modeling and implemented then as different kinds of interactions in AA4MM framework. We progressed on the specification of the meta-model which helped to define a modeling environment.

In the MS4SG projet which involves MAIA, Madynes and EDF R&D on smart-grid simulation, we developed a proof of concepts for a smart-apartment case [10].

#### 6.4.3. *Cellular automata as a foundation of complex systems*

**Participant:** Nazim Fatès.

Our research on emergent collective behavior focuses on the analysis of the robustness of discrete models of complex systems. We ask to which extent systems may resist to various perturbations in their definitions. We progressed in the knowledge of how to tackle this issue in the case of cellular automata (CA) and multi-agent systems (MAS).

We proposed an extended version of our survey on asynchronous cellular automata [3].

In collaboration with colleagues from India, we proposed a complete characterisation of the reversibility of the set of the 256 Elementary Cellular Automata with asynchronous updating [29]. These rules are known to be difficult to study in all generality and it is interesting to notice that here, asynchronism is an aid rather than an obstacle to analyse the behaviour of the systems.

With Henryk Fukś (Brock Univ., Canada), we proposed a mathematical analysis of the second-order phase transitions that are observed in the most simple asynchronous cellular automata [22].

Our work on the classification of cellular automata was presented in the AUTOMATA’14 conference and is now the topic of a collaboration with L. Gerin (École Polytechnique) [44], [20].

We are currently participating to the edition of the first book devoted to probabilistic cellular automata and to a special issue of the French-speaking journal *Technique et Science Informatique* (Lavoisier editors).

#### 6.4.4. *Revisiting wavefront construction with collective agents: an approach to foraging.*

**Participants:** François Charpillat, Olivier Simonin.

We consider here [7], the problem of coordinating a team of agents that have to collect disseminated resources in an unknown environment. We are interested in approaches in which agents collectively explore the environment and build paths between home and resources. The originality of our approach is to simultaneously build an artificial potential field (APF) around the agents' home while foraging. We propose a multi-agent model defining a distributed and asynchronous version of Barraquand et al. Wavefront algorithm. Agents need only to mark and read integers locally on a grid, that is, their environment. We prove that the construction converges to the optimal APF. This allows the definition of a complete parameter-free foraging algorithm, called c-marking agents. The algorithm is evaluated by simulation, while varying the foraging settings. Then we compare our approach to a pheromone-based algorithm. Finally, we discuss requirements for implementation in robotics.

## OAK Project-Team

# 6. New Results

## 6.1. Highlights of the Year

The year has allowed reaching important results in four research areas of the group: query-based why-not provenance with explanations , minimal query reformulations under constraints [14], Linked Open Data analytics , and RDF data management in the cloud .

BEST PAPERS AWARDS :

[6] **Extending Database Technology (EDBT)**. N. BIDOIT, M. HERSCHEL, K. TZOMPANAKI.

[10] **23rd International World Wide Web Conference**. D. COLAZZO, F. GOASDOUÉ, I. MANOLESCU, A. ROATIS.

[], [23] **The International Journal on Very Large Databases**. Z. KAOUDI, I. MANOLESCU.

## 6.2. Scalable and Expressive Techniques for the Semantic Web

A main scientific topic of the team is the design of expressive and efficient tools for analyzing and manipulating Semantic Web data, in particular RDF. Our 2014 results in this area follow three complementary directions.

First, we have finalized our model for RDF analytics and proposed a full framework in which we fully redesign, from the bottom up, core data analytics concepts and tools in the context of RDF data, leading to the first complete formal framework for warehouse-style RDF analytics. Notably, we defined *i) analytical schemas* tailored to heterogeneous, semantics-rich RDF graph, *(ii) analytical queries* which (beyond relational cubes) allow flexible querying of the data and the schema as well as powerful aggregation and *(iii) OLAP-style operations*. We implemented our RDF analytics platform on top of the KDB system and ported it on Postgres as well [10], [28]; work is ongoing to adapt it on a massively parallel RDF query evaluation platform, namely CliqueSquare (see below). In [24], we describe novel techniques for optimizing the evaluation of RDF analytical queries based on previously computed analytical query results.

Second, we continued our work on efficient evaluation of queries on RDF data, in the presence of constraints. *Reformulation-based query answering* is a query processing technique aiming at answering queries against data, under constraints. It consists of reformulating the query based on the constraints, so that evaluating the reformulated query directly against the data (i.e. without considering any more the constraints) produces the correct answer set. We have show how to optimize reformulation-based query answering in the setting of *ontology-based data access*, where SPARQL conjunctive queries are posed against RDF facts on which constraints expressed by an RDF Schema hold. The literature provides solutions for various fragments of RDF, aiming at computing the equivalent union of maximally-contained conjunctive queries w.r.t. the constraints. However, in general, such a union is large, thus it cannot be efficiently processed by a query engine. In this context, we have shown that generalizing the query reformulation language allows considering a space of reformulated queries (instead of a single possible choice), and selecting the reformulated query with lower estimated evaluation cost. We have shown experimentally that our technique enables reformulation-based query answering where the state-of-the-art approaches are simply unfeasible, while it may decrease their costs by orders of magnitude in other cases [20], [26].

Third, we have continued our work on cloud-based RDF data management. In [22], we have demonstrated CliqueSquare, a platform we developed in the team for the massively parallel processing of RDF queries. CliqueSquare enjoys the benefits of a query optimization algorithm which creates query plans as flat as possible, which in turn translates into massive opportunities for parallel processing. In [23], we have finalized our work on managing RDF data within the Amazon Web Services cloud. Finally, we have conducted a study of the existing models and algorithms published so far for the massively parrallel processing of RDF queries, which appeared as a survey in the VLDB Journal [] and was also the basis of a tutorial at the ACM SIGMOD conference.

### 6.3. Massively Distributed Data Management Systems

Work in this area concerning the massively parallel processing of Semantic Web data was covered within the respective module.

We have finalized our work on massively parallel processing of XML queries based on the Apache Flink framework, formerly known as Stratosphere from the Technical University of Berlin, which implements the PACT model (an expressive extension of MapReduce). In [21], we have addressed the problem of efficiently parallelizing the execution of complex nested data processing, expressed in XQuery. We provided novel algorithms showing how to translate such queries into PACT, a recent framework generalizing MapReduce in particular by supporting many-input tasks. We presented the first formal translation of complex XQuery algebraic expressions into PACT plans, and demonstrated experimentally the efficiency and scalability of our approach. The work has recently been accepted for publication to IEEE TKDE (to appear in 2015),

Finally, we have considered improving the performance of massively parallel data processing programs expressed using the PigLatin language. PigLatin is a popular language within the data management community interested in the efficient parallel processing of large data volumes. The dataflow-style primitives of PigLatin provide an intuitive way for users to write complex analytical queries, which are in turn compiled into MapReduce jobs. Currently, subexpressions occurring repeatedly in PigLatin scripts are executed as many times as they occur, leading to avoidable MapReduce jobs. The current PigLatin optimizer is not capable of recognizing, and thus optimizing, such repeated subexpressions. In [18], we have presented We present a novel approach for identifying and reusing common subexpressions occurring in PigLatin scripts. In particular, we lay the foundation of our reuse-based algorithms by formalizing the semantics of the PigLatin query language with extended nested relational algebra for bags. Our algorithm, named PigReuse, operates on the algebraic representations of PigLatin scripts, identifies subexpression merging opportunities, selects the best ones to execute based on a cost function, and merges other equivalent expressions to share its result; our experiments have confirmed the efficiency and effectiveness of our reuse-based algorithms and optimization strategies.

### 6.4. Advanced Algorithms for Data Querying and Transformation

We revisit in [14] the Chase&Backchase (C&B) algorithm for query reformulation under constraints. For an important class of queries and constraints, C&B has been shown to be complete, i.e. guaranteed to find all (join-)minimal reformulations under constraints. C&B is based on constructing a canonical rewriting candidate called a universal plan, then inspecting its exponentially many sub-queries in search for minimal reformulations, essentially removing redundant joins in all possible ways. This inspection involves chasing the subquery. Because of the resulting exponentially many chases, the conventional wisdom has held that completeness is a concept of mainly theoretical interest. We show that completeness can be preserved at practically relevant cost by introducing a novel reformulation algorithm that instruments the chase to maintain provenance information connecting the joins added during the chase to the universal plan subqueries responsible for adding these joins. This allows it to directly “read off” the minimal reformulations from the result of a single chase of the universal plan, saving exponentially many chases of its subqueries. We exhibit natural scenarios yielding speedups of over two orders of magnitude between the execution of the best view-based rewriting found by a commercial query optimizer and that of the best rewriting found by our algorithm.

Different types of explanations that serve as Why-Not answers have been proposed in the past and are either based on the available data, the query tree, or both. A first approach to this so called why-not provenance has been recently proposed. In [6], we show that this first approach has some shortcomings. To overcome these shortcomings, we propose Ned, an algorithm to explain data missing from a query result. NedExplain computes the why-not provenance for monotone relational queries with aggregation. This work contributes to providing necessary formalization in which the new algorithm is build. It also develops a comparative evaluation showing that it is both more efficient and effective than the state-of-the-art approach.

Solutions to answering Why-Not questions are generally more efficient and easier to interpret by developers than solutions solely based on data. However, algorithms producing such query-based explanations including ours ([6]) so far may return different results for reordered conjunctive query trees, and even worse, these results

may be incomplete. Clearly, this represents a significant usability problem, as the explanations developers get may be partial and developers have to worry about the query tree representation of their query, losing the advantage of using a declarative query language. As remedy to this problem, in [5][17], we propose to capture query based answers of Why-Not questions through operator polynomial and we devised an algorithm called Ted that produces the same complete query-based explanations for reordered conjunctive query trees.

## **6.5. Social Data Management and Crowdsourcing**

In [12], we focused on the issue of defining models and metrics for reciprocity in signed graphs. In unsigned directed networks, reciprocity quantifies the predisposition of network members in creating mutual connections. On the other hand, this concept has not yet been investigated in the case of signed graphs. We capitalize on the graph degeneracy concept to identify subgraphs of the signed network in which reciprocity is more likely to occur. This enables us to assess reciprocity at a global level, rather than at an exclusively local one as in existing approaches. The large scale experiments we perform on real world data sets of trust networks lead to both interesting and intuitive results. We believe these reciprocity measures can be used in various social applications such as trust management, community detection and evaluation of individual nodes. The global reciprocity we define in this paper is closely correlated to the clustering structure of the graph, more than the local reciprocity as it is indicated by the experimental evaluation we conducted.

As initial step towards better answering information needs in applications managing social content that is structured and possibly enriched with semantic annotations, in [19], we present a preliminary data model and an approach for answering queries over structured, social and semantic-rich content, taking into account all dimensions of the data in order to return the most meaningful results.

## ORPAILLEUR Project-Team

## 6. New Results

### 6.1. Highlights of the Year

As highlights of the year, we would like to mention several elements, an award in a competition and a best paper. In addition we would like to also mention the importance gained by two other papers.

- Yen Low, a postdoctoral fellow from Stanford and Adrien Coulet (Orpailleur team) jointly developed a prototype named *Whypothesis?* whose goal is to provide explanations on drug side effects for which the molecular mechanism remains unknown. This prototype won the “Best Application Award” at the 2014 NCBO Hackathon (National Center for Biomedical Ontology), held at Stanford University, April 26-27 ([http://www.bioontology.org/2014\\_NCBO\\_Hackathon](http://www.bioontology.org/2014_NCBO_Hackathon)).
- The paper [2] describing a first and original proposition for combining pattern structures and relational concept analysis won the best paper award at the International Conference on Formal Concept Analysis in Cluj-Napoca, Romania.
- The paper [10] published in Nucleic Acids Research describes the latest version of KBDOCK, which has had over 12,000 non-duplicate visitors since 2011.
- The paper [44] on polypharmacology represents a nice collaboration with Harmonic Pharma, and it was used for the cover issue of Journal Chemical Information (<http://pubs.acs.org/toc/jcisd8/54/3>).

BEST PAPER AWARD :

[56] **Formal Concept Analysis - 12th International Conference - Proceedings.** V. CODOCEDO, A. NAPOLI.

### 6.2. The Mining of Complex Data

**Participants:** Mehwish Alam, Aleksey Buzmakov, Melisachew Chekol, Victor Codocedo, Adrien Coulet, Elias Egho, Nicolas Jay, Florence Le Ber, Ioanna Lykourantzou, 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

Formal Concept Analysis and pattern mining are suitable symbolic methods for KDDK, that may be used for real-sized applications. Global improvements are carried 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.

#### 6.2.1. FCA and Variations: RCA, Pattern Structures and Biclustering

There are a few extensions of FCA for handling contexts involving complex data formats, e.g. graphs or relational data. Among them, Relational Concept Analysis (RCA) is a process for analyzing objects described both by binary and relational attributes [2] [131]. The RCA process takes as input a collection of contexts and of inter-context relations, and yields a set of lattices, one per context, whose concepts are linked by relations. RCA can play an important role in KDDK, especially in text mining [105].

Another extension of FCA is based on Pattern Structures (PS) [112], which allows to build a concept lattice from complex data, e.g. nominal, numerical, and interval data [119]. Since then, we worked on some experiments involving pattern structures, namely sequence mining [107], information retrieval and recommendation [58], [22], functional dependencies [50], [17] and biclustering [69], [41]. One of the next step is the adaptation of pattern structures to graph mining.



Moreover, the notion of similarity between objects is also closely related to pattern structures [102]: two objects are similar as soon as they share the same attributes (binary case) or attributes with similar values or the same description (at least in part). Combination of similarity and pattern structures is also under study, in particular for solving information retrieval and annotation problems.

In pattern mining as in FCA, one main problem is the volume of the output. One general idea is to extract patterns which show a “good behavior” w.r.t. a given measure. Such patterns or concepts are expected to have good characteristics and to provide effective knowledge. We have conducted in the framework of FCA a series of experiments on the so-called “stability measure”, showing that this measure is able to detect significant patterns [54], [53].

Finally, there is also an on-going work relating FCA and semantic web. This work focuses on the classification within a concept lattice of the answers returned by SPARQL queries. The concept lattice is then used as an index for navigating and ranking the answers w.r.t. their content and interest for a given objective [47].

### **6.2.2. Sequence Mining**

Sequence data is widely used in many applications. Consequently, mining sequential patterns and other types of knowledge from sequence data became an important data mining task. In the team, the main emphasis is on developing efficient mining algorithms for pattern classification problems. The most frequent sequences generally provide trivial information. When analyzing the set of frequent sequences with a low minimum support, the user is overwhelmed by millions of patterns.

In our recent work, we studied the notion of  $\delta$ -freeness for sequences. While this notion has extensively been discussed for itemsets, our work is the first to extend it to sequences. We defined an efficient algorithm devoted to the extraction of  $\delta$ -free sequential patterns. We presented the advantage of the  $\delta$ -free sequences and highlighted their importance when building sequence classifiers, and we showed how they can be used to address the feature selection problem in statistical classifiers which optimizes both accuracy and earliness of predictions [68].

### **6.2.3. Mining and Understanding Healthcare Trajectories**

With the increasing burden of chronic illnesses, administrative health care databases hold valuable information that could be used to monitor and assess the processes shaping the trajectory of care of chronic patients. In this context, temporal data mining methods are promising tools, though lacking flexibility in addressing the complex nature of medical events. In the thesis work of Elias Egho [15], new algorithms were designed to extract patient trajectory patterns with different levels of granularity by relying on external taxonomies [62], [34]. The algorithms rely on the general FCA framework to formalize the general notion of multidimensional healthcare trajectories. There was also another work focusing on the similarity measure among sequences. An efficient and original similarity measure was design for that purpose [8].

### **6.2.4. Video Game Analytics**

The video game industry has grown enormously over the last twenty years, bringing new challenges to the artificial intelligence and data analysis communities. We tackled this year the problem of 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. We presented a new formalism within a sequential pattern mining approach and a novel measure, the balance measure, telling how a strategy is likely to win [51]. We experimented our methodology on a real-time strategy game that is professionally played in the electronic sport community and laid plans on a future collaboration with the MIT Game Lab.

### **6.2.5. KDDK in 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 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. In addition, in his thesis work, Mohsen Sayed focused on extracting relations between named entities using graph mining methods applied to dependency graphs [67]. 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.

The notion of “Jumping Emerging Patterns” (JEP) previously used in chemistry [101], was updated and adapted in 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. a positive example, or as a simple word, i.e. a negative example. The method extracts from a training set all JEPs which are considered as hypotheses. To reduce the number of JEPs and to retain only the more significant JEPs from a linguistic point of view, JEPs are weighted and a constraint solver is used to verify the maximal coverage of the positive examples. Results are currently under evaluation.

### 6.3. KDDK in Life Sciences

**Participants:** Adrien Coulet, Marie-Dominique Devignes, Bernard Maignet, Gabin Personeni, David Ritchie, Malika Smail-Tabbone.

The 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 health care, 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.

#### 6.3.1. Inductive Logic Programming for Mining Linked Open Data

Increasing amounts of biomedical data provided as LOD offer novel opportunities for knowledge discovery in biomedicine. We proposed and published an approach for selecting, integrating, and mining LOD with the goal of discovering genes responsible for a disease [11]. The selection step relies on a set of choices made by a domain expert to isolate relevant pieces of LOD. Because these pieces are potentially not linked, an integration step is required to connect unlinked pieces. The resulting graph is subsequently mined using Inductive Logic Programming (ILP) that presents two main advantages. First, the input format compliant with ILP (first order logic) is close to the format of LOD (RDF triples). Second, domain knowledge can be added to this input and used during the induction step. We have applied this approach to the characterization of genes responsible for intellectual disability. For this real-world use case, we could evaluate ILP results and assess the contribution of domain knowledge. Our ongoing efforts explore how the combination of rules coming from distinct theories can improve the prediction accuracy [70] [16].

#### 6.3.2. 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 [118]. 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 [122]. One main challenge remains to develop a knowledge discovery approach able to mine correlations between annotations based on BioPortal ontologies, i.e. is it possible to discover interesting knowledge units within these annotations?

Then, we proposed an adaptation of FCA techniques, namely pattern structures, to explore the annotations of biomedical databases [108]. We considered documents of biomedical databases annotated with sets of ontological concepts as objects in a pattern structure. Corresponding annotations have been classified according to several dimensions, where a dimension is related to a particular aspect of domain knowledge. The pattern structure formalism was applied to classify these annotations, allowing to discover correlations between annotations but also lacks of completion in the annotations that could be fixed afterward. This adaptation of pattern structures opens many perspectives in term of ontology reengineering and knowledge discovery.

## 6.4. Structural Systems Biology

**Participants:** Marie-Dominique Devignes, Bernard Maigret, David Ritchie, Malika Smaïl-Tabbone.

**Keywords:** bioinformatics, chemistry, docking, knowledge discovery, screening, systems biology

Structural systems biology aims to describe and analyze the many components and interactions within living cells in terms of their three-dimensional (3D) molecular structures. We are currently developing advanced computing techniques for molecular shape representation, protein-protein docking, protein-ligand docking, high-throughput virtual drug screening, and knowledge discovery in databases dedicated to protein-protein interactions.

### 6.4.1. The Hex Protein Docking Program

Our *Hex* protein docking software is being more widely used than ever before. The unique polar Fourier correlation approach used in *Hex* [129] allows the expensive FFT part of its calculations to be greatly accelerated on modern graphics processors (GPUs) [130]. *Hex* is freely available for download for academic users at <http://hex.loria.fr>. A public GPU-powered server has also been created (<http://hexserver.loria.fr>) [123]. In the last four years, the server has performed some 63,700 docking runs, and the program has had some 37,000 downloads. The latest version of the program has been used successfully to dock symmetric dimers (unpublished results) in the international “CAPRI” docking experiment [115]. A manuscript on performing polar Fourier docking using symmetry constraints is in preparation with the Nano-D team at Inria Grenoble.

### 6.4.2. KBDOCK: Protein Docking Using Knowledge-Based Approaches

In order to explore the possibilities of using structural knowledge of protein-protein interactions, Anisah Ghoorah recently developed the KBDOCK system as part of her doctoral thesis project [116]. KBDOCK is available at <http://kbdock.loria.fr>. KBDOCK combines coordinate data from the Protein Data Bank [106] with the Pfam protein domain family classification [111] in order to describe and analyze all known protein-protein interactions for which the 3D structures are available. We have demonstrated the utility of KBDOCK [114] for template-based docking using 73 complexes from the Protein Docking Benchmark [117]. We recently presented results obtained using KBDOCK at the CAPRI conference on protein docking in Utrecht [115]. In late 2013, we updated KBDOCK with the latest data from Pfam and the Protein Data Bank. In 2014, an article describing the new version of KBDOCK was published in the special Database Issue of Nucleic Acids Research [10]. Since the KBDOCK web site (<http://kbdock.loria.fr>) was created in 2011, it has had over 12,000 distinct visitors.

### 6.4.3. Kpax: A New Algorithm for Multiple Flexible Protein Structure Alignments

We recently developed a new protein structure alignment approach called Kpax [128]. The approach exploits the fact that each amino acid residue has a carbon atom with a highly predictable tetrahedral geometry. This allows the local environment of each residue to be transformed into a canonical orientation, thus allowing easy comparison between the canonical orientations of residues within pairs of proteins using a novel scoring function based on Gaussian overlaps. The overall approach is two or three orders of magnitude faster than most contemporary protein structure alignment algorithms, while still being almost as accurate as the state-of-the-art TM-Align approach [134]. Kpax is now used heavily by the KBDock web server [10] to find structural templates for docking which might be beyond the reach of sequence-based homology modeling approaches. The Kpax program is also available for download at <http://kpax.loria.fr/>.

In 2014, the Kpax algorithm has been extended to allow flexible alignment and superposition of protein backbones and to perform multiple structure alignments, in analogy with multiple protein sequence alignments. Our early results show that incorporating backbone flexibility leads to much higher quality multiple alignments than can be achieved with existing algorithms.

#### 6.4.4. Polypharmacology: Developing New Uses for Old Drugs

In 2010, Violeta Pérez-Nuño joined the Orpailleur team thanks to a Marie Curie Intra-European Fellowship (IEF) award to develop new virtual screening algorithms (DOVSA). The aim of this project was to advance the state of the art in computational virtual drug screening by developing a novel consensus shape clustering approach based on spherical harmonic (SH) shape representations [126].

In 2012, Violeta joined Harmonic Pharma, a LORIA spin-out company for drug re-purposing, and we have since continued our collaborations to develop new algorithms for drug discovery and drug re-purposing. The observation that many existing drugs may be used to treat more than one disease is often referred to as “polypharmacology.” Our latest work on predicting polypharmacology uses a Gaussian clustering approach to identify groups of molecules with similar three-dimensional shapes. This work was published in the Journal of Chemical Information and Modeling [44]. An illustration from this article was used to provide the cover page for the March 2014 issue of the journal (<http://pubs.acs.org/toc/jcisd8/54/3>).

### 6.5. Around the Taaable research project

**Participants:** Valmi Dufour-Lussier, Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer.

**Keywords:** knowledge representation, description logics, classification-based reasoning, case-based reasoning, belief revision, semantic web

The Taaable project was originally created as a challenger of the Computer Cooking Contest (ICCB Conference) [4] (<http://taaable.fr>). A candidate to this contest is a system whose goal is to solve cooking problems.

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.

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 compute ingredient substitutions and to adjust the ingredient quantities [65] using engines included in the Revisor library (see § 5.4.5).

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 a meta-knowledge model to manage knowledge reliability. This model represents notions such as belief, trust, reputation and quality, as well as their relationships and rules to evaluate 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 [64].

Taaable won in 2014 the CCC originality challenge for all the open resources that the Taaable team developed during the last years for the CBR community: WikiTaaable, a semantic wiki containing cooking domain knowledge, Tuurbine, a generic ontology guided CBR engine over RDFS (see § 5.4.3), and Revisor, an adaptation engine implementing various revision operators (see § 5.4.5).

## 6.6. Some Results in Graph Theory

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

**Keywords:** graph theory, extremal graph theory, coloring, clustering

### 6.6.1. Structural and extremal graph theory

Regarding graph coloring, a conjecture of Gera, Okamoto, Rasmussen and Zhang on set coloring was solved. A *set coloring* of a graph  $G = (V, E)$  is a function  $c : V \rightarrow \{1, \dots, k\}$  such that whenever  $u$  and  $v$  are adjacent vertexes, it holds that  $\{c(x) : x \text{ neighbor of } u\} \neq \{c(x) : x \text{ neighbor of } v\}$ . In other words, there must be at least one neighbor of  $u$  that has a color not assigned to a neighbor of  $v$ , or *vice-versa*. The smallest  $k$  such that  $G$  admits a set coloring is the *set coloring number*  $\chi_s(G)$ . We confirmed the conjecture by proving that  $\chi_s(G) \geq \lceil \log_2 \chi(G) \rceil + 1$ , where  $\chi(G)$  is the (usual) chromatic number of  $G$ . This bound is tight.

Works have been started on a 12-year-old conjecture by Heckman and Thomas about the fractional chromatic number of graphs with no triangles and maximum degree at most 3. This conjecture is actually a natural generalization of a fact established by Staton in 1979. Heckman and Thomas posits that in every graph with no triangles, maximum degree at most 3 and arbitrary weights on the vertexes, there exists an independent set of weight at least  $5/14$  times the total weight of the graph.

Regarding extremal graph theory, two results have been obtained. The first one deals with permutation snarks, while the second one reads as follows.

*For every 3-coloring of the edges of the complete graph on  $n$  vertexes, there is a color  $c$  and a set  $X$  of 4-vertexes such that at least  $2n/3$  vertexes are linked to a vertex in  $X$  by an edge of color  $c$ .*

This theorem is motivated by a conjecture of Erdős, Faudree, Gould, Gyárfás, Rousseau and Schelp from 1989, which asserts that  $X$  can be of size 3 only. However, they were only able to prove that  $X$  can be of size 22. Recently, Rahil Baber and John Talbot managed to build upon our work in a very nice article: adding a new idea to our argument, they managed to confirm the conjecture.

### 6.6.2. Graph theory and other fields

Interactions of graph theory with other topics (theoretical computer science, number theory, group theory, sociology and chemistry) have been considered. Most of them are still in progress and some are published. For instance, regarding distributed computing, the purpose of our work was to question the global knowledge each node is assumed to start with in many distributed algorithms (both deterministic and randomized). More precisely, numerous sophisticated local algorithm were suggested in the literature for various fundamental problems. Noticeable examples are the MIS algorithms and the  $(\Delta + 1)$ -coloring algorithms. Unfortunately, most known local algorithms are *non-uniform*, that is, they assume that all nodes know good estimations of one or more global parameters of the network, e.g., the number of nodes  $n$ . Our work provides a rather general method for transforming a non-uniform local algorithm into a uniform one. Furthermore, the resulting algorithm enjoys the same asymptotic running time as the original non-uniform algorithm. Our method applies to a wide family of both deterministic and randomized algorithms. Specifically, it applies to almost all of the state of the art non-uniform algorithms regarding MIS and Maximal Matching, as well as to many results concerning the coloring problem.

### 6.6.3. Algorithmic Graph Theory and Clustering

Since September 2013, Mario Valencia has obtained a two years invitation (namely Inria "Délégation") for working at Inria Nancy – Grand Est, in the Orpailleur team, on graph theoretical aspects and data clustering. This research work consists in studying the modular decomposition techniques on the threshold graphs issues of the clustering process. The principal studied problem is known as the *Cluster Deletion Problem*: given a graph with real non negative edge weights, partition the vertexes into clusters (in this case cliques) in order to minimize the total weight of edges out of the clusters. Two papers were submitted to journals in 2014. In [94], we discovered a one-to-one correspondence between potential solutions of the cluster deletion problem and the minimum sum coloring problem, and use it to obtain a polynomial time algorithm to solve the cluster deletion problem in a special family of graphs called  $P_4$ -reducible graphs.

In [95], we studied the complexity of the cluster deletion problem on subclasses of chordal graphs and cographs. In particular, it is shown that the cluster deletion problem is NP-hard for unweighted chordal graphs and weighted cographs. Some polynomial-time solvable cases are also identified.

Moreover, the paper "b-coloring is NP-hard on co-bipartite graphs and polytime solvable on tree-cographs", has been accepted for publication in the journal *Algorithmica* [1].

### 6.6.4. Structural and Algebraic Graph Theory

We have also worked on the following topics. Golumbic, Lipshteyn and Stern proved that every graph can be represented as the edge intersection graph of paths on a grid, i.e., one can associate to each vertex of the graph a nontrivial path on a grid such that two vertexes are adjacent if and only if the corresponding paths share at least one edge of the grid. For a non-negative integer  $k$ ,  $B_k$ -EPG graphs are defined as graphs admitting a model in which each path has at most  $k$  bends. Circular-arc graphs are intersection graphs of open arcs of a circle. It is easy to see that every circular-arc graph is  $B_4$ -EPG, by embedding the circle into a rectangle of the grid. We proved also that every circular-arc graph is  $B_3$ -EPG (paper submitted).

We have studied the  $k$ -tuple chromatic number of the Cartesian product of two graphs  $G$  and  $H$  in [96]. We have shown that there exist graphs  $G$  and  $H$  such that  $\chi_k(G \square H) > \max\{\chi_k(G), \chi_k(H)\}$  for  $k \geq 2$ . Moreover, we have also shown that there exist graph families such that, for any  $k \geq 1$ , the  $k$ -tuple chromatic number of their Cartesian product is equal to the maximum  $k$ -tuple chromatic number of its factors.



## SMIS Project-Team

# 6. New Results

## 6.1. Flash-Based Data Management

**Participants:** Nicolas Ancaux, Matias Bjørling, Philippe Bonnet, Luc Bouganim [correspondent], Niv Dayan, Saliha Lallali, Philippe Pucheral, Iulian Sandu Popa.

There is a long tradition of work around the understanding and optimization of NAND Flash memory in the team (e.g., [7], [9]). Current work in this area covers the optimization of SSD use in DBMS engines and the design of Flash-based indexing techniques for textual and spatio-temporal data. These works on Flash-Based indexing complete the work initiated in the last years on the storage and indexing engine of PlugDB (not repeated in this report but the interested reader is referred to a DAPD'14 journal publication detailing these techniques [14]).

**Flash storage optimization.** Solid State Drives (SSDs), based on flash chips, are now the secondary storage of choice for data intensive applications. Database systems can now rely on high performance SSDs to store log, indexes and data either on servers or in the cloud. While SSDs provide increasingly high performance out of the box, maintaining high throughput and low latency as the utilization of SSDs increases and despite abrupt changes in the workload remains a challenge. This question is central for database designers and administrators, cloud service providers, and SSD constructors. The answer depends on write-amplification, i.e., garbage collection overhead. More specifically, the answer depends on how write-amplification evolves in time. We derived a mathematical expression that relates over provisioning to write-amplification. We introduced a new block manager, called Wolf, or WORKload Leveler for Flash. Wolf is able to detect and quickly adapt to changes in workload by pro-actively reallocating over-provisioned space among the groups based on their changing needs. It adapts better to stable workloads by measuring the update frequencies of groups instead of making assumptions about them. It uses a novel near-optimal closed-form expression to allocate over-provisioned space to groups.

**Flash-based keyword indexing.** As smart objects gain the capacity to acquire, store and process large volumes of data, new services emerge. However, the smart objects have to be endowed with typical data management capabilities to enable all these services. 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 smart meters to perform analytic tasks (i.e., top-k queries) over sets of events (i.e., terms) captured during time windows (i.e., files) [21]. Designing such embedded search engine is however challenging due to a combination of severe and conflicting hardware constraints (e.g., a tiny RAM combined with a NAND Flash persistent storage badly adapted to random fine-grain updates). To tackle this challenge, we introduce three design principles, namely Write-Once Partitioning, Linear Pipelining and Background Linear Merging, and show how they can be combined to produce an embedded search engine reconciling high insert/delete/update rate and query scalability. We have implemented our search engine on a development board having a hardware configuration representative for smart objects. The experimental results demonstrate the scalability of the approach and its superiority compared to state of the art methods [28]. This work is part of Saliha Lallali's Ph.D. thesis.



**Flash-based spatio-temporal indexing.** The convergence of mobile computing, wireless communications and sensors has raised the development of many applications exploiting a massive flow of spatio-temporal data such as location-based services, participatory sensing, or traffic management [15]. Among the most active research topics in this area is the spatio-temporal data indexing. Nevertheless, since a few years a new fundamental parameter has made its entry on the database scene: the NAND flash storage. However, the peculiar characteristics of flash memory require redesigning the existing data storage and indexing techniques that were devised for magnetic hard-disks. In this study we propose TRIFL, an efficient and generic TRajjectory Index for FLash. TRIFL is designed around the key requirements of 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 is supplied with 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, which makes the index appropriate for many application scenarios. The experimental evaluation shows that TRIFL outperforms the representative indexing methods on magnetic disks and flash disks. This work is part of Dai-Hai Ton That 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.

Current applications, from complex sensor systems (e.g. quantified self) to online e-markets acquire vast quantities of personal information which usually ends-up on central servers. Decentralized architectures, devised to help individuals keep full control of their data, hinder global treatments and queries, impeding the development of services of great interest. In this study, we promote the idea of pushing the security to the edges of applications, through the use of secure hardware devices controlling the data at the place of their acquisition. To solve this problem, we propose secure distributed querying protocols based on the use of a tangible physical element of trust, reestablishing the capacity to perform global computations without revealing any sensitive information to central servers. This leads to execute global treatments on an asymmetric architecture, composed of a powerful, available and untrusted computing infrastructure (server or cloud), and a large set of low powered, highly disconnected trusted devices. Given our large scale data centric applications (e.g. nationwide surveys), we discard solutions based on secure multi-party computation, which do not scale. We have studied two different computing paradigms on this architecture: our first contribution 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 [12], including vulgarization aspects [25]. Our second contribution was to study general SQL queries in this same execution context. For now, we have 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 [19]. Cost models and experiments demonstrate that this approach can scale to nationwide infrastructures [20][16]. This work is part of Quoc-Cuong To's Ph.D. thesis started in sept. 2012, and should be extended in particular to cover joins. We also plan to extend 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.

## 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 [5]. 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. Many projects and startups currently investigate this solution, like OpenPDS, CozyCloud, OwnCloud, etc. 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 [11]. On this basis, we investigate new data sharing and dissemination models, where usage and access control rules endorsed by the individuals could be enforced. In 2014, we have presented this vision at EDBT'14 [18]. Several underlying research problems and perspectives have been presented in [11]. 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 just started to explore new data sharing techniques which could be enforced in the secure personal cloud model. Athanasia Katsouraki is working on privacy issues and on adoption of the secure data engine in cooperation with the economists (CERDI) in the context of the Digital Society Institute (DSI).

## 6.4. Folk-IS

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

According to many studies, IT should become a key facilitator in establishing primary education, reducing mortality or supporting commercial initiatives in Least Developed Countries. The main barrier to the development of IT services in these regions is not only the lack of communication facilities, but also the lack of consistent information systems, security procedures, economic and legal support, as well as political commitment. In [5], we proposed the vision of trusted cells, a data platform for personal data services where the shared infrastructure (typically the cloud) is untrusted, while personal devices (such as smart phones, tablets or set-top box) are trusted execution environments. We revisited this vision to the context of LDCs. 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. As trusted cells, 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. We have published in [17] the Folk-IS vision and main principles, and in [13] a more detailed paper including technical challenges, specific to that approach and an exploitation and feasibility analysis of the Folk-IS vision.

## WIMMICS Project-Team

# 6. New Results

## 6.1. Highlights of the Year

BEST PAPERS AWARDS :

[52] **IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT**. A. G. B. TETTAMANZI, C. DA COSTA PEREIRA.

[45] **10th IEEE International Conference Beyond Databases, Architectures, and Structures (BDAS 2014)**. T. H. H. NGUYEN, N. LE THANH.

## 6.2. Argumentation Theory

### 6.2.1. *Combining Argumentation Theory and Normative Reasoning with Natural Language Processing*

**Participants:** Serena Villata, Elena Cabrio, Fabien Gandon.

We have proposed a methodology to identify and classify the semantic relations holding among the possible different answers obtained for a certain query on DBpedia language specific chapters. The goal is to reconcile information provided by language specific DBpedia chapters to obtain a consistent results set. The results of this research have been published at the LREC conference [29]. This classification has then been exploited in another work, together with Elena Cabrio and Alessio Palmero Arosio (FBK Trento, Italy), where Serena Villata has 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, she extended the results presented last year embedding the new identified relations among the different answers, using argumentation theory to reconcile the information and further improving the system's performances. A demo of the new argumentation module is available online <sup>0</sup>. This work has also been presented at the International Semantic Web Conference demo session [85].

Moreover, we have proposed, together with Alessio Palmero Arosio, a system called NLL2RDF to translate in an automated way licenses, such as GPL, in natural language into a machine-readable version using the RDF language. The system is available online <sup>0</sup>. The results of this research have been presented at the European Semantic Web conference [26].

Finally, we have published the benchmark of natural language arguments called NoDE. The benchmark is available online <sup>0</sup>. The results of this research have been presented at the 15th International Workshop on Non-Monotonic Reasoning (NMR 2014) [28] and at the 5th Conference on Computational Argumentation [27] (COMMA 2014 - demo).

### 6.2.2. *Argumentation and Legal Reasoning*

**Participant:** Serena Villata.

---

<sup>0</sup><http://qakis.org/qakis2/>

<sup>0</sup><http://www.airpedia.org/nll2rdf/>

<sup>0</sup><http://www-sop.inria.fr/NoDE/>

Together with Leendert van der Torre (University of Luxembourg), we proposed a framework for reasoning about norms using argumentation theory. Norms regulate our everyday life, and are used to assess the conformance of our behavior with respect to the regulations holding in specific contexts. Given the profound importance of norms in our lives, it is fundamental to understand which norms are valid in certain environments, how to interpret them, the legal conclusions of such norms, which norms can be derived from the existing ones, etc. In order to understand norms, people discuss about them to assess the validity or applicability of a certain norm subject to particular conditions, to derive the obligations and permissions to be enforced, or claim that a certain normative conclusion cannot be derived from the existing regulations. Several frameworks have been proposed for legal argumentation, but no comprehensive formal model of legal reasoning from arguments has been proposed yet. The goal of this work is to enrich legal argumentation with a formal account of deontic modalities. These results have been published at the 5th Conference on Computational Argumentation [55] (COMMA 2014).

Moreover, together with Guido Boella (University of Torino, Italy), Pietro Baroni and Massimiliano Giacomin (University of Brescia, Italy), Federico Cerutti (University of Aberdeen, UK), Leendert van der Torre (University of Luxembourg), we have studied also the dynamics of argumentation framework and this research has lead to a publication in the Artificial Intelligence journal [15].

## 6.3. License for the Web of Data

### 6.3.1. Reasoning about rights and licenses in the Web of Data

**Participant:** Serena Villata.

In the domain of Linked Open Data a need is emerging for developing automated frameworks able to generate the licensing terms associated to data coming from heterogeneous distributed sources. Together with Guido Governatori (NICTA, Australia) and Antonino Rotolo (University of Bologna, Italy), Serena Villata proposed and evaluated a deontic logic semantics which allows to define the deontic components of the licenses, i.e., permissions, obligations, and prohibitions, and generate a composite license compliant with the licensing items of the composed different licenses. The approach is evaluated using the SPINdle defeasible reasoner, where the proposed heuristics have been hard coded in the reasoner. The prosecution of this research line has seen the analysis of the compatibility of a set of licensing terms (always using SPINdle), the analysis of the role of licenses associated to vocabularies, and the development of the Licentia suite of services to reason over licenses and help users to deal with such kind of information. The results of this research line have been published at the International Semantic Web Conference demo session [87].

## 6.4. Logic and Cognition

### 6.4.1. Logical Foundations of Cognitive Agents

**Participants:** Andrea Tettamanzi, Serena Villata.

Together with Célia da Costa Pereira of I3S, we have investigated syntactic belief revision operators [16] and goal-generation mechanisms [30] to make the practical implementation of a general BDI (Belief-Desire-Intention) model of agency based on possibility theory. Furthermore, we took part in a joint investigation with a research team, led by Cristiano Castelfranchi, of the CNR-ISTC in Rome on the issue of trust in multi-agent systems [21]. We also employed agent-based simulation to test a theory of human stupidity proposed by the late Italian economist Carlo Cipolla [52]; our paper won the Best Paper Award at IAT 2014.

### 6.4.2. RDF Mining

**Participants:** Andrea Tettamanzi, Catherine Faron Zucker, Fabien Gandon.

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 testing candidate OWL 2 axioms against the fact contained in an RDF base and proposed a novel scoring heuristics based on falsification and possibility theory [53].

### 6.4.3. *Sentiment Analysis*

**Participant:** Andrea Tettamanzi.

Together with Célia da Costa Pereira of I3S and Mauro Dragoni of FBK, Trento, who visited our team for three months from April to June 2014, we have proposed a novel approach to concept-level sentiment analysis based on fuzzy logic. Our system [31], [89] participated in the Semantic Web Evaluation Challenge (SemWebEval) at ESWC 2014 and was the winner for Task 1 and Most Innovative Approach.

### 6.4.4. *Data and Knowledge Integration and Extraction*

**Participant:** Andrea Tettamanzi.

Together with Somsack Inthasone, Nicolas Pasquier and Célia da Costa Pereira of I3S, we developed a data warehouse collecting data for research on biodiversity [38].

### 6.4.5. *Miscellaneous*

**Participant:** Andrea Tettamanzi.

A work on electrocardiographic signal classification using evolutionary algorithms and neural networks carried out while still at the University of Milan, got published as a book chapter [65].

## 6.5. Natural Language Processing

### 6.5.1. *Bridging Natural Language Processing and the Web of Data for Multimedia Question Answering*

**Participants:** Elena Cabrio, Fabien Gandon, Yoann Moise.

Differently from search engines, the goal of Question Answering (QA) is to return precise answers to users' natural language questions, extracting information from both documentary text and advanced media content. Up to now, QA research has largely focused on text, mainly targeting factual and list questions. The goal of our work was instead to exploit structured data and metadata describing multimedia content on Linked Open Data to provide a richer and more complete answer to the user, combining textual information with other media content.

We implemented an extension of our QAKiS system<sup>0</sup> to boost the answer visualization adding multimedia content. More specifically, once QAKiS outputs the textual answer(s) to the question asked by the user, the user can click on "more details" to have further information on the retrieved entity. Three main types of additional (and multimedia) content are then displayed: *i*) additional textual information providing a description of the retrieved entity (extracted from DBpedia), and a structured Information Card containing a set of relevant properties of the entity; *ii*) images (extracted from Flickr) and relevant videos (extracted from YouTube); *iii*) entity geo-localization: a pointer on a map is shown for questions asking about a place, together with its points of interest. The results of this research have been published at ESWC 2014 - Demo/poster paper [84].

### 6.5.2. *SMILK - Social Media Intelligence and Linked Knowledge*

**Participants:** Elena Cabrio, Fabien Gandon, Fabrice Jauvat.

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. The purpose of this LabCom is indeed 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.

---

<sup>0</sup><http://qakis.org>

As a first step in this direction, during the internship of Fabrice Jauvat we have developed a prototype of a system that - given free text (in particular in the cosmetics domain, extracted from a forum, a magazine, or a Web site) - can first recognize the named entities launching in parallel the RENCO system (developed by our partner in the LabCom), and NERD<sup>0</sup>, and then connect them to DBpedia, so that additional information on the entity can be extracted. For instance, if the word "J'adore" is detected in the text, it is recognized as a named entity and connected to its DBpedia page, so that information about the fact that it is a perfume, that its brand is Dior, and so on, can be automatically extracted and shown to the user.

### 6.5.3. *Ontology-Based and Natural Language Chatbot System in the Commercial Domain*

**Participants:** Amine Hallili, Catherine Faron Zucker, Elena Cabrio, Fabien Gandon.

This work is done within a Cifre PhD Thesis collocated in the Wimmics team and with SynchroNext Company located in Nice. The work consists in modelling and implementing ontology-based natural language Chatbot in commercial domain which consists of

- The design of a commercial knowledge base using the websites' APIs and web services (e.g. Amazon1, eBay2, BestBuy3),
- Interpreting and handling links between users' natural language questions by constructing relational graph,
- Generation and visualization of textual and media answers.

A. Hallili attended the ESSLLI summer school where a poster was accepted [93].

### 6.5.4. *Editor of Formal Lexicographic Definitions*

**Participants:** Maxime Lefrançois, Romain Gugert, Alain Giboin, Fabien Gandon.

Last year a prototype of a GUI of an editor of formal dictionary definitions aimed at lexicographers was developed based on the formalism of Units Graphs and on Meaning-Text Theory. This year, the prototype was demonstrated during the IC 2014 conference [60]. The prototype was also described in a paper reporting the knowledge engineering methodology for representing lexicographic definitions it supports [39].

## 6.6. Collaborative Software Development Platforms

**Participant:** Isabelle Mirbel.

Today's Web has given rise to several platforms serving the purpose of collaborative software development. Thanks to these environments, it is possible, among others, for anyone to suggest new requirements for a software under development. A lot of requirements are thus proposed by users and it becomes difficult, after a while, for the persons in charge of the software which development is hosted by the platform to understand this large set of new requirements in its entirety. Therefore we proposed a tool to make large sets of requirements posted on collaborative software development platforms better workable despite the poor content of requirement body. Our aim was to propose an approach to automatically group similar requirements together in order to propose a limited number of requirement categories, thus improving the review process. As requirements expressed on collaborative software development platforms are usually very short and their content not very structured, we proposed to exploit relationships between stakeholders and already processed requirements to break the whole set of new requirements into meaningful categories. Our tool relies on Semantic Web languages and Formal Concept Analysis to provide a 3 steps data analysis process. The data is first extracted from the platform and translated into RDF, then stakeholders' past activities are analyzed to finally get stakeholder categories in order to improve the review of newly posted requirements.

According to the experiments that we conducted, we noticed some limitations in our approach. When the contributing stakeholders are newbies with no previous participation in any blueprint or bug and when there is no sufficient number of evaluated blueprints or bugs. To cope with this limitation, we plan to evaluate stakeholders reputation by looking at their activities on the whole collaborative software development platform (and not only the project under consideration). The results of this research have been published in [23].

<sup>0</sup><http://nerd.eurecom.fr/>



In today's software development methodologies, User Stories (US) are mostly used as primary requirements artifacts. They are used to express requirements from a final user point of view and at a low abstraction basis using natural language. Over the years, several informal templates have been proposed by agile methods practitioners or academics to guide requirements gathering. Consequently, these templates are used in an ad-hoc manner, each modeler having idiosyncratic preferences. In this context, we performed a study of templates found in literature in order to propose a unified model. We also proposed an RDFS translation of this model in order to allow the annotation of user stories, thus providing search and reasoning capabilities to agile methods practitioners. The results of this research have been published in [56].

## 6.7. Workflow Management Systems

### 6.7.1. 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 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, in the first time, we 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. In the second time we propose a formal definition of semantic constraints and an  $O(n^3)$ -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 [45], [20], [46].

### 6.7.2. Extraction Mechanisms and Semantic Enrichment of Short Messages in Social Networks

**Participants:** Amosse Edouard, Nhan Le Thanh.

In this PhD, the work is focused on text processing in social networks and the main objectives are focused on the analysis of the spatial aspect, context enrichment and spatiotemporal analysis of short text messages.

During the first half of the year, we have mainly worked on positioning the research subject beside the state of the art as well as determining relevant domain. After analyzing several works on short text analysis in many domains such as Semantic Web, Data Mining and Natural Language Processing, we have identified a lack in the representation of the spatial aspect. Indeed, the spatial properties of items shared among online communities can be seen on three different aspects: i) The location of the resources which can be identified by its URI/URL, ii) The producer's location, iii) The location related to the content of the messages.

Most existent works have considered as identical the producer's location and the event described by the content of the message, which can lead to wrong results in many cases. For example, a user can be in the United States while describing an event in Africa. The SIOC ontology is one of the most known for representing items shared among online communities; we have proposed an extension of this ontology in which the spatial aspects are clearly represented. However, there exists a big challenge in finding the relevant location that can be associated to the content of a message. We are currently working on an approach that combines NLP technics and GIS to identify the spatial location of an item by analyzing its content.

### 6.7.3. Ontology-Based Workflow Management Systems

**Participants:** Tuan Anh Pham, Nhan Le Thanh.

The main objective of the 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 in their system. This customized workflow is called *Personalized workflow*. The challenges of a SWMS is to be simple, easy to use, friendly with the user and not too heavy. But it must have all functions of a WMS. There are three major challenge in this work: How to allow the users to customize the workflow template to correspond to their requirements, but their changes must be compliance with the predefined rules in the workflow template? How to build an execution model to evaluate step by step a personalized workflow ?



#### **6.7.4. Model Spatio-Temporal Dedicated Social Networks**

**Participants:** Amel Ben Othmane, Nhan Le Thanh.

The research aims of the work are to: i) model spatio-temporal, dedicated social networks using semantic web models (ontologies) taking into account spatial, temporal, social and dedicated dimensions. ii) overcome limitations of traditional Recommender Systems and improve the quality of recommendation by exploiting context (time, location, goal, etc.) and social ties.

The following tasks, proposed in the first year planning, are completed or almost finished, and are highly relevant to the current work, despite the different initial overall aim:

- Literature Review Report in Social networks, Spatio-temporal networks, Dedicated networks, Activities modeling techniques, Recommender systems.
- Elaboration of the requirements of an "ideal system" and presentation of an initial approach. The approach named the 5ws approach tries to answer those five questions: who must do, what, when, where and why?
- Implementation of the approach with Protégé.
- Extend semantic sensor network ontology to meet our requirements. In fact we use this ontology to enrich data from sensors networks which are used to measure different metrics describing physical activities (speed, heart rates, distance, etc.).

The following developments are ongoing: Adaptation of recommender systems for activities recommendation and reusing multi-dimensional recommendation model.

### **6.8. Modeling Team Processes**

#### **6.8.1. Modeling and Assessing Coordination Processes**

**Participants:** Alain Giboin, Isabelle Mirbel.

This work is done in collaboration with Pierre Robillard (Polytechnique Montréal).

Last year we worked on an assessment method of the quality of team dynamics based on a taxonomy of episodes of interactions encountered in software development teams [99] – the CoDyMA (Collaborative Dynamics Measurement and Analysis) method. Precisely, we proposed an analysis procedure of episodes based on the Formal Concept Analysis (FCA) approach. This year, we proposed to enrich the CoDyMA method with a procedure for assessing the quality of coordination interactions and the quality of coordination artifacts within a development team. The procedure is based on the "Coordinative Artifacts" Framework [101], [100].

#### **6.8.2. Modeling Multimodal Grounding Processes in Design Teams**

**Participant:** Alain Giboin.

This work is done with Aurore Defays (Université de Liège).

Grounding is the process used by participants to a collective activity to coordinate both the content and process of their communication to be successful [96]. Grounding is also defined as the process of elaborating and maintaining the Common Ground (i.e., mutual knowledge, mutual beliefs, and mutual assumptions) necessary to participants' mutual understanding [97]. Multimodal grounding is the process of grounding using several perceptual modalities. Last year we improved the methodology of analysis of multimodal grounding proposed in [98].

### **6.9. Semantic and Temporal Analysis of Online Communities**

**Participant:** Zide Meng.

The objective of the OCKTOPUS ANR project is to increase the potential social and economic benefit of the large and quickly growing amounts of user-generated content, by transforming it into useful knowledge. Since user communities are the basic of user-generated content sites, we start with community detection problem, which is a fundamental research point in social network analysis. Based on the preliminary experience from the previous year, we made several progress this year and published the results in international conferences, specifically:

- **Topic based interested group detection:**  
By analyzing a dataset extracted from the popular question answer site "StackOverflow", we proposed a heuristic method to enrich questions' tag. We also introduced a tag tree based model to extract topics from questions' tags, then we used the detected topics to label users in order to detect interest groups. We conducted experiments on the dataset and compared with related method. Results show that the proposed method is much simple and fast. This work has been published in [43].
- **Question Answer social media management**  
We proposed a question answer social media system based on social network analysis and social media mining to manage the two main resources in question answer sites: users and contents. We also presented a vocabulary used to formalize both the level of interest and the expertise of users on topics. We tested QASM on a dataset extracted from the popular "StackOverflow" site. We showed how the formalized knowledge is used to find relevant experts for a question. This work has been published in [95].

### 6.9.1. Temporal analysis in User and Topic:

We are planning to introduce temporal analysis into our research problem. According to the previous work, the potential direction could be topic evolution and user interest evolution. We believe this work could benefit community management in question answer sites, for example topic trend detection or user interest management.

## 6.10. P2P Media Streaming

**Participant:** Gaspard Perrot.

The *Heave-Ho* project won the Inria 2014 *Boost Your Code* contest.

The goal of the project is to design an overlay network for P2P media streaming based on new HTML5 technologies such as WebRTC. While conventional Internet applications encounter problems with scaling up as the number of visitors grows, the Heave-Ho project aims to enable website's users to share the resource directly among them. The proposed solution is a perfect fit for real-time video broadcasting. In traditional server/client architecture the server can only handle a limited number of requests; if there are too many clients, some of them will not have access to the video. Using a P2P system, the video can be broadcasted to more clients. The use of sharing techniques based on user location can also cut data transfer costs directly at ISP level, thereby reducing the risk of problems such as data rate limits<sup>0</sup>.

## 6.11. Discovery Hub

**Participants:** Nicolas Marie, Fabien Gandon, Emilie Palagi, Alain Giboin.

In the context of the Discovery Hub project<sup>0</sup>, we performed an ergonomic evaluation and redesign of the graphical and textual User Interfaces (UIs) displaying, and allowing to interact with, the explanations provided by Discovery Hub to justify the results it has retrieved. We did user tests of the existing UIs and designed several mockups improving the UIs by taking users' feedback into account.

<sup>0</sup><https://www.inria.fr/en/content/view/full/61889>

<sup>0</sup><http://www.discoveryhub.co>

We also performed a user-centered evaluation of the quality of the results retrieved by the 4 algorithms of Discovery Hub. We decided to focus on the quality of the results and not on the UI. Thus, specific criteria of the quality of the results were defined in this evaluation: the surpriseness and the interestingness of the results. A result is considered as:

- surprising if the user discovers an unknown resource or relation between the topic searched and the selected result, or if she discovers something unexpected ;
- interesting if the user thinks it is similar to the topic explored or if she thinks she will remind or reuse it later on.

We are currently developing an ergonomic method for evaluating exploratory search systems (ESSs) in general. We are performing a first test of the method on Discovery Hub [41].

## 6.12. Knowledge Graphs

### 6.12.1. SPARQL Template Transformation Language

**Participants:** Olivier Corby, Catherine Faron-Zucker Song.

We finalized the design and implementation of SPARQL Template Transformation Language<sup>0</sup> (STTL) [58], [73]. STTL is an extension of SPARQL with a `template { } where { }` clause that enables a transformation to generate presentation format for RDF. For example, it is possible to generate Turtle, OWL or SPIN-to-SPARQL syntax.

We designed a new service in Corese server that returns HTML. Using STTL transformations that generate HTML, we are able to set up light weight Semantic Web servers on top of local RDF Datasets or remote Datasets such as DBpedia<sup>0</sup>.

### 6.12.2. RDF Serialization and Introspection

**Participant:** Olivier Corby.

We started a work on RDF serialization of (Java) objects for Semantic Web system introspection. In conjunction with the overloading of SPARQL named graph pattern, we are able to query the system on several internal status such as graph index, triple provenance, property path triples, etc.

### 6.12.3. OWL 2 RL

**Participant:** Olivier Corby.

We dramatically optimized Corese Inference Rule Engine and we were able to run OWL 2 RL<sup>0</sup> rule base on the FMA ontology (Fundational Model of Anatomy<sup>0</sup>) with interesting performance. The initial OWL graph contains 1.74 million triples, the final graph contains 13.46 million triples and the rule engine runs in 3 minutes.

### 6.12.4. Rules for the Web of Data

**Participants:** Oumy Seye, Olivier Corby, Catherine Faron Zucker, Fabien Gandon.

This year we focused on the validation, the update of rules bases and the optimization of the reasoning. The goal of this work is to detect some inconsistencies in selected rule bases with respect to ontology and offer users to correct this. We built a set of SPARQL queries enabling (1) to build specific rule bases for a given context or application, (2) to optimize inference engines based on rule selection with respect to target RDF data sources, (3) to validate and update rule bases. We propose another optimization of the inference engines based on graph of rules dependencies and rules application ordering. This work is published in [62].

<sup>0</sup><http://ns.inria.fr/sparql-template/>

<sup>0</sup><http://corese.inria.fr>

<sup>0</sup>[http://www.w3.org/TR/owl2-profiles/#OWL\\_2\\_RL](http://www.w3.org/TR/owl2-profiles/#OWL_2_RL)

<sup>0</sup>[http://sig.biostr.washington.edu/projects/fma/release/v3.2.1/alt\\_formats.html](http://sig.biostr.washington.edu/projects/fma/release/v3.2.1/alt_formats.html)

### 6.12.5. KGRAM

**Participants:** Olivier Corby, Fuqi Song.

We received a two years grant from Inria to support the development of the Corese platform. This action aims at enhancing Corese software for conforming to latest W3C standards and facilitating its usage in distributed environment, we integrated several open source parsers to Corese, including JSON-LD, RDFa, TriG and N-Quads. Corese now is able to process RDF dataset in these formats. A Firefox extension called RDF Triple Collector (RTC) was developed, it can extract triple from web pages (annotated using RDFa), upload triples to Corese server and query data using SPARQL endpoint. A prototype of LDP 1.1 (Linked Data Platform) is implemented using RTC as data collector.

Besides, with the purpose of improving Corese query performance and carrying out research work on distributed environment, we proposed and developed a heuristic-based query planning method within Corese. The approach includes 3 main steps: 1) generate extended SPARQL query triple pattern Graph (ESG), 2) estimate the cost of ESG using pre-defined heuristics and cost models and 3) search ESG to find a good query plan and rewrite the SPARQL query. The approach was evaluated using BSBM benchmark, the results suggest that the developed method optimized 60% of the query execution time averagely [77].

## 6.13. Sociocultural Ontologies

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

### 6.13.1. Sociocultural Ontology : Upper-level and Domain Ontologies

We propose a process of sociocultural ontology development in order to promise and preserve the culture of a country through sharing the customs and history of different localities. This can be compared with the construction of a platform straddling "corporate memory" and a "social network", but applied in the context of a country. This process is based on the Vygotskian Framework, a theory of Russian psychologist Lev Vygotsky. We worked on an upper-level ontology and mapped it on the Linked Open Data (LOD) cloud. We designed a sociocultural domain ontology for the Senegalese context and the platform design on top of Semantic Mediawiki (SMW). This allows Senegalese communities to share and co-construct their sociocultural knowledge. This work is published in [59].

### 6.13.2. Human Time Ontology

In the second step of the PhD thesis of P. F. Diallo, we focus on the consideration of the time in the modeled knowledge. The main objective of this work is to provide a vocabulary (ontology) to handle temporal information on semantic data. Thus, the first step was to create a meta-language which handles temporal knowledge representation in the socio-cultural field that can be used in a wider area. This meta-language allows 1) to model cyclic knowledge (non-convex interval), 2) knowledge about calendar, 3) convex intervals 4) modeling absolute and relative time, 5) modeling relations between intervals, 6) distinction between open and closed intervals, 7) concepts such as time stamps and 8) to set different time granularities. The second step is to propose an RDFS representation of this meta-language. Thus this representation, Human Time Ontology (HuTO), allows us to model complex time statement which are a date, an interval (convex and non-convex), relative and absolute time. HuTO allows also temporal data annotation which is the representation of temporal notions on knowledge (expressed as RDF triple) and allow to reason over it.

HuTO allows us to use a resource as a temporal marker for dating another resource. Our ontology allows the relative dating which is to determine the relative order of resources, without necessarily determining their absolute time. A major contribution of HuTO is the modeling of non-convex intervals but also requests that can treat all types of intervals. For temporal annotation data, HuTO provides an approach that can link two models: one for temporal information and another for knowledge of the modeled area. This approach facilitates the information retrieval when it is on temporal or non-temporal data. Thus HuTO can annotate resources, triples or named graphs.

## ZENITH Project-Team

# 6. New Results

## 6.1. Highlights of the Year

- Patrick Valduriez received the 2014 Innovation Prize from Inria – Académie des sciences – Dassault Systems.
- Miguel Liroz-Gistau received the best presentation award from the Grid5000 Spring School 2014 in Lyon for his talk on “Using Grid5000 for MapReduce Experiments”.
- Triton, a new common lab. (i-lab) has been created between Zenith and Beepeers (beepeers.com) to work on a platform for developing social networks in mobile/Web environments.
- 127 research groups worldwide registered to the LifeCLEF 2014 evaluation campaign chaired by Alexis Joly.

## 6.2. Big Data Integration

### 6.2.1. Probabilistic Data Integration

**Participants:** Reza Akbarinia, Naser Ayat, Patrick Valduriez.

Data uncertainty in scientific applications can be due to many different reasons: incomplete knowledge of the underlying system, inexact model parameters, inaccurate representation of initial boundary conditions, inaccuracy in equipments, error in data entry, etc.

An important problem that arises in big data integration is that of Entity Resolution (ER). ER is the process of identifying tuples that represent the same real-world entity. The problem of *entity resolution over probabilistic data* (which we call ERPD) arises in many distributed application domains that have to deal with probabilistic data, ranging from sensor databases to scientific data management. The ERPD problem can be formally defined as follows. Let  $e$  be an uncertain entity represented by multiple possible alternatives, i.e., tuples, each with a membership probability. Let  $D$  be an uncertain database composed of a set of tuples each associated with a membership probability. Then, given  $e$ ,  $D$ , and a similarity function  $F$ , the problem is to find the entity-tuple pair  $(t, t_i)$  (where  $t \in e, t_i \in D$ ) such that  $(t, t_i)$  has the highest cumulative probability to be the most similar in all possible worlds. This entity-tuple pair is called the *most probable match pair* of  $e$  and  $D$ , denoted by  $MPMP(e, D)$ .

Many real-life applications produce uncertain data distributed among a number of databases. Dealing with the ERPD problem for distributed data is quite important for such applications. A straightforward approach for answering distributed ERPD queries is to ask all distributed nodes to send their databases to a central node that deals with the problem of ER by using one of the existing centralized solutions. However, this approach is very expensive and does not scale well neither in the size of databases, nor in the number of nodes.

In [24], we propose an efficient solution for the ERPD problem. Our contributions are summarized as follows. We adapted the possible worlds semantics of probabilistic data to define the problem of ERPD based on both similarity and probability of tuples. We proposed a PTIME algorithm for the ERPD problem. This algorithm is applicable to a large class of the similarity functions, where the similarity score of two tuples depends only on their attributes i.e., context-free functions. For the rest of similarity functions (i.e., context-sensitive), we proposed a Monte Carlo approximation algorithm. We also proposed a parallel version of our Monte Carlo algorithm using the MapReduce framework. We conducted an extensive experimental study to evaluate our approach for ERPD over both real and synthetic datasets. The results show the effectiveness of our algorithms.

Another topic of interest is the integration of large astronomy data catalogs. The main challenge in such integration, besides the huge amount of catalog data to be merged, is the weak identification of sky objects, which leads to ambiguities in object matching amongst catalogs. In cite [30], we present the NACluster algorithm. NACluster considers a Euclidian metric space and distance function to drive disambiguation amongst objects in various catalogs and extends the traditional k-means algorithm to deal with the dynamic creation of new clusters, representing real sky objects. NACluster shows F-measure results steadily superior to the Q3C join operator matching results, which is its closest competitor.

### 6.2.2. *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 SQLlike language, capable of querying multiple heterogeneous data stores (relational and NoSQL) within a single query that may contain embedded invocations to each data store's native query interface. Thus, CloudMdsQL unifies a quite diverse set of data management technologies while preserving the expressivity of their local query languages. 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 multidatabase query language.

### 6.2.3. *Semantic Data Integration using Bio-Ontologies*

**Participants:** Emmanuel Castanier, Patrick Valduriez.

Biologist have adopted ontologies for several reasons: (1) to provide canonical representation of scientific knowledge; (2) to annotate experimental data to enable interpretation, comparison, and discovery across databases; (3) to facilitate knowledge-based applications for decision support, natural language processing and data integration. The challenge is to automatically process complex databases and generate mappings using relevant ontologies in a way that scales up for many resources and ontologies, while being easy to use for the biomedical community, customizable to fit specific needs and smart, in order to leverage the knowledge contained in ontologies.

The National Center for Biomedical Ontology (NCBO) has developed a popular ontology-based annotation workflow. To address the above challenge, we have integrated the NCBO annotator with our WebSmatch tool and the Biosemantic tool from IRD to perform semantic annotation using bio-ontologies [47]. The resulting tool provides very useful capabilities. First, it can convert SQL database schemas to RDF/RDFS with Biosemantic. Second, it can annotate with the NCBO annotator and WebSmatch using the NCBO resources index. Third, the NCBO annotator relies on WebSmatch to create mappings between elements of schemas and ontological concepts, and uses ontologies properties (i.e. subsumption, transitivity) to enhance matching techniques.

Unlike the bio-medical domain which has accepted ontologies as a means to manage (integrate) knowledge, the agronomic sciences is yet to exploit its full potential. To this end, we are currently developing an RDF knowledge base, Agronomic Linked Data (AgroLD) [50]. The knowledge base is designed to integrate data from various publically available plant centric data sources. The aim of AgroLD project is to collaborate with domain experts in bridging the gap between technology and its potential users to enhance biological research.

## 6.3. Distributed Indexing and Searching

### 6.3.1. *Query Reformulation in P2P Data Management Systems*

**Participant:** Esther Pacitti.



We consider peer-to-peer data management systems (PDMS), where each peer maintains mappings between its schema and some acquaintances, along with social links with peer friends. In this context, we deal with reformulating conjunctive queries from a peer's schema into other peer's schemas. Precisely, queries against a peer node are rewritten into queries against other nodes using schema mappings thus obtaining query rewritings. Unfortunately, not all the obtained rewritings are relevant to a given query, as the information gain may be negligible or the peer is not worth exploring. On the other hand, the existence of social links with peer friends might be useful to get relevant rewritings.

In [19], we propose a new notion of "relevance" of a query with respect to a mapping that encompasses both a local relevance (the relevance of the query wrt. the mapping) and a global relevance (the relevance of the query wrt. the entire network). Based on this notion, we design a new query reformulation approach for social PDMS which achieves great accuracy and flexibility. We combine several techniques: (i) social links are expressed as FOAF (Friend of a Friend) links to characterize peer's friendship; (ii) concise mapping summaries are used to obtain mapping descriptions; (iii) local semantic views are special views that contain information about mappings captured from the network by using gossiping techniques. Our experimental evaluation, based on a prototype on top of PeerSim and a simulated network demonstrate that our solution yields greater recall, compared to traditional query translation approaches proposed in the literature.

### 6.3.2. *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). Different diversification techniques have been proposed to avoid redundancy in the process of recommendation. Intuitively, the goal of recommendation diversification is to identify a list of items that are dissimilar, but nonetheless relevant to the user's interests.

The main goal of this work [39], [17] is to define a new diversified search and recommendation solution suited for scientific data (i.e., plant phenotyping, botanical data). We first propose an original profile diversification scoring function that enables to address the problem of returning redundant items, and enhances the quality of diversification compared to the state-of-the-art solutions. We believe our work is the first to investigate profile diversity to address the problem of returning highly popular but too-focused items. Through experimental evaluation using two benchmarks we showed that our scoring function presents the best compromise between diversity and relevancy. Next, to implement our new scoring function, we propose a Top-k threshold-based algorithm that exploits a candidate list to achieve diversification. However this algorithm is greedy and does not scale up well. To overcome this limitation, we propose 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 therefore the number of diversified scores to compute. Third, we propose different indexing scores (i.e., the score used to sort the items in the inverted lists) that take into account the diversification of items, and using them, we developed an adaptive indexing approach to reduce the number of accesses in the index dynamically based on the queries workload. We evaluated the performance of our techniques through experimentation. The results show that they enable to reduce the response time up to 12 times compared to a baseline greedy diversification algorithm.

We also address the problem of distributed and diversified recommendation (P2P and multi-site) that fits very well in different application scenarios. We propose a new scoring function (usefulness) to cluster relevant users over a distributed overlay. We analyzed the new clustering algorithm in details, and we studied its behavior with an experimental evaluation using different datasets. Compared with state-of-the-art solutions, we obtain major gains in recall (order of 3 times).

## 6.4. Scientific Workflows

### 6.4.1. *Reuse of Scientific Workflows*

**Participant:** Sarah Cohen-Boulakia.



With the increasing popularity of scientific workflows, public and private repositories are gaining importance as a means to share, find, and reuse such workflows. As the sizes of workflows repositories grow, methods to compare the scientific workflows stored in them become a necessity, for instance, to allow duplicate detection or similarity search. Scientific workflows are complex objects, and their comparison entails a number of distinct steps from comparing atomic elements to comparison of the workflows as a whole. Various studies have implemented methods for scientific workflow comparison and came up with often contradicting conclusions upon which algorithms work best. Comparing these results is cumbersome, as the original studies mixed different approaches for different steps and used different evaluation data and metrics.

We first contribute to the field [26] by (i) comparing in isolation different approaches taken at each step of scientific workflow comparison, reporting on an number of unexpected findings, (ii) investigating how these can best be combined into aggregated measures, and (iii) making available a gold standard of over 2000 similarity ratings contributed by 15 workflow experts on a corpus of 1500 workflows and re-implementations of all methods we evaluated.

Then, we present a novel and intuitive workflow similarity measure that is based on layer decomposition [40]. Layer decomposition accounts for the directed dataflow underlying scientific workflows, a property which has not been adequately considered in previous methods. We comparatively evaluate our algorithm using our gold standard and show that it a) delivers the best results for similarity search, b) has a much lower runtime than other, often highly complex competitors in structure-aware workflow comparison, and c) can be stacked easily with even faster, structure-agnostic approaches to further reduce runtime while retaining result quality.

#### 6.4.2. 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 [46], 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.

Also, SWfMSs generally execute a scientific workflow in parallel within one site. In [38], we propose a non-intrusive approach to execute scientific workflows in a multisite cloud with three workflow partitioning techniques. We describe an experimental validation using an adaptation of Chiron SWfMS for Microsoft Azure multisite cloud. The experiment results reveal the efficiency of our partitioning techniques, and their superiority in different environments.

#### 6.4.3. 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 [20], 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.

## 6.5. Scalable Query Processing

### 6.5.1. Big Data Partitioning

**Participants:** Reza Akbarinia, Miguel Liroz, Esther Pacitti, Patrick Valduriez.

The amount of data that is captured or generated by modern computing devices has augmented exponentially over the last years. For processing this *big data*, parallel computing has been a major solution in both industry and research. This is why, the MapReduce framework, which provides automatic distribution parallelization and fault-tolerance in a transparent way over lowcost machines, has become one of the standards in big data analysis.

For processing a big dataset over a cluster of nodes, one main step is data partitioning (or fragmentation) to divide the dataset to the nodes. In [23], we consider applications with very large databases, where data items are continuously appended. Thus, the development of efficient data partitioning is one of the main requirements to yield good performance. In particular, this problem is harder in the case of some scientific databases, such as astronomical catalogs. The complexity of the schema limits the applicability of traditional automatic approaches based on the basic partitioning techniques. The high dynamicity makes the usage of graph-based approaches impractical, as they require to consider the whole dataset in order to come up with a good partitioning scheme. In our work, we propose *DynPart* and *DynPartGroup*, two dynamic partitioning algorithms for continuously growing databases [23]. These algorithms efficiently adapt the data partitioning to the arrival of new data elements by taking into account the affinity of new data with queries and fragments. In contrast to existing static approaches, our approach offers constant execution time, no matter the size of the database, while obtaining very good partitioning efficiency. We validate our solution through experimentation over real-world data; the results show its effectiveness.

### 6.5.2. Scalable Query Processing with Big Data

**Participants:** Reza Akbarinia, Miguel Liroz, Patrick Valduriez.

We address the problem of data skew in MapReduce parallel processing framework. There are many cases where because of skew intermediate data, a high percentage of processing in the reduce side of MapReduce is done by a few nodes, or even one node, while the others remain idle. There have been some attempts to address this problem of data skew, but only for specific cases. In particular, there is no solution when all or most of the intermediate values correspond to a single key, or to a set of keys that are fewer than the number of reduce workers.

In this work, we propose FP-Hadoop, a system that makes the reduce side of MapReduce more parallel, and can efficiently deal with the problem of reduce side data skew. We extended the programming model of MapReduce 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 intermediate reduce 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 the 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.

## 6.6. Data Stream Mining

### 6.6.1. Summarizing Uncertain Data Streams

**Participants:** Reza Akbarinia, Florent Masseglia.

In recent years, there has been a growing interest for probabilistic data management. In [41], we focus on probabilistic time series where a main characteristic is the high volumes of data, calling for efficient compression techniques. To date, most work on probabilistic data reduction has provided synopses that minimize the error of representation w.r.t. 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.

### 6.6.2. An Anti-Bouncing Data Stream Model

**Participant:** Florent Masseglia.

Usage mining is a significant research area with applications in various fields. However, Web usage data is usually considered streaming, due to its high volumes and rates. Because of these characteristics, we only have access, at any point in time, to a small fraction of the stream. When the data is observed through such a limited window, it is challenging to give a reliable description of the recent usage data. In [28] we show that data intralinkings, i.e., a usage record (event) may be associated with other records (events) in the same dataset, are common for Web usage streams. Therefore, in order to have a more authentic grasp of Web usage behaviors, the corresponding data stream models for Web usage streams should be able to process such intralinkings. We study the important consequences of the constraints and intralinkings, through the "bounce rate" problem and the clustering of usage streams. Then we propose the user-centric ABS (the Anti-Bouncing Stream) model which combines the advantages of previous models but avoids their drawbacks. First, ABS is the first data stream model that is able to seize the intralinkings between the Web usage records. It is also the first user-centric data stream model that can associate the usage records for the users in the Web usage streams. Second, owing to its simple but effective management principle, the data in ABS is available at any time for analysis. Under the same resource constraints as existing models in the literature, ABS can better model the recent data. Third, ABS can better measure the bounce rates for Web usage streams. We demonstrate its superiority through a theoretical study and experiments on two real-world data sets.

### 6.6.3. Autonomic Intrusion Detection: Adaptively Detecting Anomalies over Unlabeled Audit Data Streams

**Participant:** Florent Masseglia.

In [27], we propose a novel framework of autonomic intrusion detection that fulfills online and adaptive intrusion detection over unlabeled HTTP traffic streams in computer networks. The framework holds potential for self-managing: self-labeling, self-updating and self-adapting. Our framework employs the Affinity Propagation (AP) algorithm to learn a subject's behaviors through dynamical clustering of the streaming data. It automatically labels the data and adapts to normal behavior changes while identifying anomalies. Two large real HTTP traffic streams collected in our institute as well as a set of benchmark KDD'99 data are used to validate the framework and the method. The test results show that the autonomic model achieves better results in terms of effectiveness and efficiency compared to adaptive Sequential Karhunen-Loeve method and static AP as well as three other static anomaly detection methods, namely k-NN, PCA and SVM.

## 6.7. Scalable Data Analysis

### 6.7.1. Retrieval of Large-scale Visual Entities

**Participants:** Valentin Leveau, Alexis Joly, Patrick Valduriez.

In [37], we consider the problem of recognizing legal entities in visual contents in a similar way to named-entity recognizers for text documents. Whereas previous works were restricted to the recognition of a few tens of logotypes, we generalize the problem to the recognition of thousands of legal persons, each being modeled by a rich corporate identity automatically built from web images. We therefore introduce a new geometrically-consistent instance-based classification method that has several benefits over state-of-the-art instance classification methods including an efficient training phase reduced to a simple indexing process with a linear time and space complexity, but also the easy management of multi-labeled images, the fine grained localisation of the recognized patterns or the possibility of dynamically inserting additional training images in an incremental way. Experiments show that our method achieves better results than state-of-the-art techniques while being much more scalable, notably on an automatic web crawl of 5,824 legal entities which demonstrates the scalability of the approach.

### 6.7.2. Content-based Life Species Identification in Large Multimedia Collections

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

Building accurate knowledge of the identity, the geographic distribution and the evolution of living species is essential for a sustainable development of humanity as well as for biodiversity conservation. In this context, using crowdsourced data collection and multimedia identification tools is considered as one of the most promising solution. With the recent advances in digital devices/equipment, network bandwidth and information storage capacities, the production of multimedia data has indeed become an easy task. The emergence of citizen sciences and social networking tools has actually fostered the creation of large and structured communities of nature observers (e.g. e-bird, xeno-canto, Tela Botanica, etc.) who started to produce outstanding collections of multimedia records. Unfortunately, the performance of the state-of-the-art multimedia analysis techniques on such data is still not well understood and is far from reaching the real world's requirements in terms of identification tools. We therefore created LifeCLEF [36], [35], [31], [42], a new lab of the CLEF international forum<sup>0</sup> that evaluates these challenges in the continuity of the image-based plant identification task that we organized since 2011 within the ImageCLEF<sup>0</sup> lab. LifeCLEF is organized around 3 complementary tasks (PlantCLEF, BirdCLEF, FishCLEF), each being based on large and real-world data, as well as realistic scenarios established in collaboration with biologists and environmental stakeholders. 127 research groups worldwide did registered to the 2014 pilot campaign and downloaded the data. 22 of them crossed the finish line by submitting runs and papers to the workshop.

Besides the organization of the campaign, we also participated to two tasks in order to evaluate the content-based retrieval technologies developed within ZENITH. We notably implemented a new method [34] for the bird task based on the dense indexing of MFCC features and the offline pruning of the non-discriminant ones. To make such strategy scalable to the 30M of MFCC features extracted from the tens of thousands audio recordings of the training set, we used high dimensional hashing techniques coupled with an efficient approximate nearest neighbors search algorithm with controlled quality. Further improvements were obtained by (i) using a sliding classier with max pooling, (ii) weighting the query features according to their semantic coherence, and (iii) making use of the metadata to filter incoherent species. Results did show the effectiveness of the proposed technique which ranked 3rd among the 10 participating groups (some of them with years of experience in bioacoustic).

We finally investigated new interactive identification methods in [29], by extending classical faceted search mechanisms to the use of so called visual facets. The principle is to automatically build comprehensive visual illustrations of the expert data available in classical structured botanical dataset by building a visual matching graph of the related pictures and choosing the most connected ones. Additional facets can then be built automatically by clustering the graph and solving incompleteness issues.

<sup>0</sup><http://www.clef-initiative.eu/>

<sup>0</sup><http://www.imageclef.org/>

### **6.7.3. A look inside the *PI@ntNet* experience**

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

PI@ntNet is an innovative participatory sensing platform relying on image-based plants identification as a mean to enlist non-expert contributors and facilitate the production of botanical observation data [22]. 18 months after the public launch of the iOS public application (and 6 months after the release of the Android version [32]), we carried out a self-critical evaluation of the experience with regard to the requirements of a sustainable and effective ecological surveillance tool (to appear in *Multimedia Systems* journal). Thanks to usage data analytics, we first demonstrated the attractiveness of the developed multimedia system (with more than 300K end-users and several thousands of users daily) as well as the nice self-improving capacities of the whole collaborative workflow (1.5 millions of observations were collected). We also pointed out the current limitations of the approach towards producing timely and accurate distribution maps of plants at a very large scale. We discussed in particular two main issues:

1. **Data validation bottleneck:** within the current workflow, only a few percentage of the observations are validated to avoid submerging the volunteer experts who actively do this job thanks to the collaborative web tools. There is consequently a need of smarter task assignment and recommendation mechanisms that would better balance the collaborative workload across all users and improves the serendipity.
2. **Bias of the produced data:** The temporal and geographical distribution of the observations is highly correlated with human activity. High densities of observations are more determined by population density and humans behavior than by plants density. This issue inevitably arises in any participatory sensing system but when the objective is to monitor noise nuisance or air quality, the concentration of the observations in the cities is less critical. There is therefore a need to build new data analytics methods compensating the bias through long-term statistics and the use of contextual information.

## ALICE Project-Team

# 6. New Results

## 6.1. Highlights of the Year

*Fabrication:* We proposed a novel technique to automatically generate support structures for additive manufacturing with filament based processes. The deposited filament has to be properly supported at all times, which complicates printing of overhanging shapes: a disposable support has to be generated to temporarily hold the filament deposited above. Existing techniques either generate large structures, wasting material, or generate very thin structures that are hard to print and prone to failure. In contrast, our technique optimizes a scaffolding which is made of vertical pillars and horizontal bridges – such horizontal bridges print properly as long and the filament is deposited in straight line from one pillar to the next. We showed how to formulate scaffolding generation as a minimization problem and proposed a heuristic algorithm based on an efficient plane sweeping approach. The work was published [9] in ACM Transactions on Graphics in 2014 (proceedings of SIGGRAPH 2014). It is integrated within our 3D modeler for additive manufacturing, IceSL.

*Optimal transport:* this is an active research topics in the mathematics community. Given two measures  $\mu$  and  $\nu$ , optimal transport defines a distance between  $\mu$  and  $\nu$ , as the minimum cost of “morphing”  $\mu$  into  $\nu$ . This distance (called the *Wasserstein distance*) structures the space of measures and offers new ways of solving some highly non-linear PDEs (Monge-Ampere, Fokker-Plank ...). This requires a numerical way of computing the Wasserstein distance and its gradients. We studied a semi-discrete technique [21] submitted to ESAIM J. M2AN), that optimizes power diagrams. This is to our knowledge the first numerical implementation of optimal transport for volumetric densities (computes the Wasserstein distance between a sum of Dirac masses and a piece-wise linear density supported on a tetrahedral mesh).

## 6.2. New results

This year, we obtained new results in fabrication, in geometry processing and in multi-view reconstruction.

We investigated software solutions for printing with low cost (filament) 3D printers. We proposed a solution to automatically define temporary structures that will supports the object during its creation [9]. We also strongly reduce the artefacts that are produced by multi-material printing [17]. These works allow to better understand the physics of these printers, and to come up with efficient software solutions to common drawbacks of this technology. Other contributions in fabrication are more related to the design of the printable objects, that is developed in our software IceSL. To achieve real-time rendering of CSG models, we developed a new GPU approach for single pass A-Buffer [23]. This technique is also a simple solution to handle complex rendering problems such as transparency. We also proposed [11] an efficient method for performing dilatation and erosion directly on the same representation of volume by sequence of dilatation and erosions on segments.

In geometry processing, we proposed an algorithm to compute the intersection of Voronoi cells and a simplicial complex [25]. This algorithm is fast in dimension up to  $10D$  because it doesn't require to explicit the Voronoi diagram. It comes with exact predicates and symbolic perturbation to ensure its robustness. We have also developed an algorithm [13] able to trace streamlines on triangulated surfaces in such a way that two such streamlines cannot cross or merge. This property seems obvious in the continuous case, but was very difficult to enforce with the discrete representations (triangulated surface, and floating points) manipulated by the computer. We did also revisit the Optimal Delaunay triangulation in the case of graded mesh generation [14], and we adapted our remeshing methods to Geologic applications [27].

We obtained some new results in multi-view reconstruction: a new method that expands a limited set of correspondences towards a quasi-dense map across two views [15], and an improvement of variational multi-view reconstruction obtained thanks to a simple characterization of geometric deformations [16].



## AVIZ Project-Team

## 6. New Results

### 6.1. Highlights of the Year

We had a number of highlights this year:

- Jean-Daniel Fekete was General Chair of the **IEEE VIS 2014** conference, organized for the first time ever outside of the USA, in Paris, with a record attendance.
- Aviz presented 7 articles at the IEEE VIS 2014 conference, and co-organized 3 workshops.
- Five PhD students defended this year.
- Benjamin Bach was awarded the second price in the IEEE VGTC Doctoral Dissertation Competition for his thesis "Connections, Changes, Cubes: Unfolding Dynamic Networks for Visual Exploration" [10].
- Yvonne Jansen was awarded the second price for the Gilles Kahn dissertation award for her thesis "Physical and Tangible Information Visualization" [11].
- Samuel Huron received the best paper honorable mention award at DIS 2014 for the paper "Constructive Visualization" [28].

### 6.2. Effectiveness of Staggered Animations

**Participants:** Fanny Chevalier, Pierre Dragicevic [correspondant], Steven Franconeri.

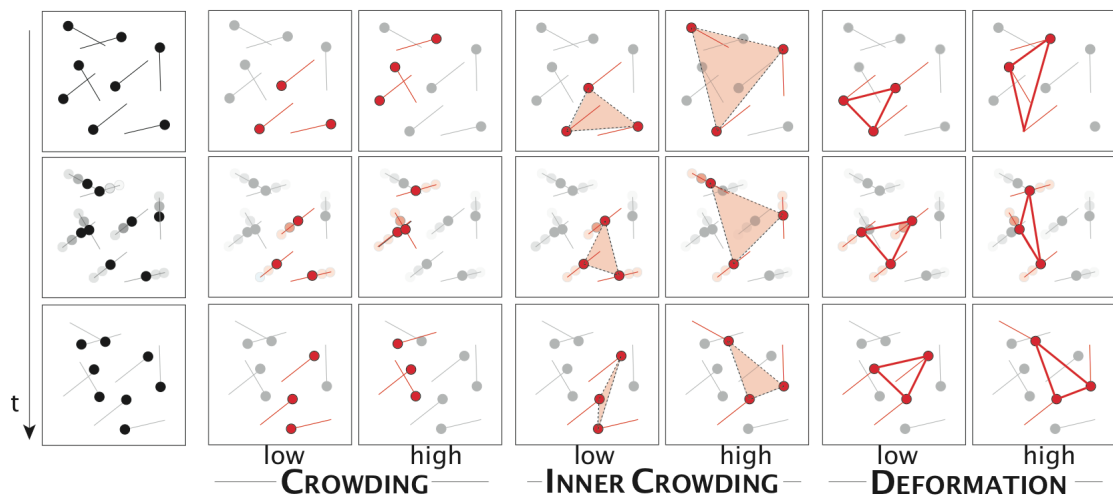


Figure 7. Illustration of the complexity metrics used in the study.



Interactive visual applications often rely on animation to transition from one display state to another. There are multiple animation techniques to choose from, and it is not always clear which should produce the best visual correspondences between display elements. One major factor is whether the animation relies on staggering—an incremental delay in start times across the moving elements. It has been suggested that staggering may reduce occlusion, while also reducing display complexity and producing less overwhelming animations, though no empirical evidence has demonstrated these advantages. Work in perceptual psychology does show that reducing occlusion, and reducing inter-object proximity (crowding) more generally, improves performance in multiple object tracking.

We empirically investigated the effectiveness of staggering [15]. We ran simulations confirming that staggering can in some cases reduce crowding in animated transitions involving dot clouds (as found in, e.g., animated 2D scatterplots). We empirically evaluated the effect of two staggering techniques on tracking tasks, focusing on cases that should most favour staggering. We found that introducing staggering has a negligible, or even negative, impact on multiple object tracking performance. The potential benefits of staggering may be outweighed by strong costs: a loss of common-motion grouping information about which objects travel in similar paths, and less predictability about when any specific object would begin to move. Staggering may be beneficial in some conditions, but they have yet to be demonstrated. Our results are a significant step toward a better understanding of animation pacing, and provide direction for further research.

More on the project Web page: [fannychevalier.net/animations](http://fannychevalier.net/animations)

### 6.3. Tablet-Based Interaction for Immersive 3D Data Exploration

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

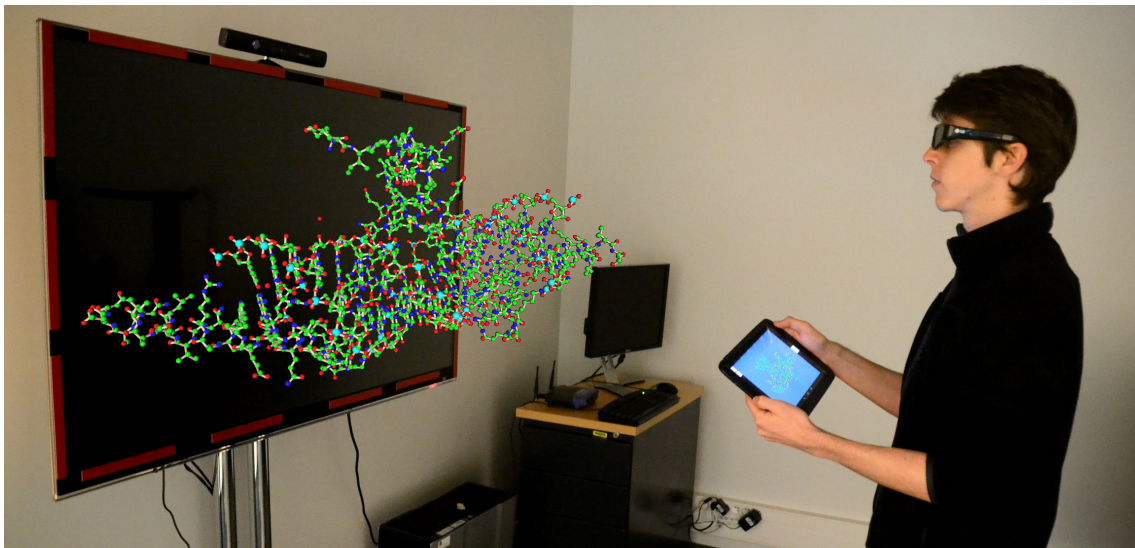


Figure 8. Illustration of the interaction setup for a combined touch-based navigation and stereoscopic viewing of 3D data.

We examined touch-based navigation of 3D visualizations in a combined monoscopic and stereoscopic viewing environment [32] (see Figure 8). We identified a set of interaction modes, and a workflow that helps users transition between these modes to improve their interaction experience. For this purpose we analyzed, in particular, the control-display space mapping between the different reference frames of the stereoscopic and

monoscopic displays. We showed how this mapping supports interactive data exploration, but may also lead to conflicts between the stereoscopic and monoscopic views due to users' movement in space; we resolved these problems through synchronization. To support our discussion, we conducted an exploratory observational evaluation with domain experts in fluid mechanics and structural biology. These experts explored domain-specific datasets using variations of a system that embodies the interaction modes and workflows; we could report on their interactions and qualitative feedback on the system and its workflow.

## 6.4. Understanding the Perception of Star Glyphs

**Participants:** Johannes Fuchs, Petra Isenberg [correspondant], Anastasia Bezerianos, Fabian Fischer, Enrico Bertini.

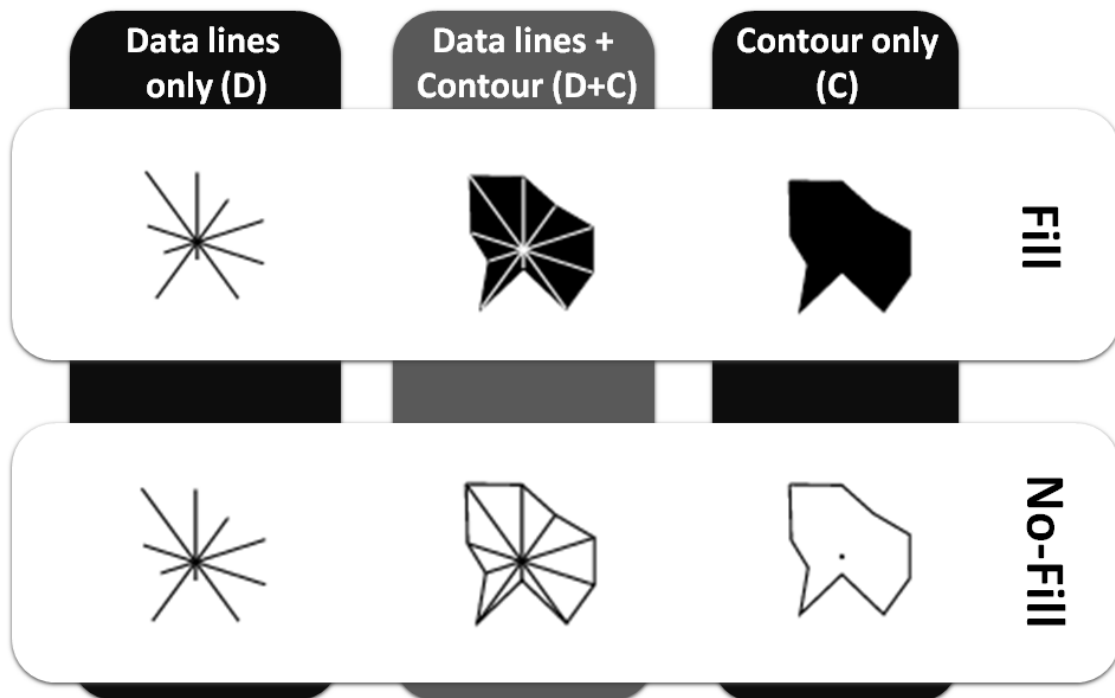


Figure 9. Illustration of the design space of the perception study.

We conducted three experiments to investigate the effects of contours on the detection of data similarity with star glyph variations [16]. A star glyph is a small, compact, data graphic that represents a multi-dimensional data point. Star glyphs are often used in small-multiple settings, to represent data points in tables, on maps, or as overlays on other types of data graphics. In these settings, an important task is the visual comparison of the data points encoded in the star glyph, for example to find other similar data points or outliers. We hypothesized that for data comparisons, the overall shape of a star glyph—enhanced through contour lines—would aid the viewer in making accurate similarity judgments. To test this hypothesis, we conducted three

experiments. In our first experiment, we explored how the use of contours influenced how visualization experts and trained novices chose glyphs with similar data values. Our results showed that glyphs without contours make the detection of data similarity easier. Given these results, we conducted a second study to understand intuitive notions of similarity. Star glyphs without contours most intuitively supported the detection of data similarity. In a third experiment, we tested the effect of star glyph reference structures (i.e., tickmarks and gridlines) on the detection of similarity. Surprisingly, our results show that adding reference structures does improve the correctness of similarity judgments for star glyphs with contours, but not for the standard star glyph. As a result of these experiments, we conclude that the simple star glyph without contours performs best under several criteria, reinforcing its practice and popularity in the literature. Contours seem to enhance the detection of other types of similarity, e.g., shape similarity and are distracting when data similarity has to be judged. Based on these findings we provide design considerations regarding the use of contours and reference structures on star glyphs.

## 6.5. Constructive Visualization

**Participants:** Samuel Huron [correspondant], Yvonne Jansen, Sheelagh Carpendale.



*Figure 10. Constructing a visualization with tokens: right hand positions tokens, left hand points to the corresponding data.*

The accessibility of infovis authoring tools to a wide audience has been identified as one of the major research challenges. A key task of the authoring process is the development of visual mappings. While the infovis community has long been deeply interested in finding effective visual mappings, comparatively little attention has been placed on how people construct visual mappings. We conducted a study designed to shed light on how people spontaneously transform data into visual representations [18]. We asked people to create, update and explain their own information visualizations using simple materials such as tangible building blocks. We learned that all participants, most of whom had no experience in visualization, were readily able to create and talk about their own visualizations. On the basis of our observations, we discussed the actions of our participants in the context of the development of their visual representations and their analytic activities. From this we suggested implications for tool design that can enable broader support for infovis authoring.

More on the project Web page: [constructive.gforge.inria.fr](http://constructive.gforge.inria.fr)

## 6.6. Multi-touch Gestures for Data Graphics

**Participants:** Wesley Willett, Qi Lan, Petra Isenberg.

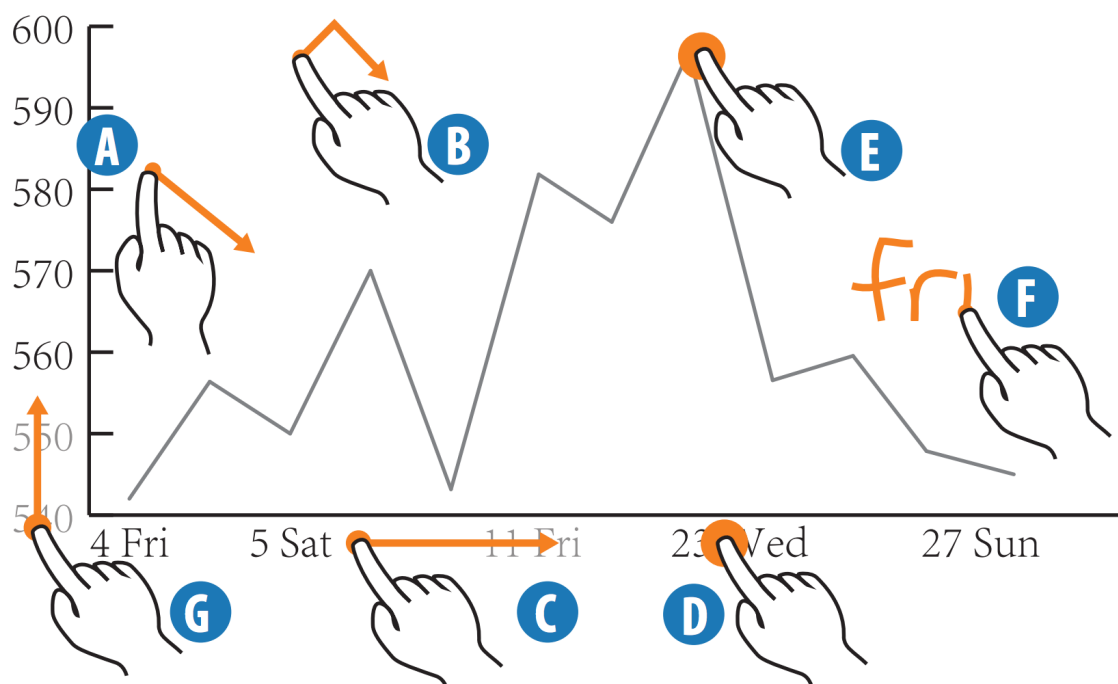


Figure 11. The most common gestures used for selecting (a) downward trends, (b) peaks, (c) ordinal ranges, (d) non-contiguous items, (e) highest points, (f) repeating dates, and (g) the lowest three points in a time series chart.

Selecting data items is a common and extremely important form of interaction with data graphics, and serves as the basis for many other data interaction techniques. However, interactive charting tools for multi-touch displays typically only provide dedicated multi-touch gestures for single-point selection or zooming. We conducted a study in which we used gesture elicitation to explore a wider range of possible selection interactions for multi-touch data graphics [35]. The results show a strong preference for simple, one-handed selection gestures. They also show that users tend to interact with chart axes and make figurative selection gestures outside the chart, rather than interact with the visual marks themselves. Finally, we found strong consensus around several unique selection gestures related to visual chart features.

## 6.7. Exploring Word-Scale Visualizations

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

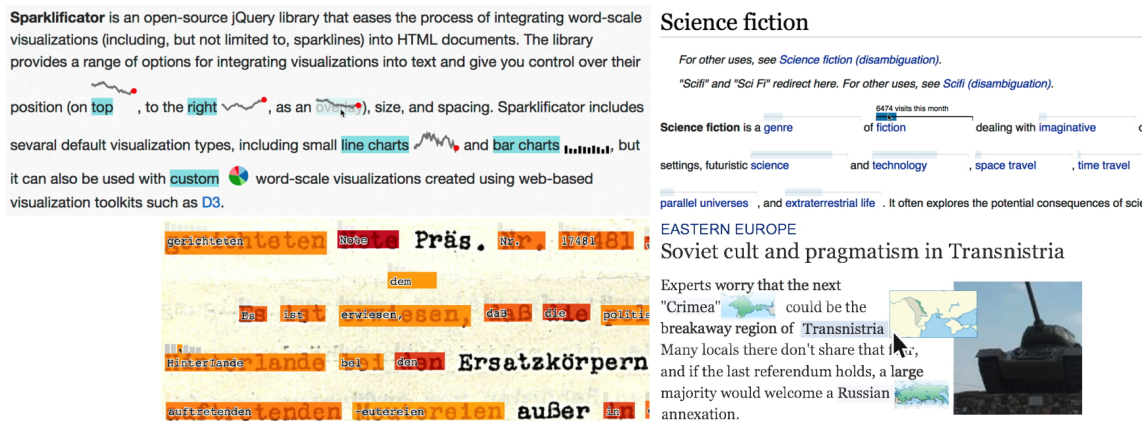


Figure 12. Four examples of the integration of word-scale visualizations into HTML documents

We presented an exploration and a design space that characterize the usage and placement of word-scale visualizations within text documents [17]. Word-scale visualizations are a more general version of sparklines—small, word-sized data graphics that allow meta-information to be visually presented in-line with document text. In accordance with Edward Tufte’s definition, sparklines are traditionally placed directly before or after words in the text. We described alternative placements that permit a wider range of word-scale graphics and more flexible integration with text layouts. These alternative placements include positioning visualizations between lines, within additional vertical and horizontal space in the document, and as interactive overlays on top of the text. Each strategy changes the dimensions of the space available to display the visualizations, as well as the degree to which the text must be adjusted or reflowed to accommodate them. We provided an illustrated design space of placement options for word-scale visualizations and identify six important variables that control the placement of the graphics and the level of disruption of the source text. We also contributed a quantitative analysis that highlights the effect of different placements on readability and text disruption. Finally, we used this analysis to propose guidelines to support the design and placement of word-scale visualizations.

More on the project Web page: [www.aviz.fr/sparklificator](http://www.aviz.fr/sparklificator)

## 6.8. Assessing Visualization Literacy

**Participants:** Jeremy Boy, Ronald A. Rensink, Enrico Bertini, Jean-Daniel Fekete.

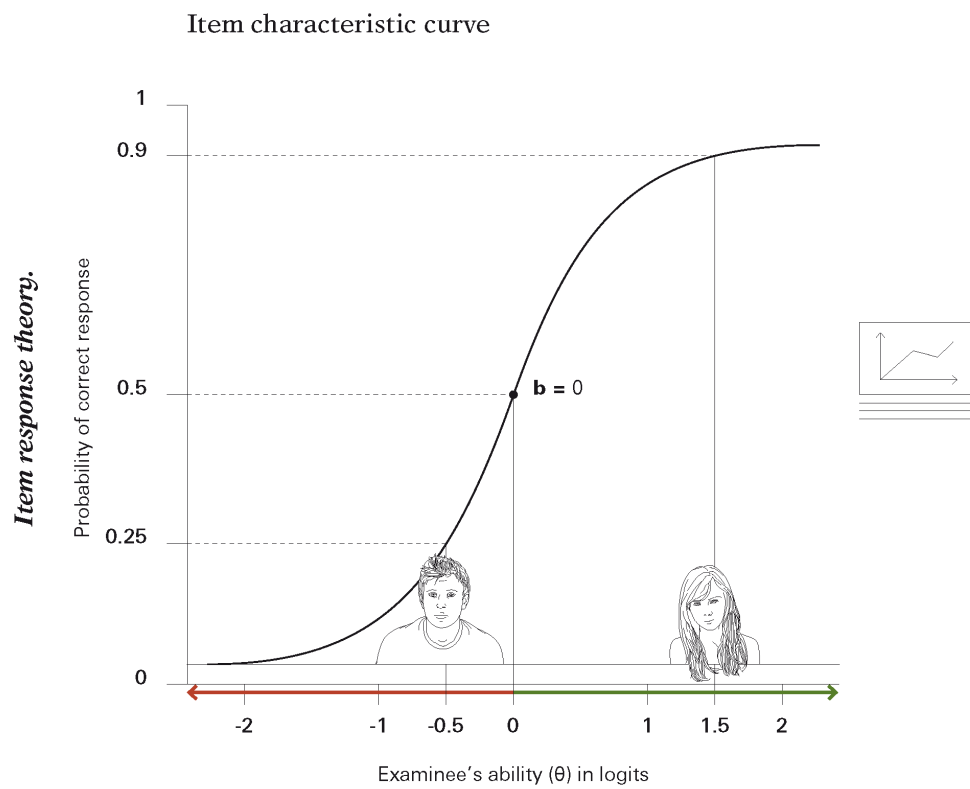


Figure 13. Example of an Item Characteristic Curve, and how people's abilities are plotted against a test-item's difficulty to determine probability of success.



We described a method for assessing the visualization literacy (VL) of a user [14]. Assessing how well people understand visualizations has great value for research (e.g., to avoid confounds), for design (e.g., to best determine the capabilities of an audience), for teaching (e. g., to assess the level of new students), and for recruiting (e. g., to assess the level of interviewees). In this project we proposed a method for assessing VL based on Item Response Theory. We described the design and evaluation of two VL tests for line graphs, and presents the extension of the method to bar charts and scatterplots. Finally, we discussed the reimplementations of these tests for fast, effective, and scalable web-based use.

More on the project Web page: [peopleviz/vLiteracy/home](http://peopleviz/vLiteracy/home).

## HYBRID Project-Team

## 6. New Results

### 6.1. Highlights of the Year

- Paper [22] from Merwan Achibet, Maud Marchal, Ferran Argelaguet and Anatole Lécuyer received the "Best Paper Award" at IEEE Symposium on 3D User Interfaces 2014 (IEEE 3DUI'14).
- Paper [26] from Jean-Baptiste Barreau, Valérie Gouranton received the "Third Best Poster Award" at International Conference on Cultural Heritage 2014.

BEST PAPERS AWARDS :

[22] **IEEE Symposium on 3D User Interfaces**. M. ACHIBET, M. MARCHAL, F. ARGELAGUET SANZ, A. LÉCUYER.

### 6.2. 3D User Interfaces

#### 6.2.1. 3D manipulation of virtual objects

**Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube**  
Maud Marchal, Collaboration with REVES

We have proposed a solution for interaction using finger tracking in a cubic immersive virtual reality system (or immersive cube) [13]. Rather than using a traditional wand device, users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks. Our solution couples finger tracking with a real-time physics engine, combined with a heuristic approach for hand manipulation, which is robust to tracker noise and simulation instabilities. A first study has been performed to evaluate our interface, with tasks involving complex manipulations, such as balancing objects while walking in the cube. The users finger-tracked manipulation was compared to manipulation with a 6 degree-of-freedom wand (or flystick), as well as with carrying out the same task in the real world. Users were also asked to perform a free task, allowing us to observe their perceived level of presence in the scene. Our results show that our approach provides a feasible interface for immersive cube environments and is perceived by users as being closer to the real experience compared to the wand. However, the wand outperforms direct manipulation in terms of speed and precision. We conclude with a discussion of the results and implications for further research

**A New Direct Manipulation Technique for Immersive 3D Virtual Environments** Thi Thuong Huyen Nguyen, Thierry Duval, Collaboration with MIMETIC

We have introduced a new 7-Handle manipulation technique [38] for 3D objects in immersive virtual environments and its evaluation. The 7-Handle technique includes a set of seven points which are flexibly attached to an object. There are three different control modes for these points including configuration, manipulation and locking/ unlocking modes. We have conducted an experiment to compare the efficiency of this technique with the traditional 6-DOF direct manipulation technique in terms of time, discomfort metrics and subjective estimation for precise manipulations in an immersive virtual environment in two consecutive phases: an approach phase and a refinement phase. The statistical results showed that the completion time in the approach phase of the 7-Handle technique was significantly longer than the completion time of the 6-DOF technique. Nevertheless, we found a significant interaction effect between the two factors (the manipulation technique and the object size) on the completion time of the refinement phase. In addition, even though we did not find any significant differences between the two techniques in terms of intuitiveness, ease of use and global preference in the result of subjective data, we obtained a significantly better satisfaction feedback from the subjects for the efficiency and fatigue criteria.

**A survey of plasticity in 3D user interfaces** Jérémy Lacoche, Thierry Duval, Bruno Arnaldi, Collaboration with b<>com

Plasticity of 3D user interfaces [33] refers to their capabilities to automatically fit to a set of hardware and environmental constraints. This area of research has already been deeply explored in the domain of traditional 2D user interfaces. Besides, during the last decade, interest in 3D user interfaces has grown. Designers find with 3D user interfaces new ways to promote and interact with data, such as e-commerce websites, scientific data visualization, etc. Because of the wide variety of Virtual Reality (VR) and Augmented Reality (AR) applications in terms of hardware, data and target users, there is a real interest in solutions for automatic adaptation in order to improve the user experience in any context while reducing the development costs. An adaptation is performed in reaction to different criteria defining a system such as the targeted hardware platform, the user's context and the structure and the semantic of the manipulated data. This adaptation can then impact the system in different ways, especially content presentation, interaction techniques modifications and eventually the current distribution of the system across a set of available devices. In [33] we present the state of the art about plastic 3D user interfaces. Moreover, we present well known methods in the field of 2D user interfaces that could become relevant for 3D user interfaces.

### 6.2.2. Navigating in virtual environments

#### **Adaptive Navigation in Virtual Environments** Ferran Argelaguet

Navigation speed for most navigation interfaces is still determined by rate-control devices (e.g. joystick). The interface designer is in charge of adjusting the range of optimal speeds according to the scale of the environment and the desired user experience. However, this approach is not valid for complex environments (e.g. multi-scale environments). Optimal speeds might vary for each section of the environment, leading to non-desired side effects such as collisions or simulator sickness. Thereby, we proposed a speed adaptation algorithm [24] based on the spatial relationship between the user and the environment and the user's perception of motion. The computed information is used to adjust the navigation speed in order to provide an optimal navigation speed and avoid collisions. Two main benefits of our approach is firstly, the ability to adapt the navigation speed in multi-scale environments and secondly, the capacity to provide a smooth navigation experience by decreasing the jerkiness of described trajectories. The evaluation showed that our approach provides comparable performance as existing navigation techniques but it significantly decreases the jerkiness of described trajectories

### 6.2.3. Novel pseudo-haptic based interfaces

#### **Toward "Pseudo-Haptic Avatars": Modifying the Visual Animation of Self-Avatar Can Simulate the Perception of Weight Lifting.** Ferran Argelaguet, Anatole Lécuyer, Collaboration with MIMETIC

We have studied how the visual animation of a self-avatar can be artificially modified in real-time in order to generate different haptic perceptions [18]. In our experimental setup, participants could watch their self-avatar in a virtual environment in mirror mode while performing a weight lifting task. Users could map their gestures on the self-animated avatar in real-time using a Kinect. We introduce three kinds of modification of the visual animation of the self-avatar according to the effort delivered by the virtual avatar: 1) changes on the spatial mapping between the user's gestures and the avatar, 2) different motion profiles of the animation, and 3) changes in the posture of the avatar (upper-body inclination). The experimental task consisted of a weight lifting task in which participants had to order four virtual dumbbells according to their virtual weight. The user had to lift each virtual dumbbells by means of a tangible stick, the animation of the avatar was modulated according to the virtual weight of the dumbbell. The results showed that the altering the spatial mapping delivered the best performance. Nevertheless, participants globally appreciated all the different visual effects. Our results pave the way to the exploitation of such novel techniques in various VR applications such as sport training, exercise games, or industrial training scenarios in single or collaborative mode.

#### **The Virtual Mitten: A Novel Interaction Paradigm for Visuo-Haptic Manipulation of Objects Using Grip Force** Merwan Achibet, Maud Marchal, Ferran Argelaguet, Anatole Lécuyer

We have proposed a novel visuo-haptic interaction paradigm called the "Virtual Mitten" [22] for simulating the 3D manipulation of objects. Our approach introduces an elastic handheld device that provides a passive haptic feedback through the fingers and a mitten interaction metaphor that enables to grasp and manipulate

objects. The grasping performed by the mitten is directly correlated with the grip force applied on the elastic device and a supplementary pseudo-haptic feedback modulates the visual feedback of the interaction in order to simulate different haptic perceptions. The Virtual Mitten allows natural interaction and grants users with an extended freedom of movement compared with rigid devices with limited workspaces. Our approach has been evaluated within two experiments focusing both on subjective appreciation and perception. Our results show that participants were able to well perceive different levels of effort during basic manipulation tasks thanks to our pseudo-haptic approach. They could also rapidly appreciate how to achieve different actions with the Virtual Mitten such as opening a drawer or pulling a lever. Taken together, our results suggest that our novel interaction paradigm could be used in a wide range of applications involving one or two-hand haptic manipulation such as virtual prototyping, virtual training or video game.



Figure 2. The Virtual Mitten : Visuo-haptic manipulation as enabled by our novel approach called the “Virtual Mitten”. Each hand holds an elastic device to control a virtual mitten (in grey) enabling the grasp of virtual objects.

**Collaborative Pseudo-Haptics: Two-User Stiffness Discrimination Based on Visual Feedback** Ferran Argelaguet, Takuya Sato, Thierry Duval, Anatole Lécuyer, Collaboration with Tohoku University Research Institute of Electrical Communication

We have explored how the concept of pseudo-haptic feedback can be introduced in a collaborative scenario [25]. A remote collaborative scenario in which two users interact with a deformable object is presented. Each user, through touch-based input, is able to interact with a deformable virtual object displayed in a standard display screen. The visual deformation of the virtual object is driven by a pseudo-haptic approach taking into account both the user in-put and the simulated physical properties. Particularly, we investigated stiffness perception. In order to validate our approach, we tested our system in a single and two-user configuration. The results showed that users were able to discriminate the stiffness of the virtual object in both conditions with a comparable performance. Thus, pseudo-haptic feedback seems a promising tool for providing multiple users with physical information related to other users’ interactions.

#### 6.2.4. Sound and virtual reality

**Sonic interaction with a virtual orchestra of factory machinery** Florian Nouviale, Valérie Gouranton, collaboration with Ronan Gagne (IMMERSIA) and LIMSI

We have conceived an immersive application where users receive sound and visual feedbacks on their interactions with a virtual environment. In this application, the users play the part of conductors of an orchestra of factory machines since each of their actions on interaction devices triggers a pair of visual and audio responses. Audio stimuli were spatialized around the listener. The application was exhibited during the 2013 Science and Music day and designed to be used in a large immersive system with head tracking, shutter glasses and a 10.2 loudspeaker configuration [43].

**Audio-Visual Attractors for Capturing Attention to the Screens When Walking in CAVE Systems** Ferran Argelaguuet, Valérie Gouranton, Anatole Lécuyer, collaboration with Aalborg University

In four-sided CAVE-like VR systems, the absence of the rear wall has been shown to decrease the level of immersion and can introduce breaks in presence. We have therefore investigated to which extent user's attention can be driven by visual and auditory stimuli in a four-sided CAVE-like system [32]. An experiment was conducted in order to analyze how user attention is diverted while physically walking in a virtual environment, when audio and/or visual attractors are present. The four sided CAVE used in the experiment allowed to walk up to 9m in straight line. An additional key feature in the experiment is the fact that auditory feedback was delivered through binaural audio rendering techniques via non-personalized head related transfer functions (HRTFs). The audio rendering was dependent on the user's head position and orientation, enabling localized sound rendering. The experiment analyzed how different "attractors" (audio and/or visual, static or dynamic) modify the user's attention. The results of the conducted experiment show that audio-visual attractors are the most efficient attractors in order to keep the user's attention toward the inside of the CAVE. The knowledge gathered in the experiment can provide guidelines to the design of virtual attractors in order to keep the attention of the user and avoid the "missing wall".

### 6.2.5. *Experiencing the past in virtual reality*

**Immersia, an open immersive infrastructure: doing archaeology in virtual reality** Valérie Gouranton, Bruno Arnaldi, collaboration with MIMETIC and Ronan Gagne (IMMERSIA)

We have first studied the mutual enrichment between archaeology and virtual reality [16]. To do so, we are considering Immersia, our open high-end platform dedicated to research on immersive virtual reality and its usages. Immersia is a node of the european project Visionair that offers an infrastructure for high level visualisation facilities open to research communities across Europe. In Immersia, two projects are currently active on the theme of archaeology. One is relative to the study of the Cairn of Carn, with the Creahh, a pluridisciplinary research laboratory of archeology and archeosciences, and one on the reconstitution of the gallo-roman villa of Bais, with the French institute INRAP.

**Virtual reality tools for the West Digital Conservatory of Archaeological Heritage** Jean-Baptiste Barreau, Valérie Gouranton, collaboration with Ronan Gagne (IMMERSIA) and INRAP

In continuation of the 3D data production work made by the WDCAH (West Digital Conservatory of Archaeological Heritage), the use of virtual reality tools allows archaeologists to carry out analysis and understanding research about their sites. We have then focused on the virtual reality services proposed to archaeologists in the WDCAH, through the example of two archaeological sites, the Temple de Mars in Corseul and the Cairn of Carn Island [27].

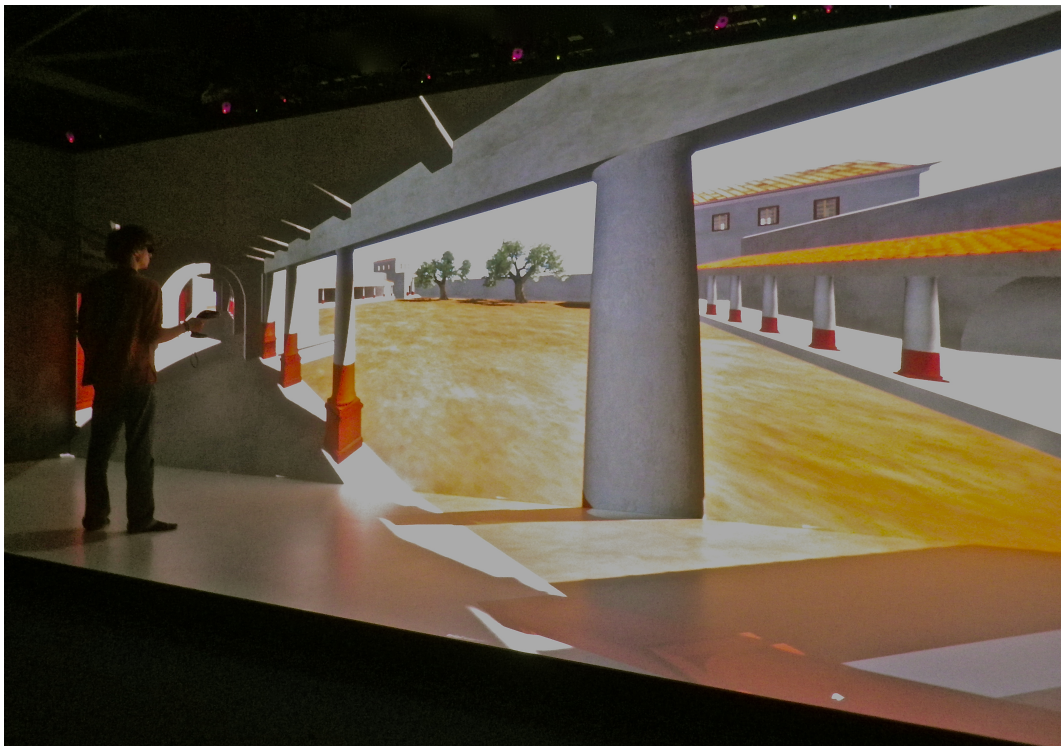
**Preservative Approach to Study Encased Archaeological Artefacts** Valérie Gouranton, Bruno Arnaldi, collaboration with Ronan Gagne (IMMERSIA) and INRAP

We have proposed a workflow based on a combination of computed tomography, 3D images and 3D printing to analyse different archaeological material dating from the Iron Age, a weight axis, a helical piece, and a fibula [39]. This workflow enables a preservative analysis of the artefacts that are unreachable because encased either in stone, corrosion or ashes. Computed tomography images together with 3D printing provide a rich toolbox for archaeologist work allowing to access a tangible representation of hidden artefacts. These technologies are combined in an efficient, affordable and accurate workflow compatible with preventive archaeology constraints.

**Combination of 3D Scanning, Modeling and Analyzing Methods around the Castle of Coatfrec Reconstitution** Jean Baptiste Barreau, Valérie Gouranton, collaboration with Ronan Gagne (IMMERSIA) and INRAP

The castle of Coatfrec is a medieval castle in Brittany constituting merely a few remaining ruins currently in the process of restoration. Beyond its great archeological interest, it has become, over the course of the last few years, the subject of experimentation in digital archeology. Methods of 3D scanning were used in order to gauge comparisons between the remaining structures and their absent hypothetical ones, resulting in the first





*Figure 3. Virtual visit of the "Temple of Mars" in the Immersia room.*



quantitative results of its kind. We have applied these methods and presented the subsequent results obtained using these new digital tools [26].

**Ceramics Fragments Digitization by Photogrammetry, Reconstructions and Applications** Jean Baptiste Barreau, Valérie Gouranton, collaboration with Ronan Gaugne (IMMERSIA) and INRAP

We have studied an application of photogrammetry on ceramic fragments from two excavation sites located north-west of France [28]. The restitution by photogrammetry of these different fragments allowed reconstructions of the potteries in their original state or at least to get to as close as possible. We used the 3D reconstructions to compute some metrics and to generate a presentation support by using a 3D printer. This work is based on affordable tools and illustrates how 3D technologies can be quite easily integrated in archaeology process with limited financial resources.

## 6.3. Physical simulation and multisensory feedback

### 6.3.1. Physically-based simulation and collision detection

**Fast collision detection for fracturing rigid bodies** Loeiz Glondu, Maud Marchal

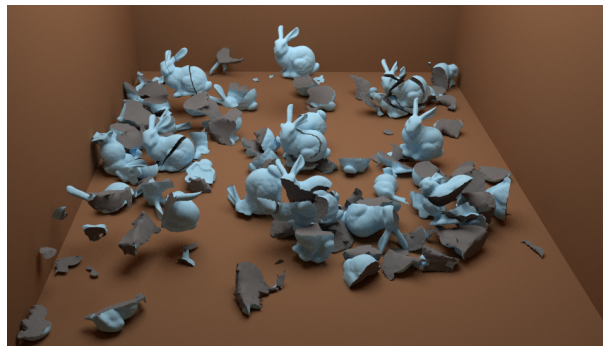


Figure 4. Fast collision detection during real-time brittle fracture simulation.

In complex scenes with many objects, collision detection plays a key role in the simulation performance. This is particularly true in fracture simulation for two main reasons. One is that fracture fragments tend to exhibit very intensive contact, and the other is that collision detection data structures for new fragments need to be computed on the fly. In [17], we present novel collision detection algorithms and data structures for real-time simulation of fracturing rigid bodies. We build on a combination of well-known efficient data structures, namely distance fields and sphere trees, making our algorithm easy to integrate on existing simulation engines. We propose novel methods to construct these data structures, such that they can be efficiently updated upon fracture events and integrated in a simple yet effective self-adapting contact selection algorithm. Altogether, we drastically reduce the cost of both collision detection and collision response. We have evaluated our global solution for collision detection on challenging scenarios, achieving high frame rates suited for hard real-time applications such as video games or haptics. Our solution opens promising perspectives for complex fracture simulations involving many dynamically created rigid objects.

This work was achieved in collaboration with Miguel Otaduy and Sara Schwartzman (URJC Madrid, Spain) and Georges Dumont (MIMETIC team).

**Collision detection: broad phase adaptation from multi-core to multi-GPU architecture** Bruno Arnaldi, Valérie Gouranton

We have presented several contributions on the collision detection optimization centered on hardware performance. We focus on the first step (Broad-phase) and propose three new ways of parallelization of the well-known Sweep and Prune algorithm. We first developed a multi-core model that takes into account the number of available cores. Multi-core architecture enables us to distribute geometric computations with use of multi-threading. Critical writing section and threads idling have been minimized by introducing new data structures for each thread. Programming with directives, like OpenMP, appears to be a good compromise for code portability. We then proposed a new GPU-based algorithm also based on the "Sweep and Prune" that has been adapted to multi-GPU architectures. Our technique is based on a spatial subdivision method used to distribute computations among GPUs. Results show that significant speed-up can be obtained by passing from 1 to 4 GPUs in a large-scale environment [12].

#### **Real-time tracking of deformable target in ultrasound images** Maud Marchal

In several medical applications such as liver or kidney biopsies, an anatomical region needs to be continuously tracked during the intervention. When using ultrasound (US) image modality, tracking soft tissues remains challenging due to the deformations caused by physiological motions or medical instruments, combined with the generally weak quality of the images. In order to overcome the previous limitation, different techniques based on physical model have been proposed in the literature. In [41], we propose an approach for tracking deformable target within 2D US images based on a physical model driven by smooth displacement field obtained from dense information. This allows to take into account highly localized deformation in the US images.

This work was achieved in collaboration with Lucas Royer and Alexandre Krupa (Lagadic team), Anthony Le Bras (CHU Rennes) and Guillaume Dardenne (IRT B-Com).

### **6.3.2. Multimodal feedback**

#### **Stereoscopic Rendering of Virtual Environments with Wide Field-of-Views up to 360** Jérôme Ardouin, Anatole Lécuyer, Maud Marchal

We propose a novel approach [23] for stereoscopic rendering of virtual environments with a wide Field-of-View (FoV) up to 360. Handling such a wide FoV implies the use of non-planar projections and generates specific problems such as for rasterization and clipping of primitives. We propose a novel pre-clip stage specifically adapted to geometric approaches for which problems occur with polygons spanning across the projection discontinuities. Our approach integrates seamlessly with immersive virtual reality systems as it is compatible with stereoscopy, head-tracking, and multi-surface projections. The benchmarking of our approach with different hardware setups shows that it is well compliant with real-time constraints, and capable of displaying a wide range of FoVs. Thus, our geometric approach could be used in various VR applications in which the user needs to extend the FoV and apprehend more visual information.

This work was achieved in collaboration with Eric Marchand (Lagadic team).

#### **A survey on bimanual haptics** Anatole Lécuyer, Maud Marchal, Anthony Talvas.

When interacting with virtual objects through haptic devices, most of the time only one hand is involved. However, the increase of computational power, along with the decrease of device costs, allow more and more the use of dual haptic devices. The field which encompasses all studies of the haptic interaction with either remote or virtual environments using both hands of the same person is referred to as bimanual haptics. It differs from the common unimanual haptic field notably due to specificities of the human bimanual haptic system, e.g. the dominance of the hands, their differences in perception and their interactions at a cognitive level. These specificities call for adapted solutions in terms of hardware and software when applying the use of two hands to computer haptics. In [21], we review the state of the art on bimanual haptics, encompassing the human factors in bimanual haptic interaction, the currently available bimanual haptic devices, the software solutions for two-handed haptic interaction, and the existing interaction techniques.

#### **Haptic cinematography** Fabien Danieau, Anatole Lécuyer

Haptics, the technology which brings tactile or force-feedback to users, has a great potential for enhancing movies and could lead to new immersive experiences. In [14] we introduce *Haptic Cinematography* which presents haptics as a new component of the filmmaker's toolkit. We propose a taxonomy of haptic effects and we introduce novel effects coupled with classical cinematographic motions to enhance video viewing experience. More precisely we propose two models to render haptic effects based on camera motions: the first model makes the audience feel the motion of the camera and the second provides haptic metaphors related to the semantics of the camera effect. Results from a user study suggest that these new effects improve the quality of experience. Filmmakers may use this new way of creating haptic effects to propose new immersive audiovisual experiences.

This work was achieved in collaboration with Marc Christie (MIMETIC team), Julien Fleureau, Philippe Guillotel and Nicolas Mollet (Technicolor).

## 6.4. Collaborative Virtual Environments

### 6.4.1. Collaborative virtual environments for training

**Collaborative Virtual Training with Physical and Communicative Autonomous Agents** Thomas Lopez, Valérie Gouranton, Florian Nouviale, Rozenn Bouville-Berthelot, Bruno Arnaldi

Virtual agents are a real asset in collaborative virtual environment for training (CVET) as they can replace missing team members. Collaboration between such agents and users, however, is generally limited. We presented a whole integrated model of CVET focusing on the abstraction of the real or virtual nature of the actor to define a homogenous collaboration model. First, we defined a new collaborative model of interaction. This model notably allows to abstract the real or virtual nature of a teammate. Moreover, we proposed a new role exchange approach so that actors can swap their roles during training. The model also permits the use of physically based objects and characters animation to increase the realism of the world. Second, we design a new communicative agent model, which aims at improving collaboration with other actors using dialog to coordinate their actions and to share their knowledge. Finally, we evaluated the proposed model to estimate the resulting benefits for the users and we show that this is integrated in existing CVET applications [20].



Figure 5. Collaborative virtual environment for training (CVET) as experienced in our Immersia VR room

**Exchange of avatars : Toward a better perception and understanding** Thomas Lopez, Rozenn Bouville-Berthelot, Florian Nouviale, Valérie Gouranton, Bruno Arnaldi

The exchange of avatars, i.e. the actual fact of changing once avatar with another one, is a promising trend in multi-actor virtual environments. It provides new opportunities for users, such as controlling a different avatar for a specific action, retrieving knowledge belonging to a particular avatar, solving conflicts and deadlocks situations or even helping another user. Virtual Environments for Training are especially affected by this trend as a specific role derived from a scenario is usually assigned to a unique avatar. Despite the increasing use of avatar exchange, users' perception and understanding of this mechanism have not been studied. We propose two complementary user-centered evaluations that aim at comparing several representations for the exchange of avatars; these are termed exchange metaphors. Our first experiment focuses on the perception of an exchange by a user who is not involved in the exchange, and the second experiment analyzes the perception of an exchange triggered by the user. Results show that the use of visual feedback globally aids better understanding of the exchange mechanism in both cases. Our first experiment suggests, however, that visual feedback is less efficient than a simple popup notification in terms of task duration. In addition, the second experiment shows that much simpler metaphors with no visual effect are generally preferred because of their efficiency [19].



Figure 6. Example of metaphor used in "Exchange of Avatars": the blue expert exchanges avatar here with the red expert.

**An interaction abstraction model for seamless avatar exchange in CVET** Rozenn Bouville-Berthelot, Thomas Lopez, Florian Nouviale, Valérie Gouranton, Bruno Arnaldi

Collaboration and interaction between users and virtual humans in virtual environments is a crucial challenge, notably for Collaborative Virtual Environments for Training (CVET). A training procedure, indeed, often involves several actors: trainees, teammates and many times a trainer. Yet, a major benefit of CVET is to propose to users to be trained even if the required number of person needed by the procedure is not available. Therefore, almost every CVET use autonomous virtual humans to replace the missing person. We have proposed to improve the effective collaboration between users and virtual humans involved in a complex



task within CVET. Using an entity called the "Shell", we are able to wrap the features common to both users and virtual humans. It gives us an abstraction level to pool the management of the main processes useful to control an avatar, interact with the environment and gather knowledge from a CVET. Besides, the Shell allows seamless exchange of avatars during a procedure. Thanks to the Shell, the exchange can be carried out at any time during a task while preserving all the data associated to a role in a procedure [29].

**#SEVEN: a Sensor Effector Based Scenarios Model for Driving Collaborative Virtual Environment**  
Guillaume Claude, Valérie Gouranton, Rozenn Bouville-Berthelot, Bruno Arnaldi

We introduced #SEVEN, a sensor effector model that enables the execution of complex scenarios for driving Virtual Reality applications. #SEVEN is based on an enhanced Petri net model which is able to describe and solve intricate event sequences. Our model also proposes several useful features for the design of collaborative scenarios for Collaborative Virtual Environments such as versatile roles and Activity Continuum. We also illustrate its usage by describing a demonstrator that presents an implementation of our model [30].

**Collaborative virtual environments for ergonomics: embedding the design engineer role in the loop**  
Thierry Duval, collaboration with Charles Pontonnier and Georges Dumont (MIMETIC).

We have proposed to define the role and duties of a design engineer involved in a collaborative ergonomic design session supported by a 3D collaborative virtual environment. For example, such a session can be used to adapt the manual task an operator must achieve in the context of an industrial assembly line. We first presented the interest of such collaborative sessions. Then we presented a related work explaining the need of proper 3DCVE and metaphors to obtain efficient collaborative ergonomic design sessions. Then we proposed a use case highlighting the type of metaphor such engineers need to have to be efficient in such a framework [40].

#### 6.4.2. Collaborative virtual environments and awareness

**Improving Awareness for 3D Virtual Collaboration by Embedding the Features of Users' Physical Environments and by Augmenting Interaction Tools with Cognitive Feedback Cues** Thierry Duval, Thi Thuong Huyen Nguyen, Valérie Gouranton, collaboration with MimeTic

The feeling of presence is essential for efficient interaction within Virtual Environments (VEs). When a user is fully immersed within a VE through a large immersive display system, his/her feeling of presence can be altered because of disturbing interactions with his/her physical environment, such as collision with hardware parts of the system or loss of tracking. This alteration can be avoided by taking into account the physical features of the user as well as those of the system hardware and embedding them in the VE. Moreover, the 3D abstract representation of these physical features can also be useful for collaboration between distant users because they can make a user aware of the physical limitations of the others he/she is collaborating with. We used the Immersive Interactive Virtual Cabin (IIVC) model to obtain this virtual representation of the user's physical environment and we illustrated how this representation can be used in a collaborative navigation task in a VE. We also presented how we can add 3D representations of 2D interaction tools in order to cope with asymmetrical collaborative configurations, providing 3D cues for a user to understand the actions of the others even if he/she is not fully immersed in the shared VE [15].

**From 3D Bimanual Toward Distant Collaborative Interaction Techniques: An Awareness Issue** Morgan Le Chénechal, Thierry Duval, Valérie Gouranton, Bruno Arnaldi, collaboration with b<>com

CVE involve the use of complex interaction techniques based on specific collaborative metaphors. The design of these metaphors may be a difficult task because it has to deal with collaborative issues that came from sparse research areas (Human-Computer Interfaces, Human-Human Interactions, Networking, Physiology and Social Psychology). Metaphors for bimanual interactions have been developed for a while essentially because it is a widely spread area of interest for common tasks. Bimanual interactions involve the simultaneous use of both hands of the user in order to achieve a goal with better performances compared to uni-manual interactions thanks to a natural skill that is proprioception. This collaborative aspect could certainly be a helpful entry point in the design of efficient collaborative interaction techniques extended from improved bimanual metaphors. However, the proprioceptive sense cannot be considered in the same way, and additional features must be

proposed to be able to collaborate efficiently. Thus, awareness is a key to let CVE be usable and the availability of collaborative feedbacks is essential to extend bimanual interactions toward collaborative ones. In this paper, we based our study on existing work on bimanual and collaborative interaction techniques trying to draw similarities between them. We emphasize common points between both fields that could be useful to better design both metaphors and awareness in CVE [34].

**A survey of communication and awareness in collaborative virtual environments** Thi Thuong Huyen Nguyen, Thierry Duval

In the domain of Collaborative Virtual Environments (CVEs), many virtual worlds, frameworks and techniques are built based on a specific and direct purpose. There is not a general and still good and efficient enough solution for all the collaborative systems. Depending on the purpose of the collaborative work, the techniques of interaction and of manipulation change from one application to another. Despite this difference between interaction techniques, they always benefit greatly from awareness features that help in explicating implicit knowledge related to one's own and others' working activities as well as to virtual workspace. In addition, people in CVEs also use communication channels to negotiate shared understandings of task goals, of task decomposition and of task progress. Therefore, awareness and communication are usually considered as "instruments" to complete collaborative tasks in the environment. However, few research work have been devoted to improving the awareness and the communication channels in CVEs for a better collaboration between users. We have studied the importance of awareness and communication in collaborative virtual environments. We have investigated different kinds of awareness which need to be carefully designed. We have discussed different communication means and how to cope with this diversity, so we can benefit from the availability of different peripheral devices and can find an effective communication means to work together. Finally, we have made some propositions to overcome these actual limitations of CVEs [37].

### 6.4.3. Collaborative virtual environments and software engineering

**When model driven engineering meets virtual reality: feedback from application to the Collaviz framework** Thierry Duval, collaboration with Arnaud Blouin and Jean-Marc Jézéquel (DIVERSE).

Despite the increasing use of 3D Collaborative Virtual Environments (3D CVE), their development is still a cumbersome task. The various concerns to consider (distributed system, 3D graphics, etc.) complexify the development as well as the evolution of CVEs. Software engineering recently proposed methods and tools to ease the development process of complex software systems. Among them, Model-Driven Engineering (MDE) considers models as first-class entities. A model is an abstraction of a specific aspect of the system under study for a specific purpose. MDE thus breaks down a complex system into as many models for different purposes, such as: generating code from models; building domain specific programming/modeling languages (DSL); generating tools such as graphical or textual editors. We have leveraged MDE for developing 3D CVEs. We showed how the Collaviz framework took benefits from a DSL we built. The benefits are multiple: 3D CVE designers can focus on the behavior of their virtual objects without bothering with distributed and graphics features; configuring the content of 3D CVEs and their deployment on various software and hardware platforms can be automated through code generation. We detailed the development process we propose and the experiments we conducted on Collaviz [31].

## 6.5. Brain-Computer Interfaces

### 6.5.1. Novel usages of BCI

**Mind-Mirror: combining BCI and augmented reality to "see your brain in action in your own head",** Anatole Lécuyer, Jonathan Mercier, Maud Marchal

Imagine you are facing a mirror, seeing at the same time both your real body and a virtual display of your brain in activity and perfectly superimposed to your real image "inside your real skull". We have introduced a novel augmented reality paradigm called "Mind-Mirror" which enables the experience of seeing "through your own head", visualizing your brain "in action and in situ" [36]. Our approach relies on the use of a semi-transparent mirror positioned in front of a computer screen. A virtual brain is displayed on screen





*Figure 7. The Mind Mirror system : Visualization of brain activity in real-time using a mirror-based augmented reality setup and an EEG headset.*

and automatically follows the head movements using an optical face-tracking system. The brain activity is extracted and processed in real-time with the help of an electroencephalography cap (EEG) worn by the user. A rear view is also proposed thanks to an additional webcam recording the rear of the user's head. The use of EEG classification techniques enables to test a Neurofeedback scenario in which the user can train and progressively learn how to control different mental states, such as "concentrated" versus "relaxed". The results of a user study comparing a standard visualization used in Neurofeedback to our approach showed that the Mind-Mirror could be successfully used and that the participants have particularly appreciated its innovation and originality. We believe that, in addition to applications in Neurofeedback and Brain-Computer Interfaces, the Mind-Mirror could also be used as a novel visualization tool for education, training or entertainment applications.

This work was achieved in collaboration with Fabien Lotte from POTIOC team (Inria-Bordeaux).

#### **Using SSVEP-based BCI with 3D stereoscopic display** Anatole Lécuyer

We have investigated the feasibility of dual-frequency Steady-State Visual Evoked Potential (SSVEP) stimulation using a 3-D display and stereoscopic glasses [44]. Dual-frequency stimulation allows for more targets to be created using a small number of frequencies, and stereoscopic vision offers a suitable medium for dual-frequency stimulation as the two views can be controlled independently. Participants were exposed to a repetitive visual stimulus flashing at different frequencies in the left and right views and the electroencephalography (EEG) trace was examined. Our results suggest that the two stimulation frequencies can still be evident in the SSVEP response. In addition, the participant ratings showed no significant differences in fatigue, annoyance, comfort or strangeness of the stimulation compared to conventional forms of stimulation. These results pave the way for further studies using stereoscopic dual-frequency stimulation and its potential for use in virtual reality and 3D videogames

This work was achieved in collaboration with Robert Leeb (EPFL, Switzerland).

#### **Passive BCI and music** Anatole Lécuyer

Passive brain-computer interfaces (passive BCI), also named implicit BCI, provide information from user mental activity to a computerized application without the need for the user to control his brain activity. We have proposed an overview of current research on passive BCIs in [45]. We have notably studied how they can be applied to the context of music creation, where they can provide novel information to adapt the music creation process, e.g., exploiting user mental concentration to adapt the music tempo.

### 6.5.2. BCI methodology

#### **Which factors drive successful BCI skill learning?** Anatole Lécuyer, Lorraine Perronnet

Brain-Computer Interfaces although very promising, suffer from a poor reliability. Rather than improving brain signal-processing alone, an interesting research direction is to guide users to learn BCI control mastery. Thus, we have introduced a set of motivational and cognitive factors which could influence the learning process, and which should be considered to improve the global performance of BCI users [47]. We base our approach on Keller's integrative theory of motivation, volition, and performance, which combines motivational (affective) and cognitive factors, to explain what makes human users learn and perform efficiently, irrespectively of the task. These factors can guide the creation of learning environments, such as BCI training protocols.

This work was achieved in collaboration with Fabien Lotte and Christian Muhl (POTIOC team, Inria-Bordeaux), Moritz Grosse-Wentrup (MPI, Tuebingen), and Reinhold Scherer (TU Graz, Austria).

#### **A methodological framework for applications combining BCI and VE** Anatole Lécuyer

We have proposed a user-centred methodological framework [46] to guide design and evaluation of applications combining Brain-Computer Interface (BCI) and Virtual Environment (VE). Our framework is based on the contributions of ergonomics to ensure these applications are well suited for end-users. It provides methods, criteria and metrics to perform the phases of the human-centred 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 applications combining BCI and VE were also discussed.

This work was achieved in collaboration with Fabien Lotte from POTIOC team (Inria-Bordeaux).

## IMAGINE Project-Team

# 6. New Results

## 6.1. Highlights of the Year

- Vector Graphics Complexes, an new structure for 2D illustration developed in collaboration with UBC, resulted into a publication at ACM SIGGRAPH [4]. This superset of multi-layers graphics and of planar maps, enable intuitive design and deformation of 2D illustrations thanks to the separation of geometry from topology.
- Our work on elastic implicit skinning, a collaboration with U. Toulouse, Victoria University, and Inria Bordeaux was accepted at ACM SIGGRAPH Asia [16]. Thanks to robust iso-surface tracking, this method captures dynamic skin siding effects and can be used with extreme bending angles.

## 6.2. User-centered Models for Shapes

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

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.

### 6.2.1. Implicit modeling

**Participants:** Antoine Bégault, Marie-Paule Cani, Michael Gleicher, Cédric Zanni.



Figure 4. Illustration from [17] showing some results of our N-ary implicit blends.

Our insight towards 3D shapes that respond in an intuitive way during both design and animation is to develop representations that clearly separate changes of structure - namely, the morphology of the shape - from changes of posture (its current 3D isometric embedding). Using skeletons is an excellent way to do so for 3D solids: the structure of a shape is represented by the topology of the skeleton, the length of its components and the shape thickness around it, while the shape posture is defined by the embedding of the skeleton in 3D space. Implicit surfaces (iso-surfaces of scalar fields) are the best mathematical model so far for generating 3D shapes from skeletons. However, a number of long standing problems - blending at distance that makes topology unpredictable, bulges at junctions, blurring of details - reduced the interest for this representation for many years. We addressed several of these issues in the last few years. Our most recent contribution is a method for enabling topology control in the case n-ary implicit blends [17]. Shapes are modeled using scale-invariant integral primitives (SCALIS) along skeletons, and blend with a plus. We use field warping to avoid unwanted blending and provide a unique control (based for instance on the angle) on the way skeleton-based primitives are allowed to blend. See Figure 4

### 6.2.2. Towards responsive assemblies

**Participants:** Stefanie Hahmann, Jean-Claude Léon, Aarohi Singh Johal.

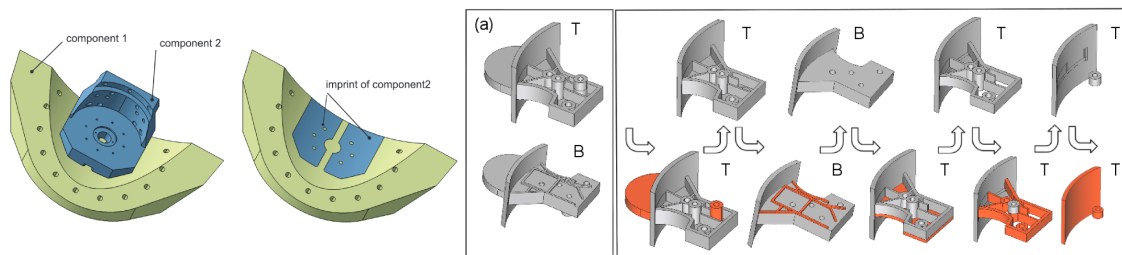


Figure 5. Illustration from [8] (Left) showing the automatic extraction of geometric interfaces, and from [2] (Right) showing the automatic generation of a construction graph.

We chose to focus on man-made objects to tackle the topic of shape assemblies, 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 Airbus group and EDF.

Assemblies representing products are most often reduced to a collection of independent CAD models representing each component. The designation of each component and information about its function are often missing. As a result, geometric interfaces between components are unknown. These interfaces are particularly useful for structural mechanics to be able to quickly generate a Finite Element model of the assembly. This is especially critical when the latter gets very complex. [8] addresses the problem of automatically generating a class of geometric interfaces for very complex assemblies. GPU-based algorithms have proved suitable to obtain reliable results on CAD models.

Precisely determining interfaces between components is also a first requirement to enrich geometric models with functional information, since a subset of functions derives from interfaces between components. Based on both geometric interfaces and on a new concept of conventional interfaces, we proposed a series of approaches [13], [3] that make use of qualitative and ontology-based reasoning to connect CAD components and their geometric interfaces to functions or to functional designations of components: this results into an intrinsic identifier of a component in an assembly that connects it to its function.

To efficiently process assemblies of components, shape analysis [40] is particularly useful to generate the dimensionally reduced models needed for structural mechanics. [2] shows that analyzing a B-Rep CAD model to derive a construction graph, i.e. a set of construction trees, can be a robust basis to generate dimensionally reduced models.

Lastly, we extended shape analysis methods to detect some sets of symmetries [9]. Recovering this knowledge and embedding it into a model is the first step towards functionality-preserving deformations of complex man-made prototypes.

### 6.2.3. Parametric shapes

**Participants:** Stefanie Hahmann, Léo Allemand-Giorgis, Tibor Stanko.



Figure 6. Illustration from [38](Left) showing our results on monotonic interpolation, and from [1](Right) with our  $G^1$  interpolation surfaces for quad meshes.

We are developing new smooth parametric surface models defined on irregular quad meshes. They are in fact a powerful alternative to subdivision surface and singularly parameterized tensor product surfaces since they combine the advantages of both, the arbitrary topology of quad meshes and the smoothness of the tensor product patches. In collaboration with G.-P. Bonneau (Maverick team) several parametric triangular surface models for arbitrary topologies have been developed in the past. A new surface spline model has been published [1] and presented at GMP'14. It solves the problem of defining a  $G^1$ -continuous surface interpolating the vertices of an arbitrary quad mesh with low degree polynomial tensor product patches. It further aims to produce shapes of very high visual quality while reducing the number of control points, see Figure 6(right).

Another contribution concerns the modeling and smoothing of shapes using the Morse-Smale complex. The Morse-Smale complex is a topological structure defined on scalar functions which extracts critical points of the function and the links between them. By encoding a hierarchy between critical points, less important critical points can be deleted in order to simplify the structure. Our goal is to reconstruct a new shape, which corresponds to the simplified structure while approximating the initial data and preserving the most salient features. We first developed a method for interpolating monotone increasing 2D scalar data with a monotone piecewise cubic  $C^1$ -continuous surface. Monotonicity is a sufficient condition for a function to be free of critical points inside its domain. We overcome the restrictive standard axial monotonicity for tensor-product surfaces and introduce sufficient conditions and two algorithms for a more relaxed monotonicity constraint [38], see a piecewise monotonic shape in Figure 6(left). Then, some preliminary results on shape reconstruction from Morse-Smale complexes have been presented as a Posterand at a national conference [35].

In collaboration with Hans Hagen and Anne Berres from University of Kaiserslautern, we investigated conditions under which shape deformations preserve surface curvatures. The work has been published as a chapter in a scientific book [39].

### 6.3. Models for Motion 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.

#### 6.3.1. Real-time physically-based models

**Participants:** Armelle Bauer, Ali Hamadi Dicko, François Faure, Matthieu Nesme.



Figure 7. Illustration from [14](Left) showing our velocity based adaptive simulation, and from [21](Right) on interactive visualization of muscle activity.

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 further improved this year the efficiency of this approach: we developed an adaptive version of frame-based elastic models, where frames get seamlessly attached to other ones during deformations when appropriate, in order to reduce computations [14], [33].

Frame-based models were successfully used to model limb movements in anatomical modeling [21]. The efficiency of this method enables us to advance towards the concept of a *Living book of anatomy*, where users move their own body and observe it through a tablet to get some visual illustration of anatomy in motion (see Figure 7).

#### 6.3.2. Specific models for virtual creatures

**Participants:** Marie-Paule Cani, Michael Gleicher.

In collaboration with Loic Barthe and Rodolphe Vaillant from IRIT (U. Toulouse), Brian Wyvill (U. Victoria) and Gael Guennebaud (Manao, Inria), we developed a new automatic method for character skinning: Based on the approximation of character limbs with Hermite RBF implicit volumes, we adjust the mesh vertices representing the skin by projecting them back, at each animation step, to their iso-surface of interest. Since the vertices start from their previous position at the last animation step, there is no need of specifying skinning weights and using another skinning method as pre-computation, as in our previous implicit skinning method. Our solution avoids the well known blending artifacts of linear blend skinning and of dual quaternions, accommodates extreme blending angles and captures elastic effect in skin deformation [16].



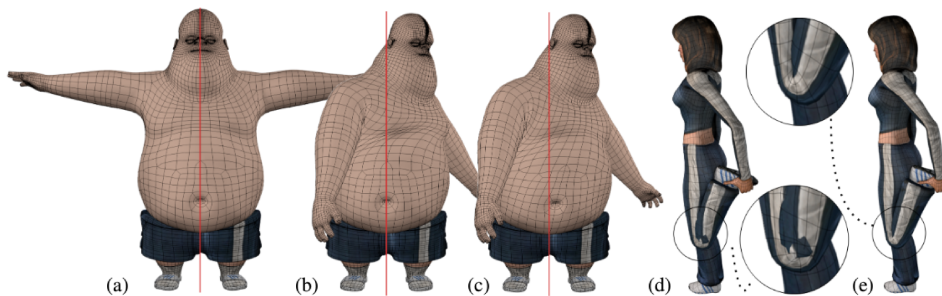


Figure 8. Illustration from [16] showing our results on dynamic skin deformations.

This year, we also studied the way character eyes and gazes are to be animated. This extensive study resulted into a state of the art report published at the Eurographics conference [32]

## 6.4. 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.

### 6.4.1. Knowledge representation through 3D ontologies

**Participants:** Armelle Bauer, Jean-Claude Léon, Olivier Palombi.

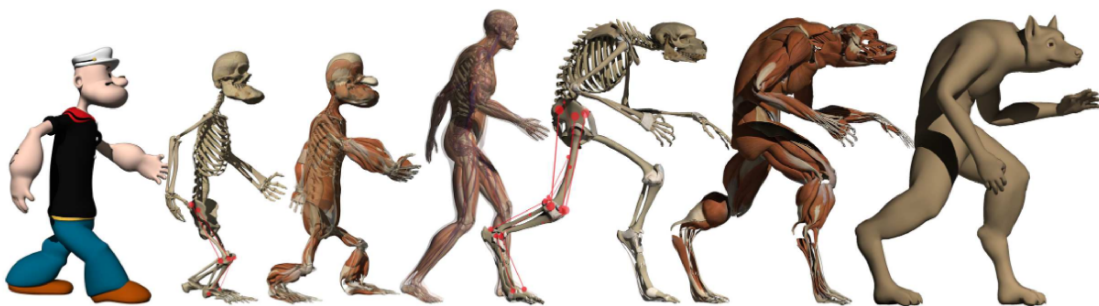


Figure 9. Illustration from [22] showing our results on ontology based anatomy generation.

We chose to develop 3D ontologies for being able to express combined knowledge on geometry, motion and function for assemblies or hierarchies of 3D objects. This is done in collaboration with a specialized group from the LIG laboratory in Grenoble. We decided to first focus these ontologies developments on two topics on which group members have a strong expertise: the anatomical domain (an interesting application test-bed for educational scenarios) and the industrial prototyping domain (where assembly scenarios can be defined).

We developed an anatomical knowledge database called My Corporis Fabrica (MyCF). We first linked functional entities defined in MyCF to the involved anatomical structures, using the musculoskeletal system as a test-bed. Based on this new formal description of the functional anatomy of limbs, we presented a novel pipeline for the construction of biomechanical simulations by combining generic anatomical knowledge with specific data which can handle complex reasoning and querying in MyCF. This resulted into a publication in the Journal of Biomedical Semantics [11]. We also used MyCF within our previous framework of anatomical transfert to set up an assistant tool for modeling and simulating anatomical structure such as bones, muscles, viscera and fat tissues easily while ensuring a correct anatomical consistency [22].

Secondly, in analysing the similarities and differences between existing ontology based description of products and virtual humans, we developed a common framework for combining 3D models and functional description to both models [15], [34].

#### 6.4.2. Virtual direction tools

**Participants:** Adela Barbulescu, Rémi Ronfard.

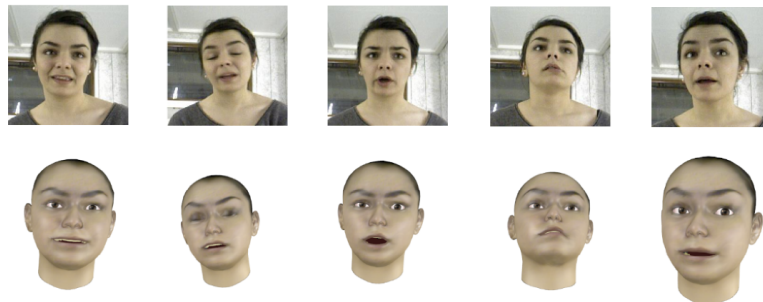


Figure 10. Illustration from [20] showing our results on expressive virtual actors.

We are developing a new approach to transfer speech signals and 3D facial expressions to virtual actors of a different identity. The converted sequences should be perceived as belonging to the target actors. This is the goal of Adela Barbulescu's thesis, co-advised by Gérard Bailly from GIPSA-lab. Our work started with conversion of speaking styles through speech signals only. This year, we started extending this approach to visual prosody and advanced on communicating social attitudes through head gestures [20].

#### 6.4.3. Virtual cinematography

**Participants:** Quentin Galvane, Vineet Ghandi, Christophe Lino, Rémi Ronfard.

Our goal is to model automatic cameras for covering 3D scenes, as well as to develop semi-automatic film editing techniques to help conveying narration. This work was first conducted on video data, enabling us to test our ideas without the need for complex 3D movies: we designed an automatic method for the identification of actors in a video, and are using it for the automatic re-framing and editing high-resolution videos shots of theater rehearsals [25].



Figure 11. Illustration from [23] showing our results on narrative-driven camera control of computer games.

In parallel, we started extending this methodology to 3D animation, in collaboration with the Mimetic group in Rennes and with Geneva University: this year, we proposed a new method for replaying first person video games with automatic camera control based on the narration [23]. We also advanced towards semi-automatic film editing: A paper was just accepted to AAI 2015. To stress the difficulty of validating film editing methods, we devoted a specific work to validation methodologies [27].

We also addressed other issues related to cinematography and narratives: We designed a pre-visualization system for 3D cinematography to be used in the Action3DS project [30]: the method makes use of 3D modeling to show what the spectators watching a 3D movie are going to see, in order to ease 3D camera control by the film director. Lastly, we worked on computer generation of narrative discourses with the university of Geneva [31].

This year, Remi Ronfard and Vineet Gandhi wrote a patent application "Dispositif de génération de rushes cinématographiques par traitement vidéo", demande de brevet français no. 1460957, déposée le 13 novembre 2014.

## 6.5. 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.

### 6.5.1. Sketch-based modeling and editing of 3D shapes

**Participants:** Marie-Paule Cani, Arnaud Emilien, Even Entem, Stefanie Hahmann, Rémi Ronfard.

While a lot of work has been done on sketch-based modeling of solid shapes, only a few methods do tackle surface models. Terrain surfaces are particular challenging: their fractal-like distribution of details makes them easy to identify, but these cannot be fully drawn by a user. In our work, users only need to draw the main silhouettes they would like to see from a first person viewpoint (enabling, for instance, an art director to set the background scene behind his actors). We generate a plausible, complex terrain that matches the sketch by deforming an existing terrain model. This is done by analyzing the complex silhouettes with cups and T-junctions in the input sketch and matching them with perceptually close features of the input terrain. The rest

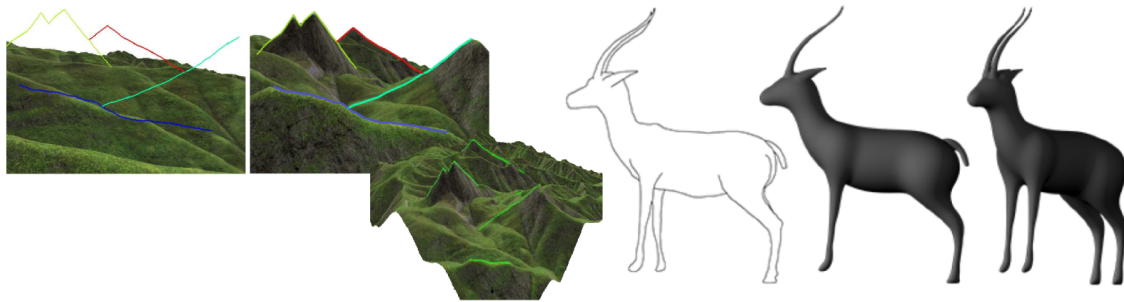


Figure 12. Illustration from [29](Left) showing our results on terrain editing, and from [6](Right) on 3D animal modeling from sketch.

of the terrain is seamlessly deformed while keeping its visual complexity and style. This work was presented at Graphics Interface 2014 [29] and extended to enable the combination of silhouettes from multiple viewpoints in the Computer and Graphics journal [12].

In collaboration with UBC, we introduced the vector graphics complex (VGC), a simple data structure that supports non-manifold topological modeling for 2D vector graphics illustrations. The representation faithfully captures the intended semantics of a wide variety of illustrations, and is a proper superset of scalable vector graphics and planar map representations. VGC nearly separates the geometry of vector graphics objects from their topology, making it easy to deform objects in interesting and intuitive ways, a premise for enabling their animation. This work was published at SIGGRAPH 2014 [4]. We also developed a method for generating 3D animals from a single sketch. This method takes a complex sketch with cups and T-junctions as input (see Figure 12), and makes use of symmetry hypotheses to analyze it into regions corresponding to the main body and to front and back limbs. The different regions are then automatically reconstructed and blended together using our implicit modeling methodologies (SCALIS surfaces).

Lastly, we designed a sketch-based interface for authoring illustrative animations. The method makes use of hierarchical motion brushes, a new concept for specifying complex hierarchical motion with a few strokes [28].

### 6.5.2. Sketching and Sculpting Motion

**Participants:** Marie-Paule Cani, Arnaud Emilien, Kevin Jordao.

We extended sculpting methods, which had been restricted so far to homogeneous geometric models of a single dimension, to the handling of complex structured shapes and to the interactive sculpting of animated environments.

We developed the first method enabling to sculpt animated content in extending our previous elastic mutable model approach. Relying on the crowd patches representation for modeling animated crowds, we extended component mutations to space-time content, enabling a user to stretch, bend or assemble populated streets while ensuring that individual character trajectories remain continuous through space and time, as well as plausible. This work, developed within Kevin Jordao's thesis co-advised with Julien Pettré from the Mimetic project-team in Rennes, was published at Eurographics 2014 [7].

We developed an interactive system for designing complex waterfall scenes: vector elements created by the user (contacts, freefalls and pools) are used to control the procedural creation of complex waterfalls and rivers that match the user intend while ensuring coherent flows and good embedding within the terrain [18].

### 6.5.3. Interaction devices and gestural patterns

**Participants:** Marie-Paule Cani, Rémi Brouet.

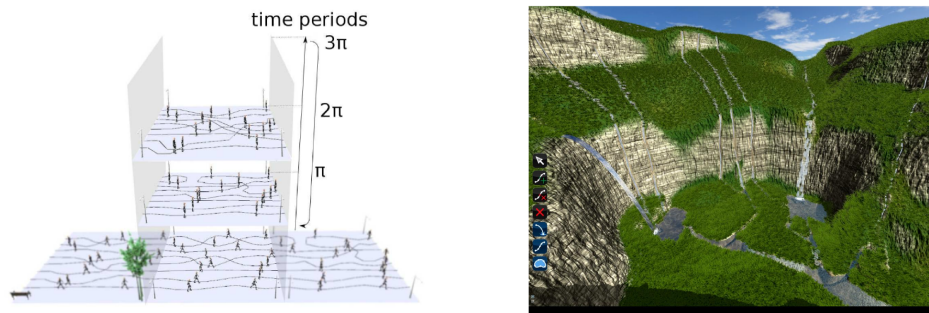


Figure 13. Illustration from [7](Left) on our crowd sculpting interface, and from [18](Right) on interactive design of complex waterfall scenes.

Our work on gestural interaction patterns for 3D design has been developed in collaboration with the HCI team from LIG laboratory towards the exploration of 2D multi-touch tables for the placement and deformation of 3D models.

We are also exploring the use of multi-touch tables for the interactive design and editing of 3D scenes. This is the topic of Rémi Brouet's PhD thesis, co-advised with Renaud Blanch from the human-computer interaction group of LIG laboratory. The main challenge here is to find out how to use 2D interaction media for editing 3D content, hence how to intuitively control the third dimension (depth, non-planar rotations, 3D deformations, etc). Our work on this topic started with a preliminary user study enabling us to analyze all possible hand interaction gestures on table-tops, and to explore the ways users would intuitively try to manipulate 3D environments, either for changing the camera position or for moving objects around. We extracted a general interaction pattern from this study. Our implementation enables both seamless navigation and docking in 3D scenes, without the need for any menu or button to change mode. We are currently extending this work to object editing scenarios, where shapes can be bent or twisted in 3D using 2D interaction and automatic mode selection only.



## IN-SITU Project-Team

# 6. New Results

## 6.1. Highlights of the Year

Wendy Mackay received the *ACM SIGCHI Lifetime Service Award*.

BEST PAPERS AWARDS :

[22] CHI '14. C. LIU, O. CHAPUIS, M. BEAUDOUIN-LAFON, É. LECOLINET, W. E. MACKAY.

## 6.2. Interaction Techniques

**Participants:** Caroline Appert, Michel Beaudouin-Lafon, Anastasia Bezerianos, David Bonnet, Olivier Chapuis [correspondant], Cédric Fleury, Stéphane Huot, Can Liu, Justin Mathew, Wendy Mackay, Halla Olafsdottir, Theophanis Tsandilas, Oleksandr Zinenko.

InSitu explores interaction and visualization techniques in a variety of contexts, including individual interaction techniques on different display surfaces that range from mobile devices to very large wall-sized displays, including standard desktop systems and tabletops.

This year, we investigated multi-touch gestures on tabletop [26], we considered the combination of Tilt and Touch on smartphone [29], we proposed novel bi-manual interaction techniques for tablets [18], we introduced a novel focus+context technique to facilitate route following [14], we introduced the *GlideCursor* to facilitate pointing on large display [15], we compared physical navigation in front of a wall-size display with virtual navigation on the desktop [22], we studied users' behavior in immersive Virtual Environments [12], we built a tool to ease the extraction and the expression of parallelism in programs [30] and we investigated the effect of contours on star glyphs [13].

In addition to providing knowledge for designers and practitioners, this set of remarkable results advances our overall knowledge regarding basic interactive phenomena, and allows to better understand how user practices will change.

*Multitouch on Tabletop* – We systematically studied how users adapt their grasp when asked to translate and rotate virtual objects on a multitouch tabletop [26]. We have shown that users choose a grip orientation that is influenced by three factors: (1) a preferred orientation defined by the start object position, (2) a preferred orientation defined by the target object position, and (3) the anticipated object rotation. We have examined these results in the light of the most recent models of planning for manipulating physical objects and explored how these results can inform the design of tabletop applications.

*Tilt & Touch* – We studied the combination of tilt and touch when interacting with mobile devices [29]. We conducted an experiment to explore the effectiveness of *TilTouch* gestures for both one-handed and two-handed use. Our results indicate the best combinations of *TilTouch* gestures in terms of performance, motor coordination, and user preferences.

*SPad* – We created *SPad* [18], a new bimanual interaction technique designed to improve productivity on multi-touch tablets: the user activates quasimodes with the thumb of the non-dominant hand while holding the device with that hand and interacts with the content with the dominant hand (figure 3 ). We conducted an iterative design process and created a tablet application that demonstrates how *SPad* enables faster, more direct and more powerful interaction without increasing complexity.



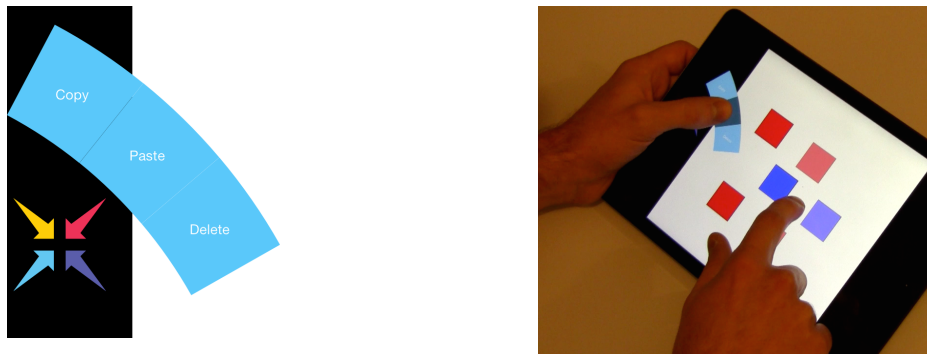


Figure 3. Left: SPad control accessible with the thumb. Three commands are accessible with a tap of the thumb, 4 menus are accessible with swipes. Right: SPad in use to paste and move objects that have just been copied.

*RouteLenses* – Millions of people go to the Web to search for geographical itineraries. Inspecting those map itineraries remains tedious because they seldom fit on screen, requiring much panning & zooming to see details. Focus+context techniques address this problem by displaying routes at a scale that allows them to fully fit on screen: users see the entire route at once, and perform magnified steering using a lens to navigate along the path, revealing additional detail. We created *RouteLenses* [14], a type of lenses that automatically adjusts their position based on the geometry of the path that users steer through (figure 4). *RouteLenses* make it easier for users to follow a route, yet do not constrain movements too strictly, leaving them free to move the lens away from the path to explore its surroundings.

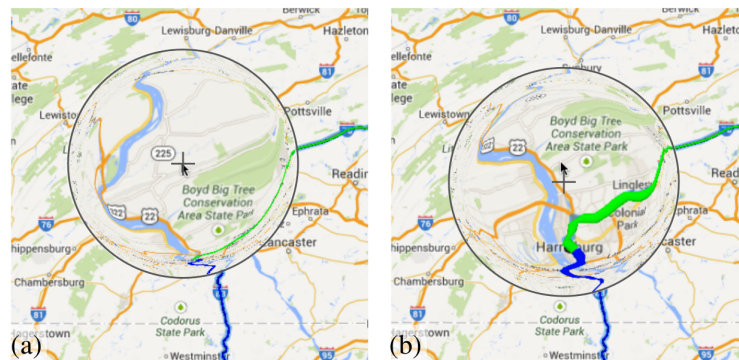


Figure 4. Following an itinerary. (a) Conventional lens: the user overshoots at a right turn in Harrisburg; losing the route that falls in the distorted region. (b) *RouteLens*: the route's attraction compensates the overshoot; the lens remains closer to the route, which remains in focus.

*GlideCursor* – Pointing on large displays with an indirect, relative pointing device such as a touchpad often requires clutching. We designed and evaluated *GlideCursor* [15], which lets the cursor continue to move during clutching gestures. The effect is that of controlling the cursor as a detached object that can be pushed, with inertia and friction similar to a puck being pushed on a table. We analyzed gliding from a practical and a theoretical perspective and conducted two studies. The first controlled experiment established that gliding

reduces clutching and can improve pointing performance for large distances. We introduced a measure called *cursor efficiency* to capture the effects of gliding on clutching. The second experiment demonstrated that participants use gliding even when an efficient acceleration function lets them perform the task without it, without degrading performance.

*Wall vs. Desktop* – The advent of ultra-high resolution wall-size displays and their use for complex tasks require a more systematic analysis and deeper understanding of their advantages and drawbacks compared with desktop monitors. While previous work has mostly addressed search, visualization and sense-making tasks, we have designed and evaluated an abstract classification task that involves explicit data manipulation [22]. Based on our observations of real uses of a wall display (figure 5 -left), this task represents a large category of applications. We conducted a controlled experiment that uses this task to compare physical navigation in front of a wall-size display (figure 5 -right) with virtual navigation using pan-and-zoom on the desktop. Our main finding is a robust interaction effect between display type and task difficulty: while the desktop can be faster than the wall for simple tasks, the wall gains a sizable advantage as the task becomes more difficult.



Figure 5. Left: Fine tuning the CHI 2013 conference schedule on the WILD display. Center: Classification task inspired by the CHI scheduling task conducted on the wall-sized display. Subjects must move misclassified red disks into containers (represented by the individual screens) of the same class. The class is represented by a very small letter at the center of each disk, forcing subjects to physically move in front of the display.

*Immersive VE* – The feeling of presence is essential for efficient interaction within Virtual Environments (VEs). When a user is fully immersed within a VE through a large immersive display system, her feeling of presence can be altered because of disturbing interactions with her physical environment. This alteration can be avoided by taking into account the physical features of the user as well as those of the system hardware. Moreover, the 3D abstract representation of these physical features can also be useful for collaboration between distant users. In [12] we presented how we use the Immersive Interactive Virtual Cabin (IIVC) model to obtain this virtual representation of the user’s physical environment and we illustrated how this representation can be used in a collaborative navigation task in a VE. We also presented how we can add 3D representations of 2D interaction tools in order to cope with asymmetrical collaborative configurations, providing 3D cues for a user to understand the actions of others even if he/she is not fully immersed in the shared VE.

*Clint* – We created *Clint*, a direct manipulation tool to ease the extraction and the expression of parallelism in existing programs [30]. Clint is built on top of state-of-the-art compilation tools (polyhedral representation of programs) in order to give a visual representation of the code, perform automatic data dependence analysis and to ensure the correctness of code transformations (figure 6 ). It can be used to rework and improve automatically generated optimizations and to make manual program transformation faster, safer and more efficient.

*Start Glyphs* – We conducted three studies using crowd-sourcing on Amazon mechanical Turk, to determine the effect of using contours on data glyphs such as star glyphs [13]. Our results indicate that glyphs without

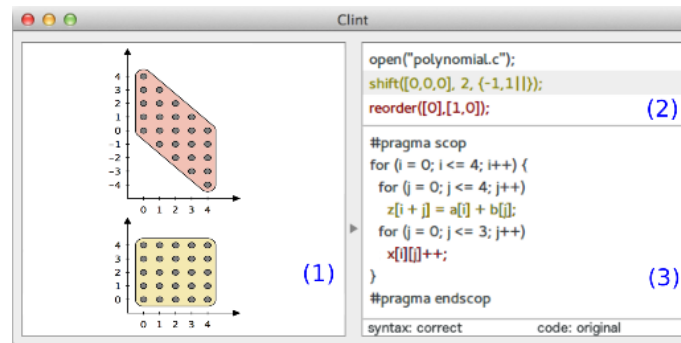


Figure 6. Clint interface includes: (1) interactive visualization, (2) editable history view, and (3) source code editor.

contours lead viewers to naturally make judgements that are data-driven. Whereas adding contours encourages shape similarity, e.g. perceiving rotated variations of glyphs as similar (even though they are not similar in data space).

### 6.3. Research Methods

**Participants:** Michel Beaudouin-Lafon, Anastasia Bezerianos, Jérémie Garcia, Stéphane Huot, Ilaria Liccardi, Wendy Mackay [correspondant], Justin Mathew.

Conducting empirical research is a fundamental part of InSitu's research activities, including observation of users in field and laboratory settings to discover problems faced by users, controlled laboratory experiments to evaluate the effectiveness of the technologies we develop, longitudinal field studies to determine how our technologies work in the real world, and participatory design, to explore design possibilities with users throughout the design process.

*Computer-aided Composition* – We designed *Polyphony* [20], a novel interface for systematically studying all phases of computer-aided composition, and then used it to observe expert creative behavior. *Polyphony* is a unified user interface that integrates interactive paper and electronic user interfaces for composing music. We asked 12 composers to use it (figure 7 -left) to compose an electronic accompaniment to a 20-second instrumental composition by Anton Webern. The resulting dozen comparable snapshots of the composition process reveal how composers both adapt and appropriate tools in their own way. In collaboration with IRCAM, we also conducted a longitudinal study where we closely collaborated with composer Philippe Leroux [19] in the creation of his piece *Quid sit musicus*. The composer used our interfaces based on interactive paper along with an OpenMusic library to generate compositional material for this work (figure 7 -right).

*Multitouch Gestures* – We created a design space of simple multitouch gestures that designers of user interfaces can systematically explore to propose more gestures to users [27]. We further considered a set of 32 gestures for tablet-sized devices, by developing an incremental recognition engine that works with current hardware technology, and empirically testing the usability of those gestures. In our experiment, individual gestures were recognized with an average accuracy of  $\sim 90\%$ , and users successfully achieved some of the transitions between gestures without the use of explicit delimiters. The goal of this work is to assist designers in optimizing the use of the rich multi-touch input channel for the activation of discrete and continuous controls, and to enable fluid transitions between controls, e.g. when selecting text over multiple views, manipulating different degrees of freedom of a graphical object or invoking a command and setting its parameter values in a row.

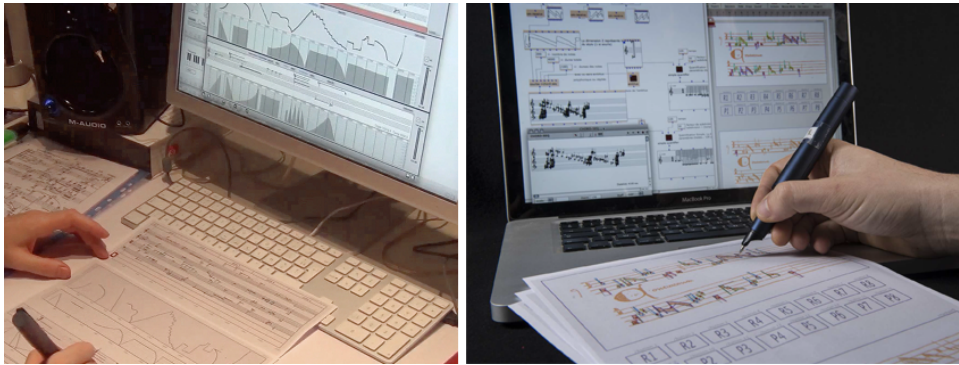


Figure 7. Left: A composer completes the composition task using Polyphony. Right: Set of tools created for Philippe Leroux' piece *Quid sit musicus*. Photo by H. Raguet, Inria.

*Spatial Audio* – We investigated the issues of spatialization techniques for object-based audio production and introduced the Spatial Audio Design Spaces framework (SpADS) [25], which describes the spatial manipulation of object-based audio. These design spaces are based on interviews with professional sound engineers and on a morphological analysis of 3D audio objects that clarifies the relationships between recording and rendering techniques that define for 3D speaker configurations. This will allow us to analyze and design novel advanced object-based controllers.

*Physical Visualizations* – We studied the design process of physical visualizations. An increasing variety of such visualizations are being built, for purposes ranging from art and entertainment to business analytics and scientific research. However, crafting them remains a laborious process and demands expertise in both data visualization and digital fabrication. We analyzed the limitations of current workflows through three real case studies and created *MakerVis*, the first tool that integrates the entire workflow, from data filtering to physical fabrication (figure 8 ). Design sessions with three end users showed that tools such as *MakerVis* can dramatically lower the barriers behind producing physical visualizations. Observations and interviews also revealed important directions for future research. These include rich support for customization, and extensive software support for materials that accounts for their unique physical properties as well as their limited supply.

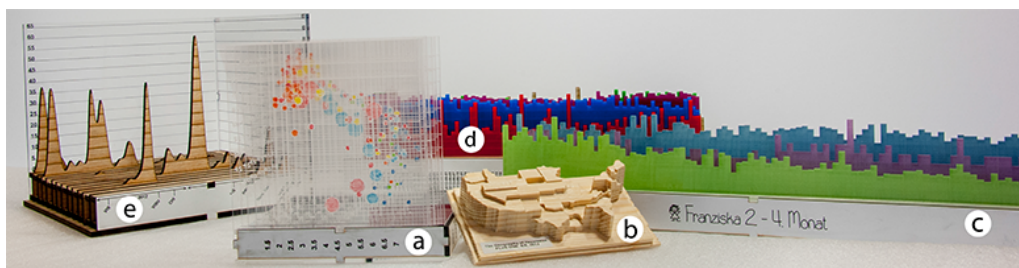


Figure 8. Physical visualizations created with *MakerVis*: a) a scatterplot created after Hans Rosling's TED talk, b) a prism map showing happiness across the US computed from Twitter sentiments, c),d),e) visualizations created by users during design sessions.



## 6.4. Engineering of interactive systems

**Participants:** Caroline Appert, Michel Beaudouin-Lafon [correspondant], Olivier Chapuis, Cédric Fleury, Stéphane Huot, Theophanis Tsandilas, Wendy Mackay.

InSitu has a long tradition of developing software tools and user interface toolkits to facilitate the creation of interactive systems. These tools allow us to better experiment with our ideas and are therefore an integral part of our research methodology. Most of them are freely available and some are used outside InSitu for research or teaching.

*Interactive Paper* – We created *PaperComposer* [31], a graphical interface builder for creating personal interactive-paper applications for musical creation. We also built an API that facilitates the development of interactive paper components for PaperComposer. The API enables developers to define new paper components that accept additional musical data with their own representation structures and interactions.

*3D Telepresence* – In the context of 3D telepresence, we studied how to transmit a 3D model of the users to a remote location. In [17] we present a 3D head reconstruction method for low cost 3D telepresence systems that uses only a single consumer level hybrid sensor (color+depth) located in front of the users. Our method fuses the real-time, noisy and incomplete output of a hybrid sensor with a set of static, high-resolution textured models acquired in a calibration phase (figure 9). A complete and fully textured 3D model of the users' head can thus be reconstructed in real-time, accurately preserving the facial expression of the user. The main features of our method are a mesh interpolation and a fusion of a static and a dynamic textures to combine respectively a better resolution and the dynamic features of the face.

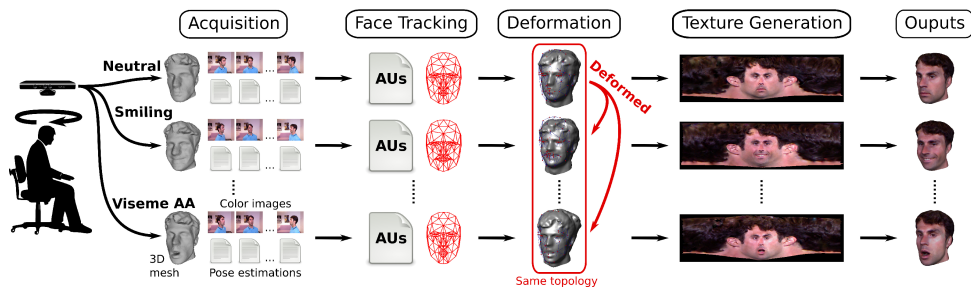


Figure 9. Acquisition step of the 3D face reconstruction: data are processed to create a complete and fully textured 3D head model for each facial expression. This set of head models are then used to improve the real-time reconstruction of the user's head during a 3D telepresence session.

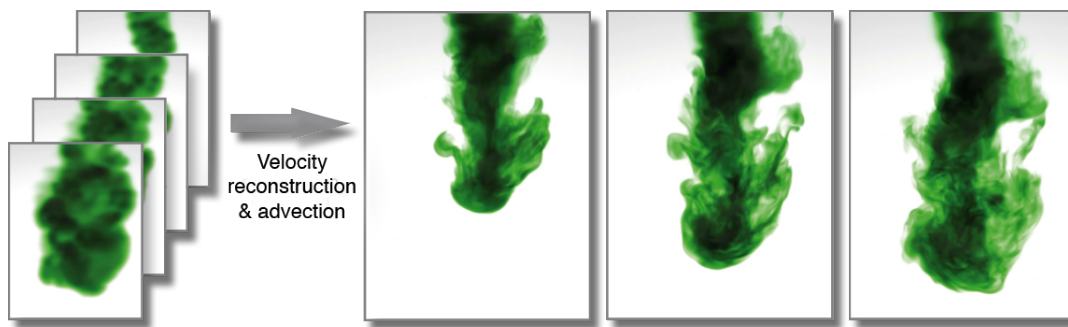
*Wall-sized displays* – We developed Smarties [16], a system that allows developers to easily add interactive support to their wall-sized display applications by using mobile devices such as tablets. The system includes an original mobile interface that can be customized by the application itself (without programming the mobile device), a communication protocol between the mobile devices and the application running on the wall-sized display, and libraries in different programming languages that implement the protocol and handle synchronization, locking and input conflicts. Synchronization between multiple mobile devices is handled by the libraries, and thus the system supports free collaboration. The mobile devices come with multiple cursor controllers, also associated with keyboards, widgets and clipboards.

## MANAO Project-Team

### 5. New Results

#### 5.1. Highlights of the Year

We are still developing our expertise in fitting techniques. As an illustration, we have solved of a long-standing problem in fluid capture: the non-invasive three-dimensional digitization of dynamic gas flows including their three-dimensional velocity fields [17] (cf. Figure 8 ). We solve the three-dimensional flow tracking problem by fitting a full 3D Navier-Stokes simulation to the acquired data. To our knowledge, this is a world-first in this area that considerably improves the results by incorporating high-level prior knowledge into the estimate. The resulting mathematical framework can be generalized easily and lends itself to editing operations. The technique has applications, e.g., in aerospace engineering. We are exploring the possibilities with ONERA, the French space agency. In fact, parts of the developed techniques have been validated by them and are now being installed in a wind tunnel facility for real-world tests.



*Figure 8. Low-resolution captures obtained by tomographic scanning (left) are used as inputs to our method which estimates physically plausible dense velocity fields. Such velocity fields fully determine the fluid state and can be applied in a variety of applications including fluid super-resolution (right) allowing capture to be integrated into pipelines for visual effects simulation.*

This year, the collaboration between Optics and Computer Graphics has grown to a now long-term project, under the initiative of the MANAO team. First, from an institutional point of view, a framework agreement has been signed the 10th of July 2014 between the IOGS and Inria. This is an important and institutional recognition of the potential trans-disciplinary impacts of our work. Second, we have begun to set-up the COEL experimentation facility inside the LP2N laboratory. It has been made possible thanks to the support of the "Région Aquitaine" and upcoming supports from l'"Initiative d'excellence de l'université de Bordeaux". With this trans-disciplinary experimentation facility – rather unique in Europe – we can now put into practice a long-term vision of the researches that we want to achieve.

In term of visibility, we managed to published our first paper in the Optics scientific community [15], highlighting our trans-disciplinary research. We have also been part of the final and transnational exhibition of the V-Must.net network of excellence: Keys2Rome - <http://keys2rome.eu>. It was launched simultaneously in Rome, Sarajevo, Amsterdam and Alexandria on September 23, 2014. The exhibition uses immersive technology to present and connect these regional cultures within the Roman Empire, highlighting their diversity and commonality over centuries of Roman rule. Our spatial augmented reality solution [21] was included in this event.



## 5.2. Analysis and Simulation

### 5.2.1. Importance Sampling of Realistic 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 environment maps and 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 [74]. This is a foundation work that has led to new sampling techniques for physically-based rendering with time-varying environment lighting [73] and light field light sources. The results show that physically accurate rendering with realistic light sources can be achieved in real time.

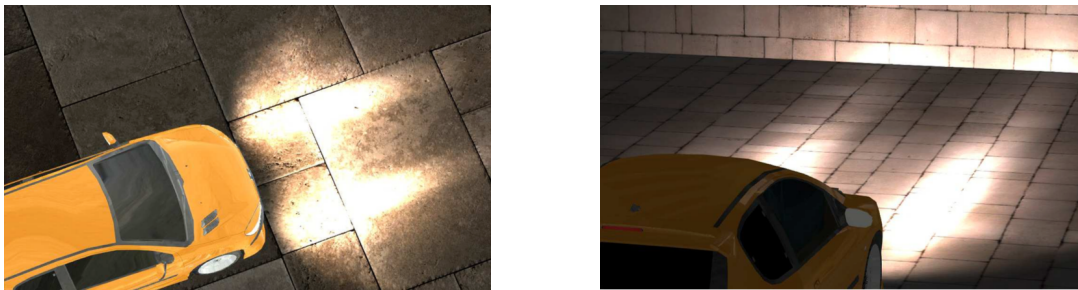


Figure 9. Our new light importance sampling technique estimates direct lighting interactively (7-9 fps) with only 200 samples per pixel that are distributed among the different images of the light field luminaire. The car headlights are represented by the same light field composed of  $11 \times 9$ 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

### 5.2.2. Frequency Analysis of Light Scattering and Absorption

We have proposed [14] an innovative analysis of absorption and scattering of local light fields in the Fourier domain, and derived the corresponding set of operators on the covariance matrix of the power spectrum of the light field. This analysis brings an efficient prediction tool for the behavior of light along a light path in participating media. We leverage this analysis to derive proper frequency prediction metrics in 3D by combining per-light path information in the volume. Our key contribution is to show that analyzing local light fields in the Fourier domain reveals the consistency of illumination in such media, and provides a set of simple and useful rules to be used to accelerate existing global illumination methods.

## 5.3. Acquisition and Display

### 5.3.1. Three-Dimensional, Dynamic, Full State Fluid Capture and Manipulation

Participant: I. Ihrke

We have explored [17] the connection between fluid capture, simulation and proximal methods, a class of algorithms commonly used for inverse problems in image processing and computer vision. Our key finding is that the proximal operator constraining fluid velocities to be divergence-free is directly equivalent to the pressure-projection methods commonly used in incompressible flow solvers. This observation lets us treat the inverse problem of fluid tracking as a constrained flow problem all while working in an efficient, modular framework. In addition it lets us tightly couple fluid simulation into flow tracking, providing a global prior that significantly increases tracking accuracy and temporal coherence as compared to previous techniques. We demonstrate how we can use these improved results for a variety of applications, such as re-simulation, detail enhancement, and domain modification. We furthermore give an outlook of the applications beyond fluid tracking that our proximal operator framework could enable by exploring the connection of deblurring and fluid guiding.

### 5.3.2. *Measurements and Analysis of Retro-reflective Materials*

**Participants:** L. Belcour, R. Pacanowski

We have compared [15] performance of various analytical retro-reflecting BRDF models to assess how they reproduce accurately measured data of retro-reflecting materials. We have also introduced a new parametrization, the back vector parametrization, to analyze retro-reflecting data and we have shown that this parametrization better preserves the isotropy of data. Furthermore, we have updated existing BRDF models to improve the representation of retro-reflective data. This work was supported by the development of the ALTA library [23].

### 5.3.3. *Kaleidoscopic Imaging*

**Participants:** I. Reshetouski, I. Ihrke

Kaleidoscopes have a great potential in computational photography as a tool for redistributing light rays. In time-of-flight imaging the concept of the kaleidoscope is also useful when dealing with the reconstruction of the geometry that causes multiple reflections. Our work [13] is a step towards opening new possibilities for the use of mirror systems as well as towards making their use more practical. The focus of this work is the analysis of planar kaleidoscope systems to enable their practical applicability in 3D imaging tasks. We have analyzed important practical properties of mirror systems and developed a theoretical toolbox for dealing with planar kaleidoscopes. Based on this theoretical toolbox, we have explored the use of planar kaleidoscopes for multi-view imaging and for the acquisition of 3D objects [90]. The knowledge of the mirrors positions is crucial for these multi-view applications. On the other hand, the reconstruction of the geometry of a mirror room from time-of-flight measurements is also an important problem. We therefore employ the developed tools for solving this problem using multiple observations of a single scene point.

### 5.3.4. *Interactive Spatial Augmented Reality*

**Participants:** B. Ridel, P. Reuter, X. Granier

We have proposed the *Revealing Flashlight* [21], a new 6-degree-of-freedom interaction and visualization technique in spatial augmented reality that helps to reveal the details of cultural heritage artifacts. We locally and interactively highlight them by projecting an expressive visualization. The Revealing Flashlight can be used by archaeologists, for example, to help decipher inscriptions in eroded stones, or by museums (cf. Figure 10 ) to let visitors interactively discover the features and meta-information of cultural artifacts. A permanent exhibition is now running at the Allard Pierson Museum, and others museums are asking us to set-up similar installations. It was part of the final trans-European showcase of the V-MusT.net project.

## 5.4. **Rendering, Visualization & Illustration**

### 5.4.1. *Computing Smooth Surface Contours with Accurate Topology*

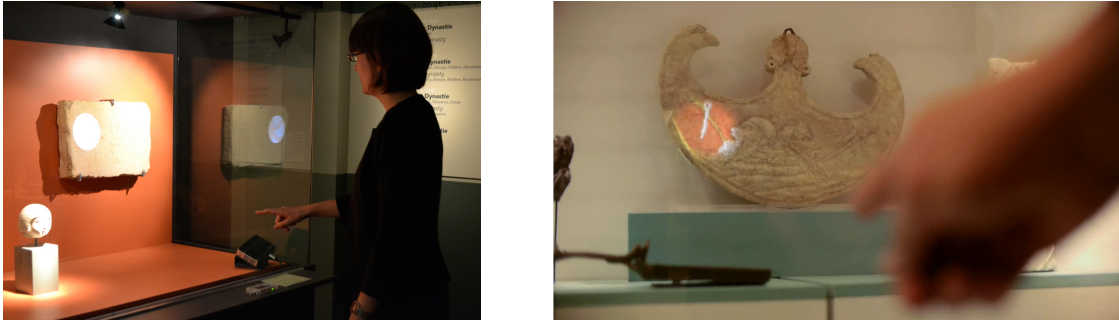


Figure 10. “The Revealing Flashlight” lets visitors explore ancient artifacts interactively. (Left) Allard Pierson Museum - Amsterdam. (Right) Keys2Rome exhibition in Museo dei Fori Imperiali - Roma.

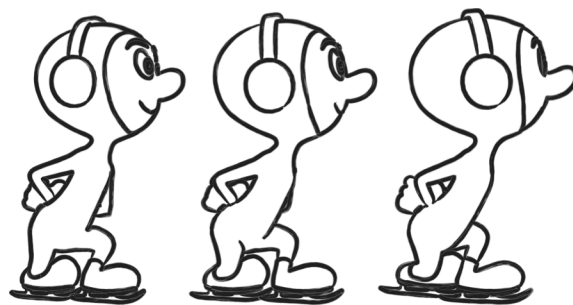


Figure 11. Contours stylized with tapered strokes [16]. Our method avoids classical breaks and gaps, producing more coherent animated strokes. Red © Pixar

We have introduced [16] a method for accurately computing the visible contours of a smooth 3D surface for stylization. This is a surprisingly difficult problem, and previous methods are prone to topological errors, such as gaps in the outline. Our approach is to generate, for each viewpoint, a new triangle mesh with contours that are topologically-equivalent and geometrically close to those of the original smooth surface. The contours of the mesh can then be rendered with exact visibility. The core of the approach is Contour-Consistency, a way to prove topological equivalence between the contours of two surfaces. Producing a surface tessellation that satisfies this property is itself challenging; to this end, we introduce a type of triangle that ensures consistency at the contour. We then introduce an iterative mesh generation procedure, based on these ideas. This procedure does not fully guarantee consistency, but errors are not noticeable in our experiments. Our algorithm can operate on any smooth input surface representation; we use Catmull-Clark subdivision surfaces in our implementation.

## 5.5. Editing and Modeling

### 5.5.1. Tomography-Based Volume Painting

**Participant:** I. Ihrke

Although volumetric phenomena are important for realistic rendering and can even be a crucial component in the image, the artistic control of the volume's appearance is challenging. Appropriate tools to edit volume properties are missing, which can make it necessary to use simulation results directly. Alternatively, high-level modifications that are rarely intuitive, e.g., the tweaking of noise function parameters, can be utilized. We have introduced [18] a solution to stylize single-scattering volumetric effects in static volumes. Hereby, an artistic and intuitive control of emission, scattering and extinction becomes possible, while ensuring a smooth and coherent appearance when changing the viewpoint. Our method is based on tomographic reconstruction, which we link to the volumetric rendering equation. It analyzes a number of target views provided by the artist and adapts the volume properties to match the appearance for the given perspectives. Additionally, we describe how we can optimize for the environmental lighting to match a desired scene appearance, while keeping volume properties constant. Finally, both techniques can be combined. We demonstrate several use cases of our approach and illustrate its effectiveness.

### 5.5.2. Implicit Skinning

**Participant:** G. Guennebaud

In collaboration with IRIT (Toulouse), we extended our *implicit skinning* method to a new approach for interactive character skinning called *elastic implicit skinning*. The method simulates skin contacts between limbs as well as the effect of skin elasticity (Figure 12). In addition, we go a step further towards the automation of the rigging process: our method doesn't require the definition of skinning weights. Elastic implicit skinning takes the best features of the recent implicit skinning method, and makes it robust to extreme character movements. While keeping the idea of implicit skinning, namely approximate the character by 3D scalar fields in which mesh-vertices are appropriately re-projected, we depart from the processing pipeline used so far. Implicit skinning is history independent and uses an initial skinning solution (e.g., linear blending or dual quaternions) to correct vertex positions at each frame. Our new approach is history dependent; the mesh directly tracks the iso-surfaces of the scalar field over time. Technically our solutions include: new implicit surface composition operators and a tangential relaxation scheme derived from the as-rigid-as possible energy. This work [101] has been presented at SIGGRAPH Asia this year.

### 5.5.3. Multi-scale Editing

**Participant:** G. Guennebaud

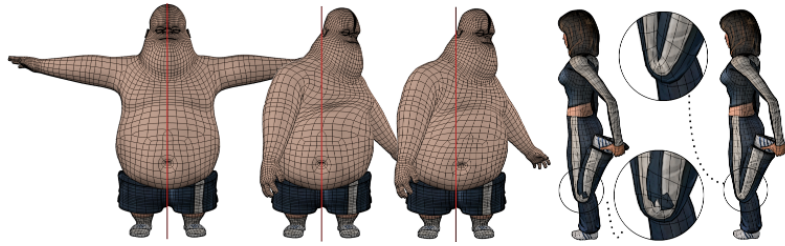


Figure 12. Illustration of the benefits of our novel elastic implicit skinning technique.

In the continuation of our Growing Least Square approach [5] for the multi-scale analysis of shape, we developed a novel tool that enables the direct editing of surface features in large point-clouds or meshes [19]. This is made possible by a novel multi-scale analysis of unstructured point-clouds that automatically extracts the number of relevant features together with their respective scale all over the surface. Then, combining this ingredient with an adequate multi-scale decomposition allows us to directly enhance or reduce each feature in an independent manner. Our feature extraction is based on the analysis of the scale-variations of locally fitted surface primitives combined with unsupervised learning techniques. Our tool may be applied either globally or locally, and millions of points are handled in real-time. The resulting system enables users to accurately edit complex geometries with minimal interaction.

#### 5.5.4. Manipulation of Anisotropic Highlights

**Participants:** B. Raymond, P. Barla, G. Guennebaud, X. Granier

We have developed [20] a system for the direct editing of highlights produced by anisotropic BRDFs, which we call anisotropic highlights. We first provide a comprehensive analysis of the link between the direction of anisotropy and the shape of highlight curves for arbitrary object surfaces. The gained insights provide the required ingredients to infer BRDF orientations from a prescribed highlight tangent field. This amounts to a non-linear optimization problem, which is solved at interactive framerates during manipulation. Taking inspiration from sculpting software, we provide tools that give the impression of manipulating highlight curves while actually modifying their tangents. Our solver produces desired highlight shapes for a host of lighting environments and anisotropic BRDFs.

## MAVERICK Project-Team

# 5. New Results

## 5.1. Highlights of the Year

The impacting PhD work [3] of Eric Heitz on *appearance filtering* (see section 5.5.1 ) has received a very good reception in both academic and industrial world, including several “best paper” prizes in 2013 and 2014, invitation to participate to the Siggraph Course on Photorealistic Rendering [13], and statements of importance and/or integration by reference peoples and CG companies.

BEST PAPER AWARD :

[] **Computer Graphics Forum**. E. HEITZ, E. D'EON.

## 5.2. Visual perception

### 5.2.1. *The effects of surface gloss and roughness on color constancy for real 3-D objects*

**Participants:** Jeoren J. M. Granzier, Romain Vergne [contact], Karl Gegenfurtner.

Color constancy denotes the phenomenon that the appearance of an object remains fairly stable under changes in illumination and background color. Most of what we know about color constancy comes from experiments using flat, matte surfaces placed on a single plane under diffuse illumination simulated on a computer monitor. Here we investigate whether material properties (glossiness and roughness) have an effect on color constancy for real objects. Subjects matched the color and brightness of cylinders (painted red, green, or blue) illuminated by simulated daylight (D65) or by a reddish light with a Munsell color book illuminated by a tungsten lamp. The cylinders were either glossy or matte and either smooth or rough. The object was placed in front of a black background or a colored checkerboard as shown in Figure 6. We found that color constancy was significantly higher for the glossy objects compared to the matte objects, and higher for the smooth objects compared to the rough objects. This was independent of the background. We conclude that material properties like glossiness and roughness can have significant effects on color constancy [7].

## 5.3. Visualization

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

In computer visualization we have worked on two topics: topology for visualization and perception for visualization.

In topology for visualization we have worked on scalar field visualization methods taking into account the topology of the data. In [14] We have derived theoretical results on monotonic interpolation of scalar data. Our method enables to interpolate given topological data such as minima, maxima and saddle points at the corners of a rectangular domain without adding spurious extrema inside the function domain, as illustrated in Figure 7

We have collaborated to a state of the art chapter on Uncertain Visualization [15], in which we described the evaluation of visualization methods based on visual perception.

Furthermore we have worked on two topics related to geometry for visualization. In [6] we introduce a method for interpolating a quad mesh using G1-continuous polynomial surfaces. We plan to use this method in the future for displaying isosurfaces of higher order data. In [11] we have published a method for reconstructing interfaces in highly complex assemblies, as illustrated in Figure 8. This method has been developed in order to visualize data arising from simulation of complex mechanical assemblies, within the ANR project ROMMA, closed in January 2014.



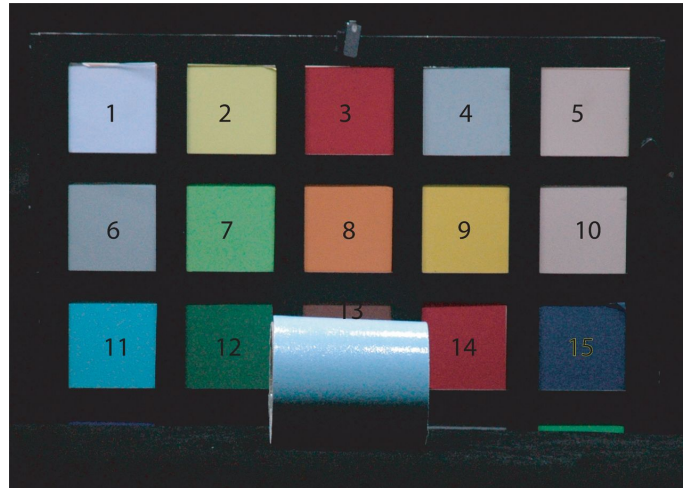


Figure 6. Color perception depends on material properties. This image represents one stimulus used in our experiment to compare the effect of glossiness on color constancy.

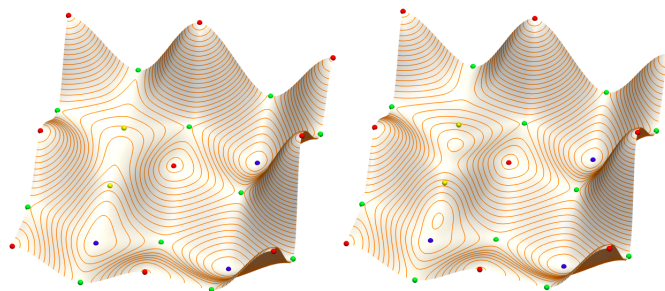


Figure 7. Local maxima (red), minima (blue), saddles (green) and regular (yellow) vertices are interpolated by a  $C1$  piecewise cubic interpolant. Left: no unwanted local extrema exist in the interior of the cubic patches. Right: partial derivatives too large in size are chosen for the yellow regular vertices implying that additional unwanted local extrema appear inside the cubic polynomial patches.

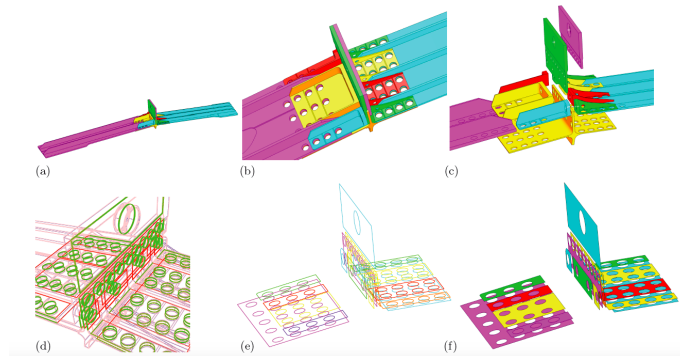


Figure 8. Aircraft part for assembling the wings with the body of an aircraft (model courtesy of EADS). (a,b) two views of the components, (c) exploded view, (d) ray casting results, (e) boundary reconstruction, (f) nal interfaces

## 5.4. Image creation and editing

### 5.4.1. 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.

### 5.4.2. Color transfer guided by summary statistics

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

Modifying the colors of an image is an attractive way to edit its ambiance and mood. In practice, manually and directly tuning the color distribution of an image is challenging and tedious. Color transfer methods offer an intuitive alternative by automatically changing an image colors according to a target image. Existing transfer methods mostly rely on global matching processes to reshape and map the color histogram of the source image as close as possible to the target histogram. However, they offer no control over where the colors of the target will be transferred in the source image: they only tend to match colors that have similar intensities and chromaticities. This can lead to unexpected results, especially when some elements do not have the same colors in the two images, but share similar features. In this work, we propose to implicitly segment input images before transferring colors. Instead of relying on colors only, we use a summary of statistics to describe the underlying texture properties of each pixel. This provides a measure of pixel similarity which is then used to guide and ensure the transfer to be done between similar features (see Figure reffig:color for a preliminary result).



Source

Target

Transfer

Figure 9. By transferring colors between statistically similar pixels, meaningful colors are transferred from the target to the source image.

## 5.5. Complex scenes

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

### 5.5.1. Surfacic appearance pre-filtering

**Participants:** Eric Heitz, Fabrice Neyret [contact].

Here, we deal with complex surfaces represented by microfacets and material attributes.

Among the various correlations between material ingredients forming the BRDF, we published an extended version of the work on correlation between surface attribute (like color) and visibility [9], and Eric published an comprehensive interpretation of the microfacet model in a journal in the field of physics [8]. He also adapted his microfacet approach to the efficient BRDF sampling for path tracing – published in EGSR/CGF [10], see Figure 11, and he was invited to participate to the prestigious Siggraph Course “Physically Based Shading in Theory and Practice” [13]. This work is now implemented in various professional and standard software and thus settled a new standard. Eric defended his PhD on September, 26 2014 [3]

### 5.5.2. Volumetric appearance pre-filtering

**Participants:** Guillaume Loubet, Fabrice Neyret [contact].

Here, we deal with complex density distributions. The first target is galactic material in the scope of the *veRTIGE / Galaxy ANR* project, but the long term goal is more general since at long distance complex surfaces or scattered objects can more efficiently be represented as volumetric distributions.

The usual hypothesis in CG is that volumes are homogeneous distribution of matter. But star and (dark) dust distributions are fractal, not homogeneous. This breaks all the existing equations accounting for large scale opacity and lighting of volumes of such material.

first, we developed a new procedural noise able to easily mimic such fractal distributions according to astrophysical models (see fig 12 ,a). Then we studied how to reproduce the same opacity (fig 12 ,b) and reflectivity (fig 12 ,c) for various level of details – this is still ongoing work.

Moreover, volumetric material is often concentrated into bodies, with a boundary delimited by a density jump or gradient. We studied the macroscopic light behavior in such configurations (fig 12 ,d).

## 5.6. Realistic rendering

Note that Cyril Soler defended his HDR “*Models and Analyses for Image Synthesis*”, Université Joseph-Fourier, on June 2014.

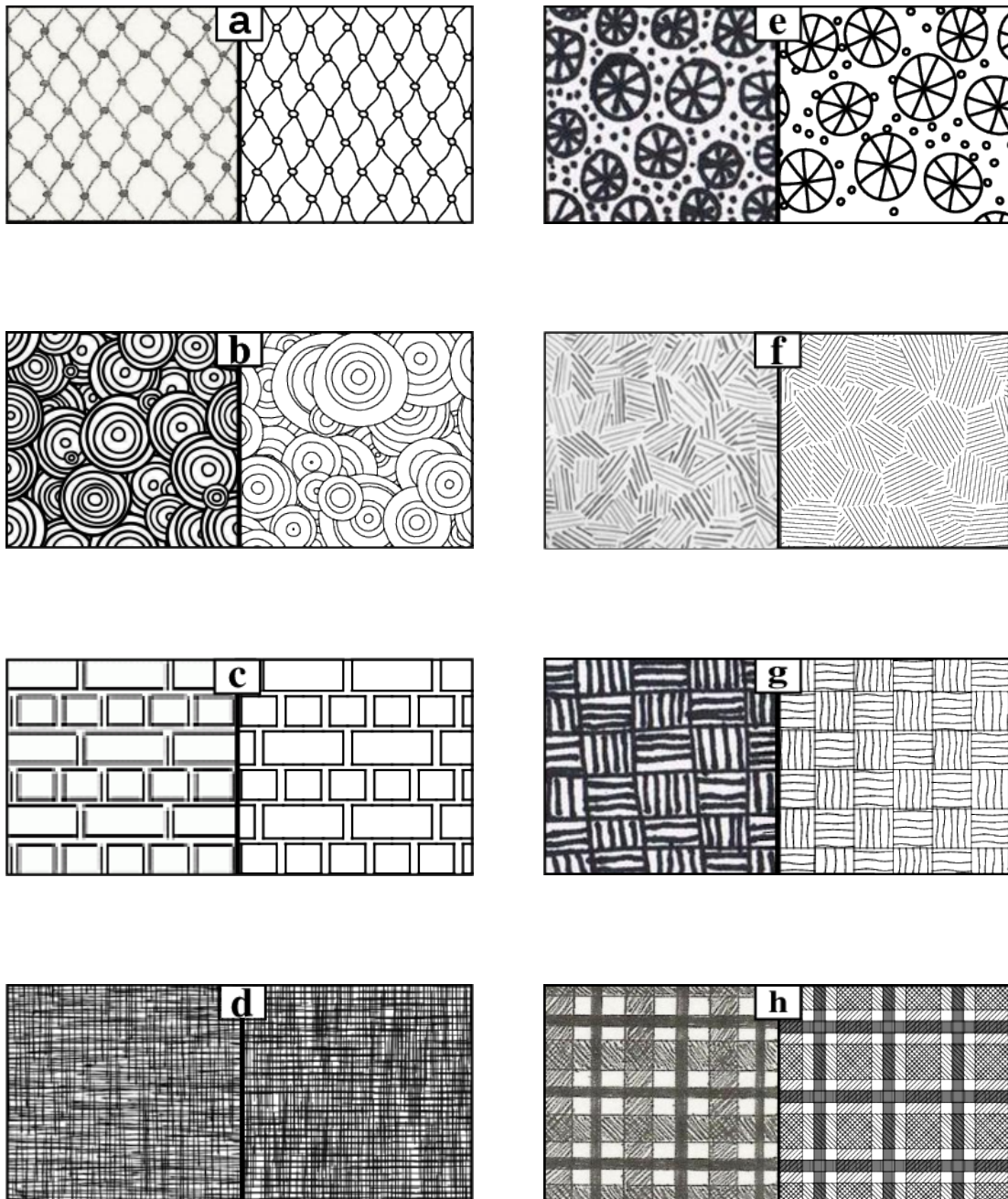


Figure 10. **Element textures commonly used.** These textures can be found in professional art (d,g,h), casual art (a,e,f), technical productions such as Computer-Assisted Design illustration tools (c), and textile industry (b). For each example, we show a hand-drawn image (left), and our synthesized 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” [20]; (a,e) Profusion Art [profusionart.blogspot.com]; (f) Hayes’ Art Classes [hayesartclasses.blogspot.com]; (c) CompugraphX [www.compugraphx.com]; (b) 123Stitch [www.123stitch.com].



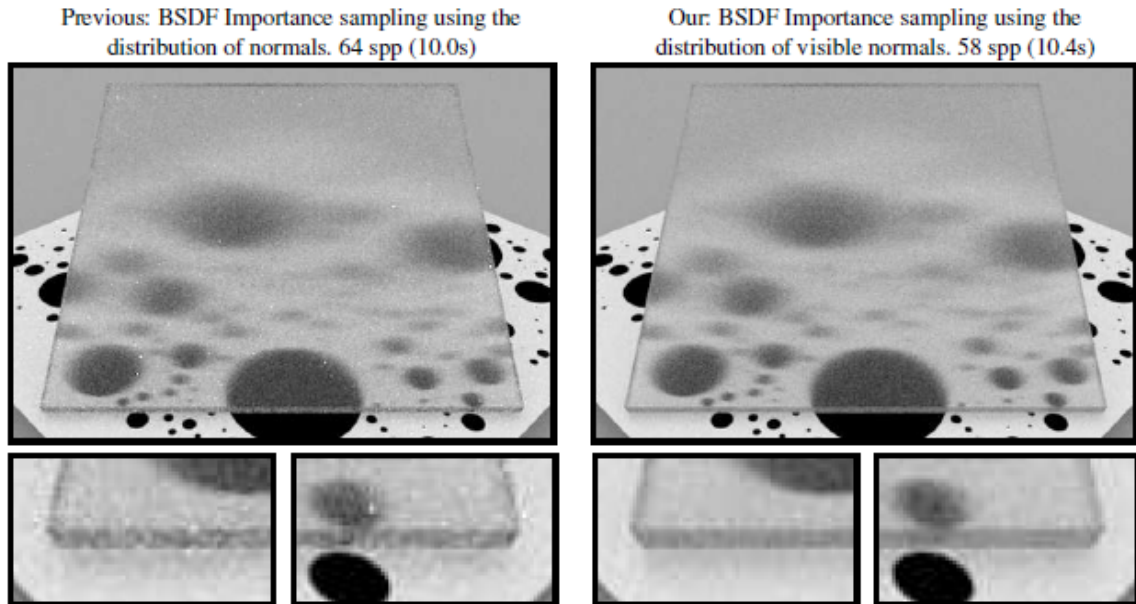


Figure 11. A dielectric glass plate ( $n = 1.5$ ) with anisotropic GGX roughness ( $a_x = 0.05$ ,  $a_y = 0.4$ ) on all faces (with the Smith masking function). For a similar sample budget and the same render time, our method (right) significantly reduces the variance and converges faster than the common technique used in previous work (left).

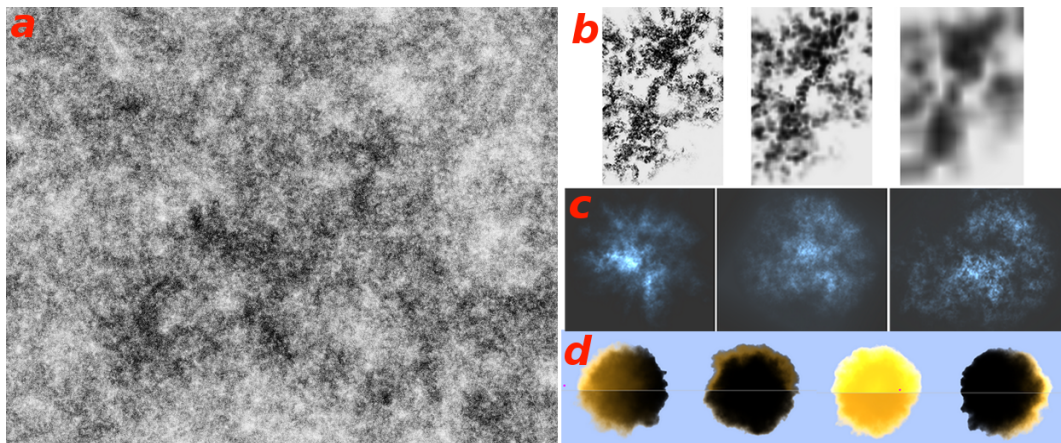


Figure 12. a: Our new fractal procedural noise. b: Multiscale opacity. c: Multiscale reflectance. d: Light reflection at volumetric bodies boundary with gradient (top) or jump (bottom) density distribution, with different light direction (left to right).

### 5.6.1. Single Scattering in participating media with refractive boundaries

**Participant:** Nicolas Holzschuch [contact].

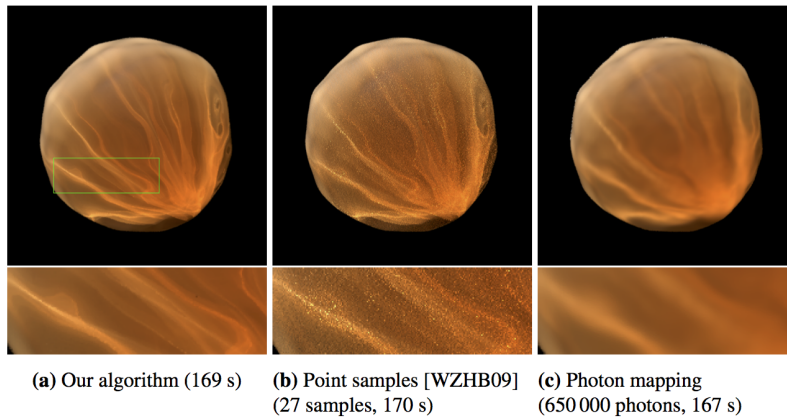


Figure 13. Single scattering: comparison between our algorithm and existing methods (equal computation time) on a translucent sphere illuminated by a point light source from behind.

Volume caustics are high-frequency effects appearing in participating media with low opacity, when refractive interfaces are focusing the light rays (see Figure 13). 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 accepted for publication at Computer Graphics Forum [10].

### 5.6.2. A Local Frequency Analysis of Light Scattering and Absorption

**Participants:** Laurent Belcour, Kavita Bala, Cyril Soler [contact].

We proposed a novel analysis of absorption and scattering of local light fields in the Fourier domain in the neighborhood of light paths. This analysis aims at predicting the changes over the distribution of light energy, so as to allow efficient sampling and integration methods of diffused light in participating media. Our analysis explains that absorption increases frequency since it acts as a continuous visibility mask over the local light field, and that scattering lowers frequencies as it operates a low pass convolution filter in the directional domain. In order to combine this analysis with our previous work on covariance tracing—and therefore use it to improve existing algorithms for path tracing in participating media—we derived new sampling metrics all based on a common prediction of the 3D covariance of the fluence in the volume. We demonstrate indeed that the covariance of the fluence can efficiently be computed by combining the 4D covariance matrices of light fields in the neighborhood of light paths, and that it can be used to compute effective metrics (1) for the variance of energy collected along camera rays, (2) for determining the shape and size of reconstruction kernels in screen space, and (3) for drastically improving the convergence of density estimation methods. For the later, we propose an improvement of the method of Progressive Photon Beams. This work has been published in ACM Transactions on Graphics and presented at Siggraph'2014 in Vancouver [5].



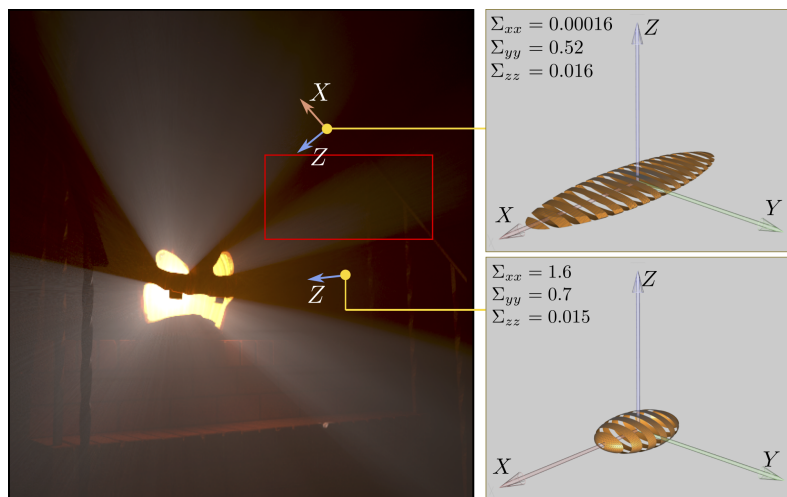


Figure 14. Predictions of the covariance of the Fourier spectrum of the fluence in the volume computed using our Fourier analysis of scattering and absorption.

## MIMETIC Project-Team

# 6. New Results

## 6.1. Highlights of the Year

### 6.1.1. *Link between performance and risk of injury*

**Participants:** Richard Kulpa [contact], Benoit Bideau, Michaël Ropars.

In our previous biomechanical analysis of the tennis serve, we have demonstrated that the energy flow is a pathomechanical factor, that means that it can increase joint constraints (and thus risk of injury) while not increasing performance. Nevertheless, the definition and evaluation of energy flow is still a complex scientific challenge. We have proposed to compare the energy flow during the serve between injured and non-injured tennis players by investigating the relationships between the quality and magnitude of energy flow, the ball velocity and the peaks of upper limb joint kinetics [11]. The results showed that ball velocity increased and upper limb joint kinetics decreased with the quality of energy flow from the trunk to the ‘hand+racket’. Injured players showed a lower quality of energy flow through the upper limb kinetic chain, a lower ball velocity and higher rates of energy absorbed by the shoulder, the elbow than non-injured players. These findings imply that an effective energy flow through the kinetic chain by using a proper serve technique is necessary for reducing overuse joint injury risks.

### 6.1.2. *ACM SIGGRAPH Course on crowd simulation*

**Participant:** Julien Pettré [contact].

Crowds for entertainment or safety applications purposes are most of the time simulated using microscopic algorithms. In contrast with other types of approaches, microscopic simulators are able to generate continuous and smooth trajectories for individual agents. They are based on models of local interactions between agents. The crowd motion result from the combination of all local motion and interactions. The fact that the resulting crowd motion is emergent makes difficult anticipating the simulation results. Many motion and interaction models have been proposed to design a plethora of simulation algorithms: force-based models, rule-based models, coupled or not with flow-based models, etc. Each type of interaction models will actually result into specific crowd motions as well as agents trajectories. Unfortunately, not all have the desired properties: oscillations, jerky trajectories, residual collisions or deadlocks are often observed in simulations. From this point of view, the course [28] presents the many recent progresses in crowd simulation algorithms since the introduction of velocity-based algorithms, as well as the impact on the level of realism and the visual quality of simulated crowd motions. It also presents the impact on various kind of applications.

### 6.1.3. *Immersive basketball playing*

**Participants:** Franck Multon [contact], Alexandra Covaci, Anne-Hélène Olivier.

The paper has received the best paper award of the ACM VRST 2014 Conference in November 2014. This paper addressed the problem of perception of distances in immersive environments when dealing with precision distant tasks, such as basketball free throw. The work has been done in collaboration with University of Brasov in Romania, thanks to the FP7 VISIONAIR infrastructure project. The main results of this work tend to show that third person perspectives enabled subjects to perform the task with similar movements than in real world, compared to first person perspective. Third person perspective consists in placing the camera at a different place from the eye point of view, as in many videogames. On the opposite first person perspective consists in place the camera at the place of the user’s eyes in scale 1:1, as if the user was colocalized in the virtual environment. We also demonstrated an adaptation to the task in immersive environments, which is a key information for future development of training methods using VR. We have been invited to submit an extended version of the paper to the IEEE Computer and Graphics journal for 2015.

BEST PAPERS AWARDS :

[21] ACM Symposium on Virtual Reality Software and Technology VRST. A. COVACI, A.-H. OLIVIER, F. MULTON.

## 6.2. VR and Sports

**Participants:** Richard Kulpa [contact], Benoit Bideau, Franck Multon.

Previous works in MimeTIC have shown the advantage of using VR to design and carry-out experiments on perception-action coupling in sports, especially for duels between two opponents. However the impact of using various technical solutions to carry-out this type of experiment in sports is not clear. Indeed immersion is performed by using interfaces to capture the motion/intention of the user and to deliver various multi-sensory feedbacks. These interfaces may affect the perception-action loop so that results obtained in VR cannot be systematically transferred to real practice.

Most of the applications in VR provide the user with visual feedbacks in which the avatar of the user can be more or less simplified (sometimes limited to a hand or the tools he is carrying). In first person view in caves the user generally does not need accurate avatars as he can perceive his real body but some authors have shown that the perception of distances is generally modified. Some authors have also demonstrated that first person view was less efficient than third person view with avatars when performing accurate tasks such as reaching objects in constrained environments. We proposed an experiment to evaluate which type of feedback was the most appropriate one for complex precision tasks, such as basketball free-throw. In basketball free-throw the user has to throw a ball into a small basket placed at over 4.5m far from him. Thus perception of distance is actually a key point in such a task. Beginners and experts carried-out a first experiment in real in order to measure their motion and performance in real situation. Then beginners were asked to perform free throws with a real ball in hands, but in three conditions in a Cave (Immersia): 1) first person view (see Figure 4), 2) third-person view with the visual feedback of the ball's position, and 3) third-person view the virtual ball and additional rings modeling the perfect trajectory for the ball to get in the basket. Results show that significant differences exist in ball speed between first person view condition compared to real condition whereas no differences exist in third-person view conditions. If we focus on successful throws only, ball speed in the last condition 3) was very similar to real condition whereas all the other VR conditions (1) and 2)) lead to significant differences compared to real situation. In all VR conditions the height of ball release was significantly higher in VR compared to real situation. These results show that VR conditions lead to adaptations in the way people perform such a precision task, especially for ball speed and height of ball release. However this difference is significantly higher with first person view and tends to zero in condition 3). Future works will tend to evaluate new conditions with avatars and complementary points of view (such as lateral and frontal views together as suggested by some authors). It will also be important to more clearly understand the problem of perception of distances in such an environment. This work has been performed in cooperation with University of Brasso in Romania [21]. This paper has received the best paper award of the ACM VRST 2014 Conference.

Another key feedback is the external forces associated with the task. In most sports applications such forces are strongly linked to performance. However delivering these forces in virtual environments is still a challenge as it required haptic devices that could affect the way the users perform the task (with a different grip compared to real situation and limitations in dynamic response of the device). Pseudohaptic has been introduced in the early 2000. It consists in using visual feedbacks to make people perceive the forces linked to a task. However this approach has not been tested for whole-body interaction. In collaboration with Hybrid team in Inria Rennes, we studied how the visual animation of a self-avatar could be artificially modified in real-time in order to generate different haptic perceptions. In our experimental setup participants could watch their self-avatar in a virtual environment in mirror mode. They could map their gestures on the self-animated avatar in real-time using a Kinect. The experimental task consisted in a weight lifting with virtual dumbbells that participants could manipulate by means of a tangible stick. We introduce three kinds of modification of the visual animation of the self-avatar: 1) an amplification (or reduction) of the user motion (change in C/D ratio), 2) a change in the dynamic profile of the motion (temporal animation), or 3) a change in the posture of the avatar (angle of inclination). An example is depicted in Figure 5. Thus, to simulate the lifting of a "heavy" dumbbell, the



Figure 4. First-person view condition in the basket free-throw.

avatar animation was distorted in real-time using: an amplification of the user motion, a slower dynamics, and a larger angle of inclination of the avatar. We evaluated the potential of each technique using an ordering task with four different virtual weights. Our results show that the ordering task could be well achieved with every technique. The C/D ratio-based technique was found the most efficient. But participants globally appreciated all the different visual effects, and best results could be observed in the combination configuration [9]. Our results pave the way to the exploitation of such novel techniques in various VR applications such as for sport training, exercise games, or industrial training scenarios in single or collaborative mode.



Figure 5. Weight discrimination task: the animation of the avatar showed a lifting effort according to the weight of the virtual dumbbell and the user has to rank the feedbacks from the lightest to the heaviest mass.

### 6.3. Motion Sensing and analysis

**Participant:** Franck Multon [contact].

Sensing human activity is a very active field of research, with a wide range of applications ranging from entertainment and serious games to personal ambient living assistance, including rehabilitation. MimeTIC aims at proposing original methods to process raw motion capture data in order to compute relevant information according to the application.

In rehabilitation, we have collaborated with University of Montreal, Saint-Justine Hospital which main activity is rehabilitation of children with pathologies of the pyramidal control system. In this domain, defining metrics and relevant measurement to diagnose pathologies and to monitor patients during treatment is a key point. In gait, most of the previous works focus on gait spatio-temporal parameters (such as step length, frequency, stride duration, global speed...) which could be measured with two main families of systems: 1) one-point measurement with a force plate, one accelerometer or dedicated devices (such as a Gait Ride), or 2) multi-point measurement systems with motion sensors or markers placed over the patient's skin. The former provides the clinician with compact but incomplete knowledge whereas the latter provides him with numerous data which are sometimes difficult to analyze and to get (specific technical skills are required). The first step to any type of analysis is to detect the main gait events, such as foot strikes and toe offs. In treadmill walking, widely used in rehabilitation as it enables the clinician to analyze numerous gait cycles in a limited place with a controlled speed, automatically detecting such gait events requires complex devices with specific technical skills (such as calibration and post-processing with motion capture systems) [1].

Recent papers have demonstrated that low-cost and easy-to-use depth cameras (such as a Kinect from Microsoft) look promising for serious applications requiring motion capture. However there exist some confusion between the feet and the ground at foot strike and foot off leading to bad estimation of the gait cycle events. We have proposed an alternative approach that consists in using the strong correlation between knee and foot trajectories to deduce foot strikes thanks to knee movements. Indeed, the maximal distance between knees along the longitudinal axis provides us with very accurate gait events detection compared to previous works. We have validated this detection events on walking patterns that were also altered by placing a 5cm-sole below the left (resp. right) foot of the subject to create asymmetry. The results show that this gait cycle event detection based on depth images is as accurate as using reference methods based on accurate motion capture systems.

## 6.4. VR and Ergonomics

**Participants:** Charles Pontonnier [contact], Georges Dumont, Pierre Plantard, Franck Multon.

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

We proposed in collaboration with Thierry Duval (Lab-Stic, Telecom Bretagne, Brest) a new architecture for information sharing and bridging in collaborative virtual environments in application to ergonomics studies. We particularly presented this year how we implemented the design engineer role in the collaborative environment [30], [29]. We are currently evaluating the complete framework for collaborative ergonomics by defining use-cases and trying to find the best design mode to efficiently solve this problem. Moreover, we have developed and evaluated some manipulation techniques, such as the 7-handle technique which is particularly efficient to manipulate large objects in an immersive environments [6], [27]. A demonstration of this technique has been presented during the ICAT-EGVE conference [37].

We also contributed in the on-site motion analysis field. Microsoft Kinect is a promising tool to evaluate human poses without markers, calibration and manual post-processing. It has been applied to a wide set of applications, such as entertainment, rehabilitation, sports analysis and more recently in ergonomics. In MimeTIC we wish to develop innovative approaches based on a Kinect in order to assess the potential risks of musculoskeletal disorders. However analyzing humans in work places is challenging because of many potential occlusions and displacements of the user. Hence it is a key point to evaluate to which extent this method could be applied to real work places, in real work condition. Most of previous works aiming at evaluating the Kinect sensor generally focus on simple 2D poses. In this work we proposed to evaluate the reliability of Kinect measurements for assessing the movement of operators in ergonomic studies with complex 3D upper-limb poses [38]. To this end we asked subjects to perform complex 3D arm motions concurrently measured with a Kinect and a Vicon motion capture system. The results demonstrated that most of the poses were correctly estimated with the Kinect but specific poses are badly reconstructed, led to errors going up 30°. Hence, experimenter should take this information into account when using a Kinect on a work place in order to avoid experimenting these bad results.

At last, we proposed a new approach for the use of virtual reality with haptics in the Product Development Process loop for testing deformable parts by introducing the user in the loop and proposing a two-stage deformation simulation method for real time haptic interaction. Such an approach is of importance to let the designer be able to handle and validate the design of a product or a workstation respecting multiple constraints, e.g. ergonomics, bulk or productivity. This approach has been fully detailed in a book chapter published this year [34]

## 6.5. Virtual Human Animation

**Participants:** Franck Multon [contact], Julien Pettré, Steve Tonneau.

A common issue in three-dimensional animation is the creation of contacts between a virtual creature and the environment. Contacts allow force exertion, which produces motion. This paper addresses the problem of computing contact configurations allowing to perform motion tasks such as getting up from a sofa, pushing an object or climbing. We propose a two-step method to generate contact configurations suitable for such tasks. The first step is an offline sampling of the range of motion (ROM) of a virtual creature. The ROM of the human arms and legs is precisely determined experimentally. The second step is a run time request confronting the samples with the current environment. The best contact configurations are then selected according to a heuristic for task efficiency. The heuristic is inspired by the force transmission ratio. Given a contact configuration, it measures the potential force that can be exerted in a given direction. The contact configurations are then used as inputs for an inverse kinematics solver that will compute the final animation. Our method is automatic and does not require examples or motion capture data. It is suitable for real time applications and applies to arbitrary creatures in arbitrary environments. Various scenarios (such as climbing, crawling, getting up, pushing or pulling objects) are used to demonstrate that our method enhances motion autonomy and interactivity in constrained environments [15], [32]. In Figure 6, a character is able to select the most appropriate constraints to pull a heavy cupboard by putting a foot on an obstacle to maximize the force ratio.

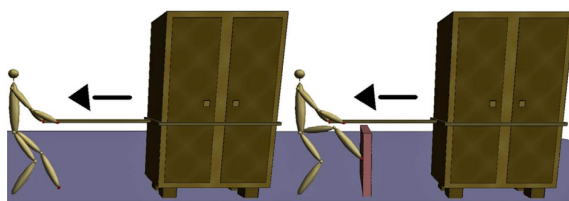


Figure 6. First-person view condition in the basket free-throw.

## 6.6. Biomechanics for avatar animation

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

Bio-inspired controllers and planners are compelling for avatar animation. We are currently engaging several works on the subject within the frame of the ENTRACTE project.

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. A first result has been obtained in defining and validating a bio-inspired limb controller based on a linearizing loop of a neuromuscular complex. Application on a one-dof limb has been validated by comparing the muscle activation shapes obtained in simulation with standard records of biceps and triceps activation [3].



## 6.7. Semantically consistent hierarchical decomposition of virtual urban environments

**Participants:** Carl-Johan Jorgensen, Fabrice Lamarche [contact].

When planning a path in their environment, pedestrians do not consider every detail at once. Instead, people first plan a coarse path, choosing streets to travel to reach their goal. Local decisions such as where to cross a street or on which side to pass by a pole are taken during navigation. In computer science, hierarchical representations of an environment are often used to reduce the computation cost of the planning algorithm. Such a representation also enables smarter navigation behaviours. Indeed, it offers the opportunity to delay the local planning until relevant information is available. It also enables a quick recovery from unexpected events, as the high-level path might stay valid even if unexpected events alter the lower-level path.

We proposed a method that automatically generates a three-level hierarchical representation of an informed urban environment. In this hierarchy, each level is a semantically coherent partition of the navigation areas and can be used to plan paths at different levels of abstraction. This representation is used in a path planning process that delays some decisions until relevant information is perceived. This algorithm uses path options to smartly adapt the path when unexpected events occur.

## 6.8. Shoulder biomechanics

**Participants:** Armel Crétual [contact], Michaël Ropars.

At first sight, in clinical practice, shoulder mobility is frequently evaluated through mono-axial amplitude. Interestingly, for diagnosing shoulder hyperlaxity or frozen shoulder, external rotation of the arm whilst at the side (ER1) is commonly used. We first gave a definition of hyperlaxity, as described actually in the literature, and its link with shoulder instability and treatment. After looking for an optimized way to examine external rotation of the shoulder, we proposed the definition of a novel index to quantify global shoulder mobility, the Shoulder Configuration Space Volume (SCSV) corresponding to the reachable volume in the configuration space of the shoulder joint [4]. Then, this new index was examined through correlation to shoulder signs of hyperlaxity [14].

## MINT Project-Team

# 6. New Results

## 6.1. Highlights of the Year

- “Adoiraccourcix : sélection de commandes sur écrans tactiles multi-points par identification des doigts” [31] received the *best paper award* from the IHM 2014 conference;
- “L’ordinateur portable comme instrument de musique” [41] received the *best demo award* from the IHM 2014 conference.

## 6.2. Impact of form factors and input conditions on absolute indirect-touch pointing tasks

Absolute indirect interaction maps the absolute position of a device’s end-effector to the absolute position of a remote on screen object. Despite its long-time use with graphics tablets and growing use in research prototypes, little is known on the influence of form factors and input conditions on pointing performance with such a mapping. The input and display can have different sizes and aspect ratios, for example. The on-screen targets can vary in size. Users can look solely at the display or at the input device as well. They can also hold the input device in certain cases, or let it rest on a table. We ran two experiments designed to investigate the influence of all these factors on absolute indirect-touch pointing performance [20], [11].

The first experiment focused on input device size and input conditions and revealed that users get higher performance when they can look at the input surface (even if nothing is displayed on it). In addition we found that the smallest target size users can acquire in motor space is not constant across different input dimensions but degrades as the input size increases. The second experiment focused on scale effects and aspect ratio and revealed users’ performance is not affected by scale but that aspect ratio matters: similar input and output aspect ratios lead to better performance. This findings led us to list four main recommendations for the design of touch input surfaces with applications supporting absolute direct interaction.

## 6.3. Direct and indirect multi-touch interaction on a wall display

Multi-touch wall displays allow to take advantage of co-located interaction (direct interaction) on very large surfaces. However interacting with content beyond arms’ reach requires body movements, introducing fatigue and impacting performance. Interacting with distant content using a pointer can alleviate these problems but introduces legibility issues and loses the benefits of multi-touch interaction. We introduced WallPad [30], [11], a widget designed to quickly access remote content on wall displays while addressing legibility issues and supporting direct multi-touch interaction (Figure 1 ). To support multi-touch on such a wall display, we developed a custom system using front diffuse illumination and 4 cameras. Our system can detect 50+ simultaneous contacts with a precision between 3 and 5 mm.

## 6.4. Sketching dynamic and interactive illustrations

We collaborated with Autodesk Research in Toronto (as a scientific consultant) on a project whose focus was to design and develop tools that enable artists to bring life to illustrations with subtle, continuous animation effects and infusing interactive behavior to the drawings. We believe designers, artists and creators should be able to communicate with computers the way they think about art and animation. This motivated Autodesk to develop interfaces that facilitate powerful ways of thinking and content creation with freeform sketching and direct manipulation, thus offering an alternative to complex professional animation tools. Our design combines the complementary affordances of humans and computers by utilizing by-example phenomena, thus preserving expressiveness and personal style, yet reducing tedium.

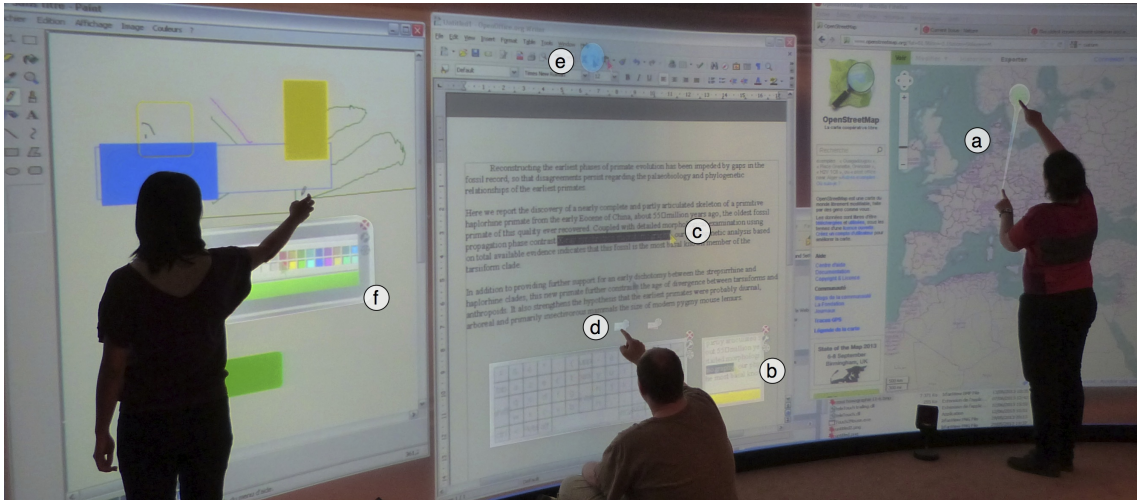


Figure 1. A user creating a WallPad (a). Another using two WallPads to select (b) some text (c) and click (d) on Copy/Paste buttons (e) beyond reach. A third user accessing a color palette located at the bottom of the screen through a WallPad (f).

The outcome of the collaboration is Kitty [23], a sketch-based tool for authoring dynamic and interactive illustrations (Figure 2). Artists can sketch animated drawings and textures to convey the living phenomena, and specify the functional relationship between its entities to characterize the dynamic behavior of systems and environments. An underlying graph model, customizable through sketching, captures the functional relationships between the visual, spatial, temporal or quantitative parameters of its entities. As the viewer interacts with the resulting dynamic interactive illustration, the parameters of the drawing change accordingly, depicting the dynamics and chain of causal effects within a scene. The generality of this framework makes our tool applicable for a variety of purposes, including technical illustrations, scientific explanation, infographics, medical illustrations, children’s e-books, cartoon strips and beyond. A user study demonstrates the ease of usage, variety of applications, artistic expressiveness and creative possibilities of our tool.

Kitty is a follow up of a previous project, Draco [50], a prototype sketch-based interface that allows artists and casual users alike to add a rich set of animation effects to their drawing, seemingly bringing illustrations to life such as a school of fish swimming, tree leaves blowing in the wind, or water rippling in a pond. Draco was realized before Fanny Chevalier joined Inria. Kitty is the result of a collaboration between Autodesk Research (inventor) and Inria (scientific consultant). A patent has been filed by Autodesk Research for Kitty, and the company is currently developing a commercial application based on the research prototype.

## 6.5. The not-so-staggering effect of staggered animations

Interactive visual applications often rely on animation to transition from one display state to another. There are multiple animation techniques to choose from, and it is not always clear which should produce the best visual correspondences between display elements. One major factor is whether the animation relies on staggering—an incremental delay in start times across the moving elements. It has been suggested that staggering may reduce occlusion, while also reducing display complexity and producing less overwhelming animations, though no empirical evidence has demonstrated these advantages. We empirically evaluated the effect of two staggering techniques on tracking tasks, focusing on cases that should most favour staggering [14]. We found that introducing staggering has a negligible, or even negative, impact on multiple object tracking performance. The potential benefits of staggering may be outweighed by strong costs: a loss of common-motion grouping

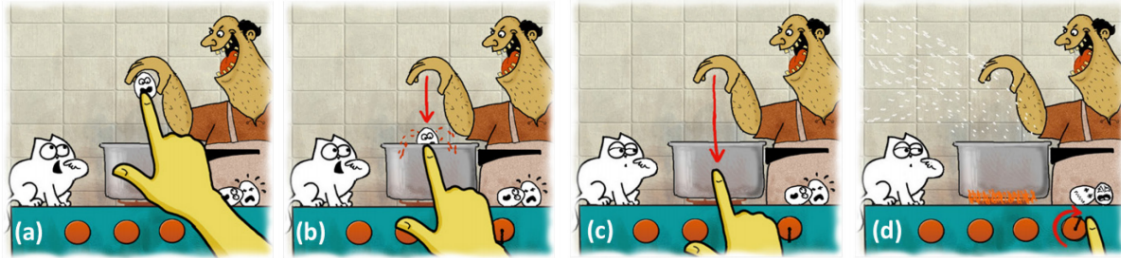


Figure 2. Example of a dynamic interactive illustration authored with Kitty. (a) Objects in the scene are interactive: the egg held by the cook can be dragged down, as if falling into the pot, triggering subsequent animations, such as soup splashes (b) and closing of the cat's eyelids (c). Turning the knob increases the fire and steam (d). The resulting dynamic illustration captures the living nature of the scene, where the gas stove flames burn and steam emits from the pot.

information about which objects travel in similar paths, and less predictability about when any specific object would begin to move.

## 6.6. Flexible contextual retrieval of chosen documents and windows

Users of Personal Computers interact with a large number of resources to do their work. To handle their different tasks, they need their documents to be readily available, and as the number of activities and documents increase, systems must offer adequate support for quick retrieval of these resources. The Hotkey Palette [29] is a quick retrieval facility that we designed that uses hotkeys and makes them visible and configurable through a quasi-modal always-available on-screen keyboard. This facility contributes to the state of the art in three ways. It extends on-screen keyboard interaction by providing feedback on the state of the linked resources, it provides persistent and integrated access to local windows and files and other online resources, and it provides flexible control over contextualization by leveraging existing resource hierarchies.

## 6.7. Multi-touch command selection using finger identification

Hotkeys are a critical factor of performance for expert users in WIMP interfaces. Multi-touch interfaces, by contrast, do not provide such efficient command shortcuts. Adoiraccourcix leverages finger identification to introduce quick command invocation integrated with direct manipulation in this context (Figure 3 ). We illustrated its use in a vectorial drawing application and ran preliminary user studies comparing it to classical user interfaces. Results suggest that once mastered, it provides very powerful means of interaction [31], [44], [43].



Figure 3. Partial illustration of the Adoiraccourcix' logics.

## 6.8. Impact of the localization and activation of mode switchers

Input devices have a limited number of buttons and degrees of freedom, but they are used to control many functionalities. Modes and quasi-modes makes it possible to map several actions to the same input. For example keys of a keyboard either input a letter or trigger a command. Delimiters allow users to switch between the modes. On the keyboard, the default mode is often text entry and pressing the Ctrl key switches the mode to command mode. This choice was made at a time when the mouse was not widespread. In [33], we explored the possibility to place mode switchers on the mouse and experimented the benefits. We showed that there is a performance benefit if the current tasks are essentially mouse-based. In particular we showed that using mode switchers on the mouse reduces homing the dominant hand between the mouse and the keyboard.

## 6.9. A serial Architecture for a collaborative robot

The haptic magnifier consists in using a serial architecture, where a motor is inserted between a tool and a user's hand (figure 4). By this way, the tool's speed  $v_o$  can be changed relatively to user's speed  $v_i$ , by controlling motor's speed. The haptic rendering of a load can then be changed, and fine details can be more easily detected.

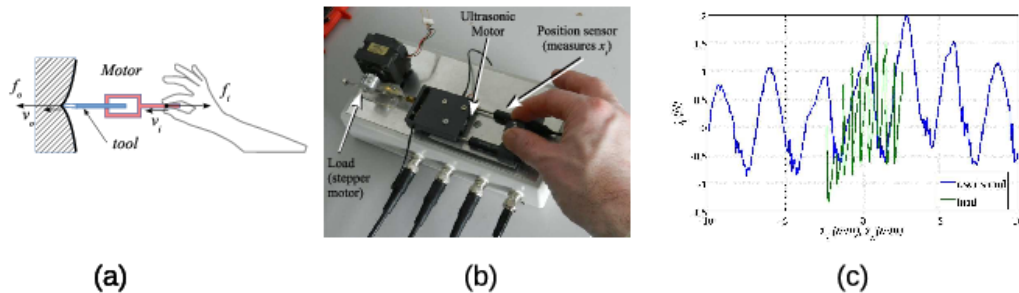


Figure 4. The Haptic Magnifier; (a) the serial architecture with a motor inserted to achieve a haptic magnifier; (b) the implementation with an ultrasonic Motor; and (c) the resulting rendering at load's end and user's end.

The haptic magnifier is built up with an ultrasonic motor, whose characteristic is low speed - high torque. So the tool and the end-effector can be directly connected to the motor, leading to a lightweight architecture. The user's study presented in [21] have shown that the precision in using the tool could be improved during a freehand manipulation.

## 6.10. Mimetic Interaction Spaces: Controlling Distant Displays in Pervasive Environments

Pervasive computing is a vision that has been an inspiring long-term target for many years now. Interaction techniques that allow one user to efficiently control many screens, or that allow several users to collaborate on one distant screen, are still hot topics, and are often considered as two different questions. Standard approaches require a strong coupling between the physical location of input device, and users. We propose to consider these two questions through the same basic concept, that uncouples physical location and user input, using a mid-air approach. We present the concept of mimetic interaction spaces (MIS), a dynamic user-definition of an imaginary input space thanks to an iconic gesture, that can be used to define mid-air interaction techniques. We describe a participative design user-study, that shows this technique has interesting acceptability and elicit some definition and deletion gestures. We finally describe a design space for MIS-based interaction, and show how such concept may be used for multi-screen control, as well as screen sharing in pervasive environments [26].



### 6.11. Match-Up & Conquer: A Two-Step Technique for Recognizing Unconstrained Bimanual and Multi-Finger Touch Input

We present a simple, two-step technique for recognizing multi-touch gesture input independently of how users articulate gestures, i.e., using one or two hands, one or multiple fingers, synchronous or asynchronous stroke input. To this end, and for the first time in the gesture literature, we introduce a preprocessing step specifically for multi-touch gestures (Match-Up) that clusters together similar strokes produced by different fingers, before running a gesture recognizer (Conquer). We report gains in recognition accuracy of up to 10% leveraged by our new preprocessing step, which manages to construct a more adequate representation for multi-touch gestures in terms of key strokes. It is our hope that the Match-Up technique will add to the practitioners toolkit of gesture preprocessing techniques, as a first step toward filling today's lack of algorithmic knowledge to process multi-touch input and leading toward the design of more efficient and accurate recognizers for touch surfaces. [27]

### 6.12. Understanding Users's perceived Difficulty of Multi-Touch Gesture Articulation

We show that users are consistent in their assessments of the articulation difficulty of multi-touch gestures, even under the many degrees of freedom afforded by multi-touch input, such as (1) various number of fingers touching the surface, (2) various number of strokes that structure the gesture shape, and (3) single-handed and bimanual input. To understand more about perceived difficulty, we characterize gesture articulations captured under these conditions with geometric and kinematic descriptors computed on a dataset of 7,200 samples of 30 distinct gesture types collected from 18 participants. We correlate the values of the objective descriptors with users' subjective assessments of articulation difficulty and report path length, production time, and gesture size as the highest correlators (max Pearson's  $r=0.95$ ). We also report new findings about multi-touch gesture input, e.g., gestures produced with more fingers are larger in size and take more time to produce than single-touch gestures; bimanual articulations are not only faster than single-handed input, but they are also longer in path length, present more strokes, and result in gesture shapes that are deformed horizontally by 35% in average. We use our findings to outline a number of 14 guidelines to assist multi-touch gesture set design, recognizer development, and inform gesture-to-function mappings through the prism of the user-perceived difficulty of gesture articulation.[28]

### 6.13. Dynamic Modelling of Electrostatic Attraction

Electrostatic attraction may be used to modulate the apparent friction coefficient between a fingertip and a surface to create a tactile stimulator. In this work, we want to propose an accurate modelling of the force generation. For that purpose, a specific experimental test bench has been manufactured, as presented in figure 5

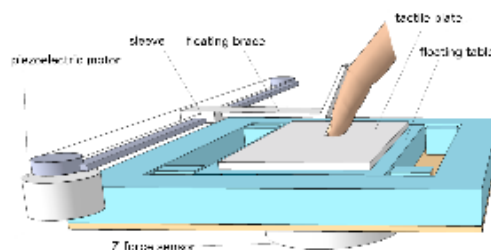


Figure 5. Representation of the measurement system, the finger is moved on the plate by the motor.



Then, an investigation on the current modeling were carried out, with a focus on the temporal evolution and frequency dependence of the stimulus. More particularly, we considered the charge lost through the stratum corneum. Indeed, lost charges is gathered on the surface of the insulator as free surface charge, for this reason it no longer participates to the generation of the force on the finger, and consequently, to the measured force (Fig. 6). This happens because the charges on the surface of the insulator are no longer mechanically bounded to the finger and the insulator sustains the induced electrostatic force.

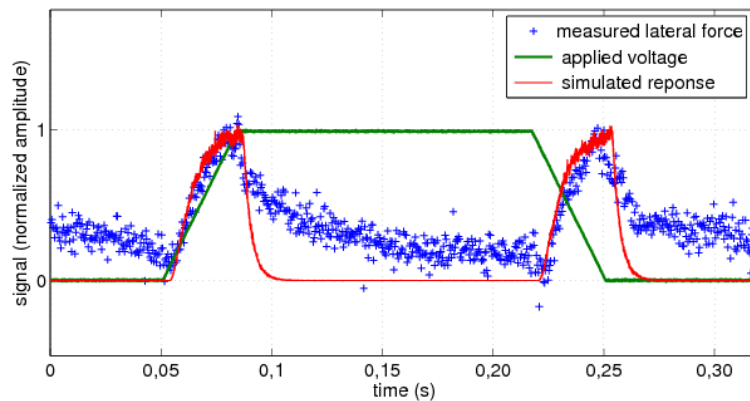


Figure 6. Charge configuration at the border of the stratum corneum (SC) and insulator (I). The conductive part of the system is represented like the electrode of a capacitor. (1) Initial configuration on the charge when the voltage  $v$  is applied. (2) Discharge through the stratum corneum with the two equivalent capacitors. (3) Final configuration of the charges after the transient.

The improvement of the modeling is proposed to take into account this major effect, and then, it is checked with an experimental set-up and compared with literature results.

## 6.14. Coupling between Electro vibration and squeeze film for tactile stimulation

Electro vibration and squeeze film effect are two different principles which modify user perception of a surface. The first is generated by a polarization of a finger approaching a high voltage supplied plate, and the latter by an ultrasonic vibrating plate. Their compatibility on the same stimulator has been analysed and their concomitant has been proven as well as the increased range of sensations [34]. A joint model has been proposed to describe the behaviour of the friction when both principles are merged. For the analysis, a specific experimental test bench has been built to measure the forces induced, as shown in figure 7.

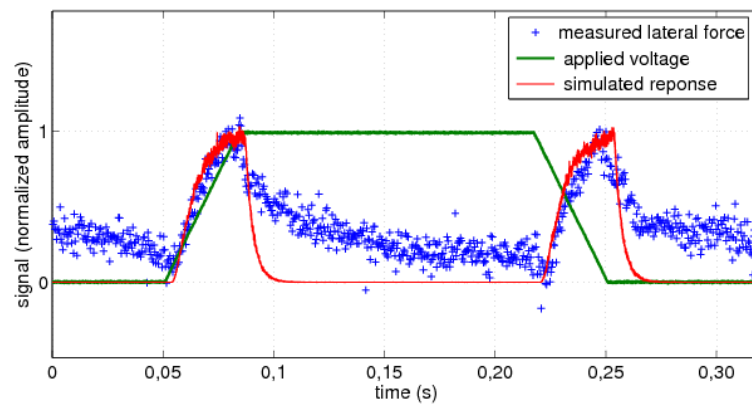


Figure 7. The experimental setup, and the recorded friction modulation.

## POTIOC Project-Team

# 6. New Results

## 6.1. Highlights of the Year

- Acceptance of the ANR project "ISAR" (Interacting with Spatial Augmented Reality) lead by Martin Hachet (Potioc)
- Publication of "Teegi" (Tangible EEG Interface) at UIST14 [15] and more than 13000 views on vimeo until December 2014 (<http://vimeo.com/potioc/teegi>)

## 6.2. Teegi -Tangible EEG Interface- and MindMirror for interactive visualization of brain activities

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

Typical brain activity visualization tools are usually hard to understand and interpret for novice users. With advances in neurotechnologies (notably BCI) and HCI/AR, we explored the design of new ways to visualize our own brain activity in real-time, for which we proposed two new systems.

We designed Teegi, a Tangible EEG Interface that enables novice users to get to know more about something as complex as brain signals, in an easy, engaging and informative way [15]. To this end, we have designed a new system based on a unique combination of spatial augmented reality, tangible interaction and real-time neurotechnologies (see Figure 5 ). With Teegi, a user can visualize and analyze his or her own brain activity in real-time, on a tangible character that can be easily manipulated, and with which it is possible to interact. Users can also reveal some specific EEG phenomena (e.g., sensorimotor rhythms) still using a tangible approach by placing dedicated "mini-teegi" (small pupets) in a designated area on the interaction zone. The whole system has been designed with educational psychology tools in mind to ensure an efficient learning. An explorative study has shown that interacting with Teegi seems to be easy, motivating, reliable and informative. Overall, this suggests that Teegi is a promising and relevant training and mediation tool for the general public.

In addition, together with colleagues from Inria Rennes (team Hybrid), we introduced a novel augmented reality paradigm called "the Mind-Mirror" which enables such an experience of seeing "through your own head", visualizing your brain "in action and in-situ" [23]. Our approach relies on the use of a semi-transparent mirror positioned in front of a computer screen. A virtual brain is displayed on screen and automatically follows the head movements thanks to an optical face-tracking system. The brain activity is extracted and processed in real-time thanks to an EEG cap wore by the user. A rear view is also proposed thanks to an additional web-cam recording the rear of user's head (see Figure 6 ).

## 6.3. Interaction in mobile augmented reality

**Participants:** Asier Marzo, Benoît Bossavit, Martin Hachet.

Nowadays, handheld devices are capable of displaying augmented environments in which virtual content overlaps reality. To interact with these environments it is necessary to use a manipulation technique. The objective of a manipulation technique is to define how the input data modify the properties of the virtual objects. Current devices have multi-touch screens that can serve as input. Additionally, the position and rotation of the device can also be used as input creating both an opportunity and a design challenge. In this project we compared three manipulation techniques which namely employ multi-touch, device position and a combination of both. A user evaluation on a docking task revealed that combining multi- touch and device movement yields the best task completion time and efficiency. Nevertheless, using only the device movement and orientation is more intuitive and performs worse only in large rotations. This work has been presented at the ACM Symposium on Spatial User Interaction 2014 [21].

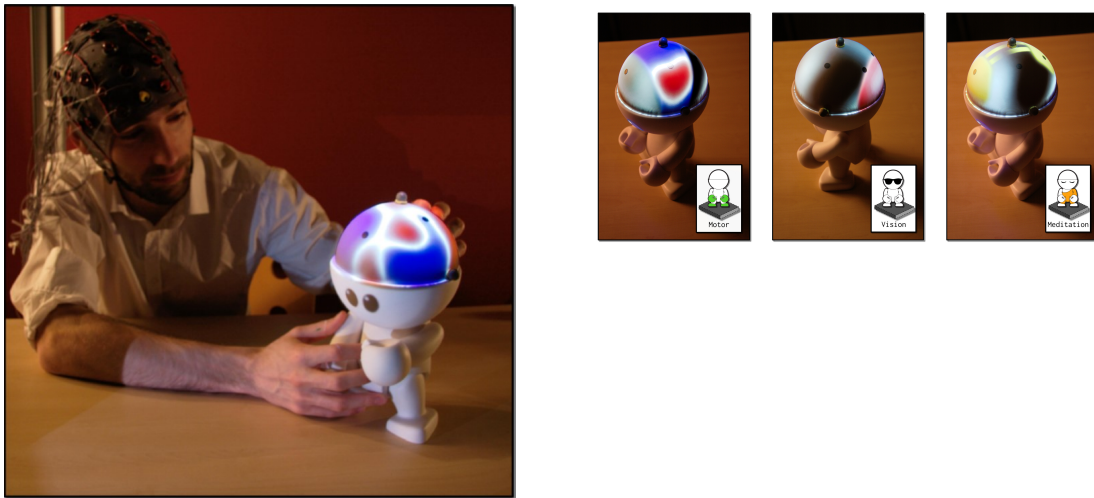


Figure 5. Teegi : a Tangible EEG Interface based on augmented reality.



Figure 6. The Mind Mirror, a new real-time visualization tool of the user's own brain activity based on augmented reality.



Figure 7. Touch (left), tilte (middle) and AR interaction techniques.

In this project we have furthermore evaluated controls based on Augmented Reality (AR), Tilt and Touch for a Point and Shoot Mobile Game (see Figure 7). A user study (n=12) was conducted to compare the three controls in terms of player experience and accuracy. Tilt and AR controls provided more enjoyment, immersion and accuracy to the players than Touch. Nonetheless, touch caused fewer nuisances and was playable under more varied situations. Despite the current technical limitations, we suggest to incorporate AR controls into the mobile games that supported them. Nowadays, AR controls can be implemented on handheld devices as easily as the more established Tilt and Touch controls. However, this study is the first comparison of them and thus its findings could be of interest for game developers. This work has been presented at ISMAR - MASH'D [22].

## 6.4. CurSAR: Interacting with Spatial Augmented Reality with 2D Input Devices

**Participants:** Renaud Gervais, Jérémy Frey, Martin Hachet.

Spatial Augmented Reality (SAR) opens interesting perspectives for new generations of mixed reality applications. Compared to traditional HCI contexts, there is little work that studies user performance in SAR. We did an experiment that compared pointing in SAR versus pointing in front of a screen using standard pointing devices (mouse and tablet). The results showed that the users tend to interact in SAR in a way that is similar to the screen condition, without a big loss of performance.

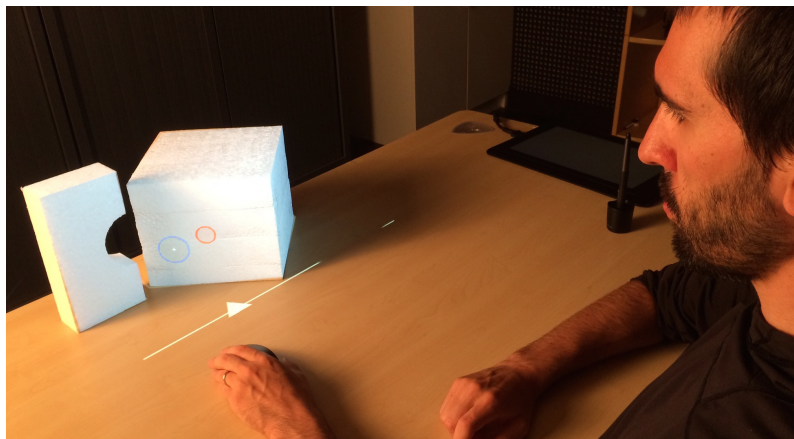


Figure 8. Pointing in spatial augmented reality

## 6.5. Creative Coding on Objects

**Participants:** Renaud Gervais, Jérémy Laviolle, Asier Marzo, Martin Hachet.

In a near future scenario, we will replace some of our everyday objects with counterparts in form of Computational Objects (COs). COs look similar to the original object; however, inside them there are input sensors, output devices such as displays and a CPU. Furthermore, COs still convey the context and meaning that the original object had. For instance, a clock is associated with time and thus users could expect its CO version to display time-related data. We suggest that any user should be able to easily code new appearances and behaviors for his or her own objects. Using creative coding as a base, we propose to add the notions of affordances and conventions to this programming context. Moreover, we suggest that COs could be used as a creativity support tool although modifying their behavior beyond conventions could confuse the user. Finally, we reckon that with the proper tools, users could also make physical modifications to COs. For example, a retractile cord can be attached to the clock and be used to pull data out and display them in a linear layout.

This work has been presented as a poster at TEI 2014 [30].

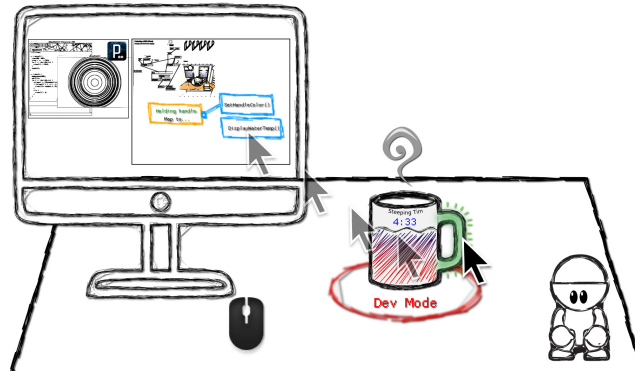


Figure 9. Sketch for a creative coding on objects scenario

## 6.6. Physiological sensors: bridging human-computer interaction

**Participants:** Jérémy Frey, Dennis Wobrock, Aurélien Appriou, Christian Mühl, Fabien Lotte, Martin Hachet.

Physiological sensors are not limited to research and medical facilities anymore. They are getting more and more affordable and they become widely accessible to users, as denoted by the popularity of smartphone apps and wearables that track heart rate during fitness activities. Before long, we may see a wide range of sensors embedded into consumer electronic devices. This trend has already started with the arrival of "smartwatches" that could – among other things – detect users' heart beats covertly.

We anticipated this opportunity in order to increase engagement in human-computer interaction, more specifically in human-agent interaction. In [14] we demonstrated that we could increase the social presence of embodied agents – that is, of virtual beings – by simply mirroring the heart rate of users. The "similarity-attraction" effect induces positive emotions toward persons or things that look like us or react as we do. An agent that is associated to a heart beat at the same pace as the user is found more sympathetic. The "similarity-attraction" effect, applied to physiological computing, could help with little effort to improve the acceptance of embodied agents and robots by the general public. (See Figure 10 for setup).

Furthermore, we have taken advantage of physiological sensors in order to evaluate different sorts of human-computer interfaces prior to their release. First, we showed that we could reliably estimate the user's mental workload levels from his/her EEG signals, across different contexts involving different levels of social stress [10]. Then, based on those results, we used a combination of electrocardiography (measure of heart beats), galvanic skin response (measure of sweat on the skin) and electroencephalography (EEG, measure of brain activity) to assess the workload of users during 3D manipulation tasks. The first preliminary results seem to indicate that we might be able to discriminate the parts of the interaction that provokes a high cognitive stress, hence that needs to be improved. This work is in line with the evaluation of visual comfort. We presented earlier this year a pilot study documenting how different virtual depths could cause different levels of discomfort [17], and how this discomfort translates to EEG activity.

Pervasive technologies and physiological computing may be a key component to bridge the gap that too often keeps dividing machines and general public. We believe that it'll help make computers more enjoyable and more usable.





Figure 10. Experimental setup used to study the physiological "similarity-attraction" effect in human-agent interaction.

## 6.7. Training Approaches for Brain-Computer Interfaces

**Participants:** Alison Cellard, Martin Hachet, Camille Jeunet, Fabien Lotte, Christian Mühl, Julia Schumacher.

While recent research on Brain-Computer Interfaces (BCI) has highlighted their potential for many applications, they remain barely used outside laboratories due to a lack of robustness. Spontaneous BCI (i.e., mental imagery-based BCI) often rely on mutual learning efforts by the user and the machine, with BCI users learning to produce stable EEG patterns (spontaneous BCI control being widely acknowledged as a skill) while the computer learns to automatically recognize these EEG patterns, using signal processing. Most research so far was focused on signal processing, mostly neglecting the humans in the loop.

Indeed, even if it has been advocated in one of our previous publications (see activity report 2013) that current human training approaches for spontaneous BCI are most likely inappropriate, based on theoretical models, we still needed practical confirmations that users' modest performances at controlling a BCI could be partly due to these inappropriate training protocols. Thus, in our work, we proposed to study standard BCI training protocols without EEG signals, i.e., without a BCI [31]. In particular, we studied how people could learn to do two simple motor tasks using the same training tasks and feedback as the one given to motor imagery BCI users. More precisely, we asked subjects to learn to draw on a graphic tablet a triangle and a circle (the correct size, angles and speed of drawing of these two shapes being unknown to the subject) that can be recognized by the system, using a synchronous training protocol and an extending bar as feedback, like for motor imagery based BCI training. Our results show that most subjects (out of  $N=20$  subjects) improved with this feedback and practice (i.e., the shapes they drew were increasingly more accurately recognized by the system), but that 15% of them completely failed to learn how to draw the correct shapes, despite the simplicity of the motor tasks. This suggests that part of BCI illiteracy/deficiency is likely due to the training protocols currently used.

From the huge variability in users' performances at BCI mastery emerged the following question: Why do some people manage to learn using these protocols and others do not? Our hypothesis here was that these protocols are not adapted to some users' profiles. Thus, we designed an experiment in which we looked for correlations between the personality and cognitive profile of the users and their ability to learn to control a MI-

BCI. Our current results (N=18) show that 1) performances are strongly correlated with users' spatial abilities and 2) we can reliably predict these performances using a model including different psychological factors (like abstractedness, self-reliance or tension). These results are very encouraging as they could lead to reflections about 1) exercises to improve users' spatial abilities and 2) solutions to take into account users' cognitive and personality profiles in BCI training approaches.

Furthermore, it is more and more claimed that visual feedback is not ideal for BCIs as they are conceived for interaction situations in which the visual channel is often overtaxed. Thus, tactile feedback might appear to be more relevant. In order to test this hypothesis, we proposed a study aiming at comparing a standard visual feedback with an equivalent tactile feedback in an appealing training environment containing visual distractors (to mimic an interaction context in which the visual channel is overtaxed). Users had to learn to perform motor-imagery tasks as well as a counting task, and received either a visual or vibrotactile feedback (see Figure 11). Our main result (N=18) is the fact that people receiving tactile feedback perform significantly better (at Motor-Imagery and counting task). This kind of result should encourage the BCI community to replace standard BCI protocols by more motivating training environments and multimodal feedback.

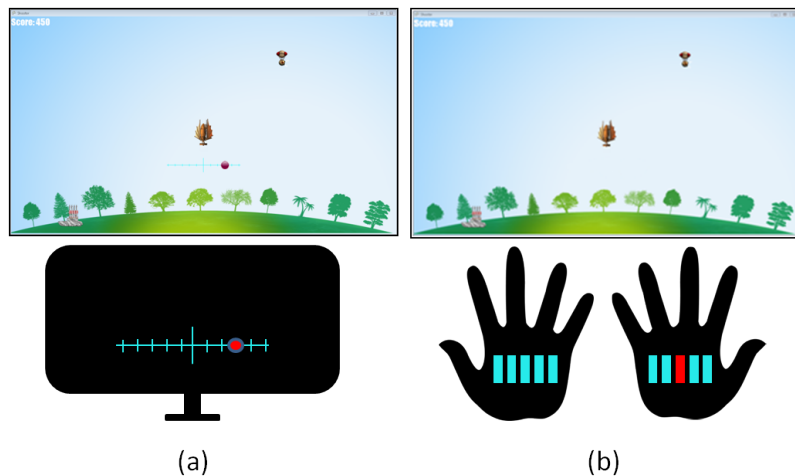


Figure 11. Illustration of the combination of an appealing training environment and a vibrotactile feedback (right, compared to a standard visual feedback, left)

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.

## 6.8. EEG signal processing

**Participants:** Alison Cellard, Nicoletta Caramia, Fabien Lotte.

Spatial filters are powerful tools for EEG classification for BCI design, able to reduce spatial blurring effects. In particular, optimal spatial filters have been designed to classify EEG signals based on band power features. Unfortunately, there are other relevant EEG features for which no optimal spatial filter exists. This is the case for Phase Locking Value (PLV) features, which measure the synchronization between 2 EEG channels.

Therefore, we proposed to create such a pair of optimal spatial filters for PLV-features [13]. To do so, we optimized a functional measuring the discriminability of PLV-features based on a genetic algorithm. An evaluation of our algorithm on a motor imagery EEG data set showed that using optimized spatial filters led to higher classification performances, and that combining the resulting PLV features with traditional methods boosts the overall BCI performances.

We also wrote a chapter that is an introductory overview and a tutorial of signal processing techniques that can be used to recognize mental states from EEG signals in BCI [26]. More particularly, this chapter presented how to extract relevant and robust spectral, spatial and temporal information from noisy EEG signals (e.g., Band Power features, spatial filters such as Common Spatial Patterns or xDAWN, etc.), as well as a few classification algorithms (e.g., Linear Discriminant Analysis) used to classify this information into a class of mental state. It also briefly touched on alternative, but currently less used approaches.

## 6.9. Navigation techniques in 3D digital cities on mobile touch devices

**Participants:** Jacek Jankowski, Thomas Hulin, Martin Hachet.

This project was part of "Villes transparentes" research project in collaboration with Mappy (Solocal group) and Vectuel - VirtuelCity initiated in 2013. It aimed at characterizing today's most common interaction techniques for street-level navigation in 3D digital cities, for mobile touch devices, in terms of their efficiency and usability. To do so, we conducted a user study, where we compared target selection (Go-To), rate control (Joystick), position control, and stroke-based control navigation metaphors (see Figure 12). The results suggest that users performed best with the Go-To interaction technique. The subjective comments showed a preference of novices towards Go-To and expert users towards the Joystick technique. This work has been published at the 3DUI 2014 conference [18].

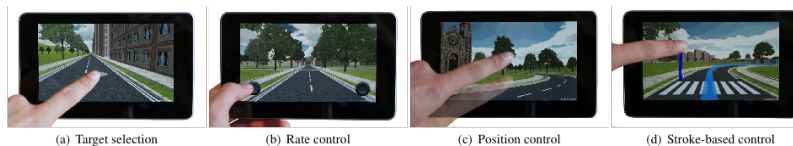


Figure 12. Four techniques for navigating in a 3D city on a mobile touch device.)

## REVES Project-Team

# 6. New Results

## 6.1. Highlights of the Year

Our work on sketch-based modeling for product designers (Sec. 6.4.4 ) has received significant attention. It appeared on the news page of University of British Columbia <http://news.ubc.ca/2014/08/13/powerful-math-creates-3-d-shapes-from-simple-sketches/> and our video has been watched more than 7000 times on Youtube <http://youtu.be/tbUlJHJv4Rg>. We filed a patent on this technology and we have contacts with several companies about a potential transfer.

Our poster on *C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics* [ ] received the 3rd place in the ACM's Graduate Student Research Competition at SIGGRAPH 2014. This work is a collaboration with George Alex Koulieris and Katerina Mania from the Technical University of Crete and Douglas Cunningham from the Technical University of Cottbus.

BEST PAPER AWARD :

[ ] **Computer Graphics Forum**. G. A. KOULIERIS, G. DRETTAKIS, D. W. CUNNINGHAM, K. MANIA.

## 6.2. Plausible and Realistic Image Rendering

### 6.2.1. 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. 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 labeling, allowing us to compute high quality reflectance and shading layers. We then create shadow-caster geometry that preserves shadow silhouettes. Combined with the intrinsic layers, this approach allows 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 is part of an industrial partnership with Autodesk and is under revision for ACM Transactions On Graphics.

### 6.2.2. Compiler and Tiling Strategies for IIR Filters

**Participants:** Gaurav Chaurasia, George Drettakis.

We present a compiler for parallelizing IIR or recursive filters. IIR filters are frequently used for  $O(1)$  convolutions, but they cannot exploit GPUs because they are very hard to parallelize and also exhibit poor memory locality which hinders performance on both CPUs and GPUs. We present algorithmic tiling strategies for IIR filters which overcome these limitations. Tiled IIR filters are notoriously hard to implement and hence largely ignored by programmers and hardware vendors. We present a compiler front-end that supports intuitive functional specification and tiling of IIR filters. We demonstrate that different tiling strategies may be optimal on different platforms and filter parameters; our compiler can express the exhaustive set of alternatives in just 10-20 lines of code. This enables programmers to easily explore a large variety of trade-offs at different levels of granularity, thereby making it easier and more likely to discover the optimal implementation, while also producing intuitive and maintainable code. Our initial results show that our compiler is as terse as vendor provided libraries, but it allows exploiting the algorithmic advantages of tiling which cannot be provided by any precompiled library.

For example, our compiler can compute a nearly 8 times faster summed area table ( $4096 \times 4096$  image) in 20 lines of code including a fully customized CUDA schedule, as compared to 10 lines in NVIDIA Thrust which does not allow tiling or customizing the CUDA schedule.

This ongoing work is a collaboration with Jonathan Ragan-Kelley (Stanford University), Sylvain Paris (Adobe) and Fredo Durand (MIT).

### 6.2.3. Video based rendering

**Participants:** Abdelaziz Djelouah, George Drettakis.

In this project our objective is to propose a new algorithm for novel view synthesis in the case of dynamic scene. The main difference compared to static image-based rendering is the limited number of viewpoints and the presence of the extra time dimension. In a configuration where the number of cameras is limited, segmentation becomes crucial to identify moving foreground regions. To facilitate the difficult task of multi-view segmentation, we currently target scenes captured with stereo cameras. Stereo pairs provide important information on the geometry of the scene while simplifying the segmentation problem.

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

### 6.2.4. Temporally Coherent Video De-Anaglyph

**Participants:** Joan Sol Roo, Christian Richardt.

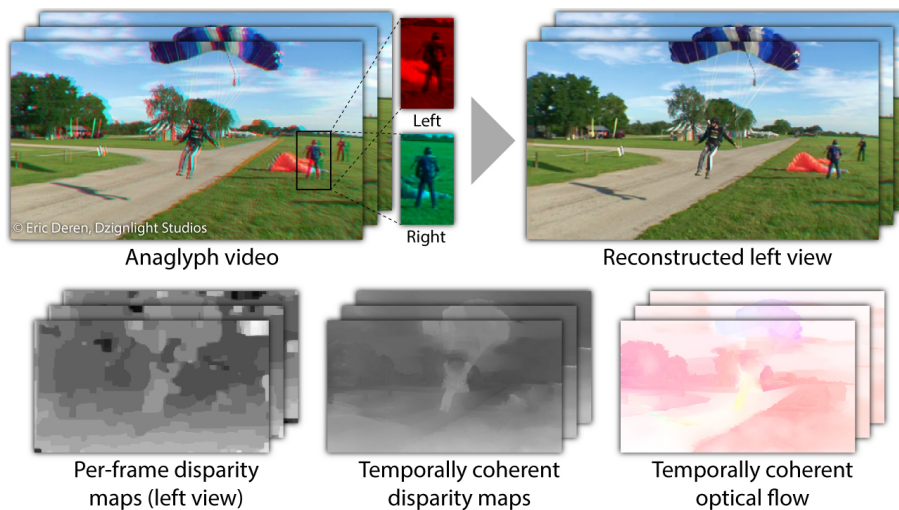


Figure 4. Top: We convert anaglyph videos (left) to temporally coherent full-color stereo videos (right). Bottom: Our approach starts with rough, per-frame disparity maps (left) and produces temporally coherent disparity maps and optical flow (center and right) that are used for reconstructing the stereo views.

This work investigates how to convert existing anaglyph videos to the full-color stereo format used by modern displays. Anaglyph videos only contain half the color information compared to the full-color videos, and the missing color channels need to be reconstructed from the existing ones in a plausible and temporally coherent fashion. In our approach, we put the temporal coherence of the stereo video results front and center (see Figure 4). As a result, our approach is both efficient and temporally coherent. In addition, it computes temporally coherent optical flow and disparity maps that can be used for various post-processing tasks. As a practical contribution, we also make the source code of our implementation available online under CeCILL-B license.



This work was carried out by Joan Sol Roo during his internship in the summer of 2013. The work was presented as a talk and poster at SIGGRAPH 2014 [19].

### 6.2.5. Probabilistic Connection Path Tracing

**Participants:** George Drettakis, Stefan Popov.

Bi-directional path tracing (BDPT) with Multiple Importance Sampling (MIS) 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 project, is that we can find better connections to an eye sub-path by considering multiple light sub-paths at once and creating connections probabilistically only with the most promising ones. We do this by storing light paths, and estimating probability density functions (PDF) of the discrete set of possible connections to all light paths. This has two key advantages: we efficiently create connections with high quality contributions by Monte Carlo sampling, and we reuse light paths across different eye paths. We also introduce a caching scheme for PDFs by deriving a low-dimensional approximation to sub-path contribution.

This ongoing work is a collaboration with Fredo Durand from MIT and Ravi Ramamoorthi from the University of California San Diego in the context of the CRISP associate team.

### 6.2.6. Unified Color and Texture Transfer for By-Example Scene Editing

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

Color and texture transfer methods are at the heart of by-example image editing techniques. Color transfer well represents the change of overall scene appearance; however it does not represent the change of texture and shape. On the other hand, by-example texture transfer expresses the texture change but it often destroys the target scene structure. We seek the best combination of by-example color and texture transfer to combine these transfer methods so as to selectively work where each method is suitable. Given the source and exemplar pair, the proposed algorithm learns local error metrics which describe if local change between the source and exemplar is best expressed by color or texture transfer. The metric provides us with a local prediction of where we need to synthesize textures using a texture transfer method. This work is a collaboration with Alexei Efros from UC Berkeley in the context of the associate team CRISP.

### 6.2.7. Improved Image-Based Rendering

**Participants:** Rodrigo Ortiz Cayan, Abdelaziz Djelouah, George Drettakis.

Image-based rendering algorithms based on warping present strong artifacts when rendering surfaces at grazing angle. We are working on a new IBR algorithm that overcomes this problem by rendering superpixel segments as piece-wise homography transformations. The input to our method is a set of images calibrated and a 3D point cloud generated from multi-view stereo reconstruction. In pre-processing we robustly fit planes to superpixel segments that contain reconstruction information and then propagate plausible depth and normal information for image-based rendering. Novel views are obtained by re-projecting superpixel segments as homography from different input views, then adaptively blending them according to distortion and confidence estimations.

### 6.2.8. Structured Procedural Textures

**Participants:** Kenneth Vanhoey, George Drettakis.

Textures form a popular tool to add visual detail to shapes, objects and scenes. Manual texture design is however a time-consuming process. An alternative is to generate textures from an input exemplar (*i.e.* an acquired photograph) automatically. The difficulty is to synthesize textures of arbitrary size from a single input, preferably with no repetition artifacts. State of the art synthesis techniques can be categorized in two: copy-based techniques and procedural noise-based ones. The first copy pixels using iterative algorithms. The latter deduce a continuous mathematical function from the exemplar, and evaluate it on the space to be textured. They have the advantage of continuity (no resolution-dependence, minimized memory storage, etc.) and fast



local evaluation suitable for parallel GPU implementation. They are however tedious to define and manipulate. Current state of the art methods are limited to reproducing Gaussian patterns, that is, textures with no or few structure.

We investigate how to go beyond this limit. Noise-based methods constrain the Fourier power spectrum of a texture-generating noise function to resemble the spectrum of the exemplar. By also constraining the phase of the Fourier spectrum to resemble the exemplar, an exact reproduction is obtained, thus lacking variety and showing maximal repetition. By randomizing the phases, an unstructured "same-looking" image is obtained. 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).

In this project, we proceed by investigating the phase spectrum of an image. It contains the structure but identifying how and where is difficult. To characterize structure, we will exploit the splatting process of local random-phase noise and exhibit possible correlations between local phases and spatial placement.

This ongoing work is a collaboration with Ian Jermyn from Durham University.

## 6.3. Perception for Plausible Rendering

### 6.3.1. *An Automated High Level Saliency Predictor for Smart Game Balancing*

**Participant:** George Drettakis.

Successfully predicting visual attention can significantly improve many aspects of computer graphics: scene design, interactivity and rendering. Most previous attention models are mainly based on low-level image features, and fail to take into account high level factors such as scene context, topology, or task. Low-level saliency has previously been combined with task maps, but only for predetermined tasks. Thus, the application of these methods to graphics (*e.g.*, for selective rendering) has not achieved its full potential.

In this work, we present the first automated high-level saliency predictor incorporating two hypotheses from perception and cognitive science that can be adapted to different tasks. The first states that a scene is comprised of objects expected to be found in a specific context as well objects out of context which are salient (scene schemata) while the other claims that viewer's attention is captured by isolated objects (singletons). We proposed a new model of attention by extending Eckstein's Differential Weighting Model. We conducted a formal eye-tracking experiment which confirmed that object saliency guides attention to specific objects in a game scene and determined appropriate parameters for a model. We presented a GPU-based system architecture that estimates the probabilities of objects to be attended in real-time (Figure 5). We embedded this tool in a game level editor to automatically adjust game level difficulty based on object saliency, offering a novel way to facilitate game design. We perform a study confirming that game level completion time depends on object topology as predicted by our system.

This work is a collaboration with George Alex Koulieris and Katerina Mania from the Technical University of Crete and Douglas Cunningham from the Technical University of Cottbus. The work was published in the ACM Transactions on Applied Perception (TAP) Journal [15] and presented as a Talk at SIGGRAPH 2014 in Vancouver.

### 6.3.2. *C-LOD: Context-aware Material Level-of-Detail applied to Mobile Graphics*

**Participant:** George Drettakis.

Attention-based Level-Of-Detail (LOD) managers downgrade the quality of areas that are expected to go unnoticed by an observer to economize on computational resources. The perceptibility of lowered visual fidelity is determined by the accuracy of the attention model that assigns quality levels. Most previous attention based LOD managers do not take into account saliency provoked by context, failing to provide consistently accurate attention predictions.



Figure 5. A low level saliency algorithm indicates that the most salient area of the image is the dark area behind the chair. Our tool highlights the vase at a consistent/singleton location as the most salient object in the image.

In this work, we extended a recent high level saliency model with four additional components yielding more accurate predictions: an object-intrinsic factor accounting for canonical form of objects, an object-context factor for contextual isolation of objects, a feature uniqueness term that accounts for the number of salient features in an image, and a temporal context that generates recurring fixations for objects inconsistent with the context. We conducted a perceptual experiment to acquire the weighting factors to initialize our model. We then designed C-LOD, a LOD manager that maintains a constant frame rate on mobile devices by dynamically re-adjusting material quality on secondary visual features of non-attended objects. In a proof of concept study we established that by incorporating C-LOD, complex effects such as parallax occlusion mapping usually omitted in mobile devices can now be employed, without overloading GPU capability and, at the same time, conserving battery power. We validated our work via eye-tracking (Figure 6 )

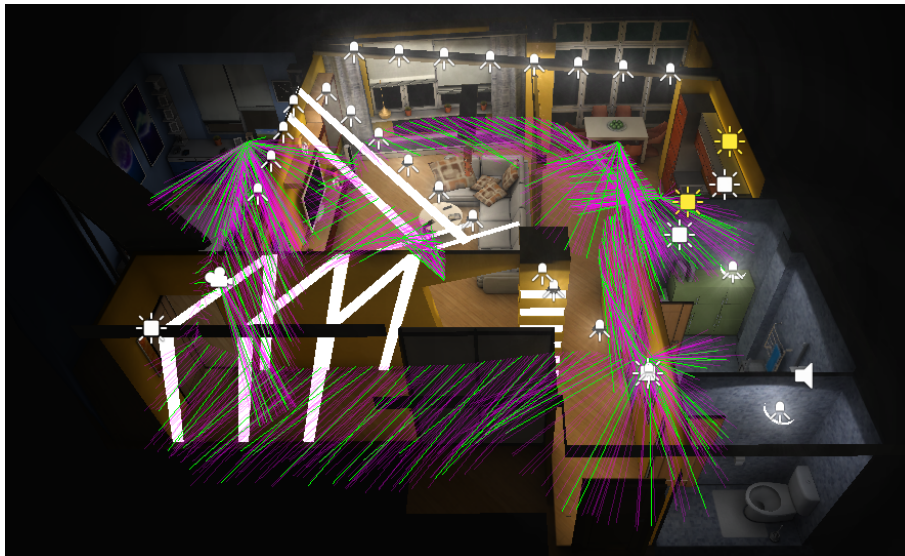


Figure 6. Our validation tool indicates the subject's gaze point with magenta colored beams. The green beams indicate predictions by our attention model.

This work is a collaboration with George Alex Koulieris and Katerina Mania from the Technical University of Crete and Douglas Cunningham from the Technical University of Cottbus. The work was published in a special issue of Computer Graphics Forum [] and was presented at the Eurographics Symposium on Rendering 2014 in Lyon. It was also presented as a poster at SIGGRAPH 2014 in Vancouver winning the 3rd place in the ACM's Graduate Student Research Competition.

## 6.4. Interaction and Design for Virtual Environments

### 6.4.1. Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube

**Participants:** Emmanuelle Chapoulie, George Drettakis.

We present a solution for interaction using finger tracking in a cubic immersive virtual reality system (or immersive cube). Rather than using a traditional flystick device, users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks. Our solution couples finger tracking with a real-time physics engine, combined with a heuristic approach for hand manipulation, which is robust to tracker noise and simulation instabilities. We performed a first study to evaluate our interface with tasks involving complex manipulations, such as balancing objects while walking in the cube. The users finger-tracked manipulation was compared to manipulation with a 6 degree-of-freedom flystick, as well as with carrying out the same task in the real world. Users were also asked to perform a free task, allowing us to observe their perceived level of presence in the scene. Our results showed that our approach provides a feasible interface for immersive cube environments and is perceived by users as being closer to the real experience compared to the flystick. However, the flystick outperforms direct manipulation in terms of speed and precision.

This work is a collaboration with Maria Roussou and Evanthia Dimara from the University of Athens, Maud Marchal from Inria Rennes, and Jean-Christophe Lombardo from Inria Sophia Antipolis. The work has been published in the journal Virtual Reality [13].



Figure 7. A user balancing a tray with both hands [13].

We have also worked on a followup study in which we examine a much more controlled context, studying only very limited movements, in 1D, 2D and 3D. To do this we designed specific devices that can be instantiated both in the virtual world and as physical objects. We compared finger manipulation to wand and to real configurations; the study demonstrated the feasibility of such a controlled comparison for the study of finger-based interaction. This work is in collaboration with InSitu, specifically F. Tsandilas, W. Mackay and L. Oehlberg, and has been accepted for publication in 2015 at IEEE 3DUI.

#### 6.4.2. *Reminiscence Therapy using Image-Based Rendering in VR*

**Participants:** Emmanuelle Chapoulie, George Drettakis, Rachid Guerchouche, Gaurav Chaurasia.

We present a novel VR solution for Reminiscence Therapy (RT), developed jointly by a group of memory clinicians and computer scientists. RT involves the discussion of past activities, events or experiences with others, often with the aid of tangible props which are familiar items from the past; it is a popular intervention in dementia care. We introduced an immersive VR system designed for RT, which allows easy presentation of familiar environments. In particular, our system supports highly-realistic Image-Based Rendering in an immersive setting. To evaluate the effectiveness and utility of our system for RT, we performed a study with healthy elderly participants to test if our VR system could help with the generation of autobiographical memories. We adapted a verbal Autobiographical Fluency protocol to our VR context, in which elderly participants were asked to generate memories based on images they were shown. We compared the use of our image-based system for an unknown and a familiar environment. The results of our study showed that the number of memories generated for a familiar environment is higher than the number of memories obtained for an unknown environment using our system. This indicates that IBR can convey familiarity of a given scene, which is an essential requirement for the use of VR in RT. Our results also showed that our system is as effective as traditional RT protocols, while acceptability and motivation scores demonstrated that our system is well tolerated by elderly participants.

This work is a collaboration with Pierre-David Petit and Philippe Robert from the CMRR in Nice. The work has been published in the Proceedings of IEEE Virtual Reality [18].

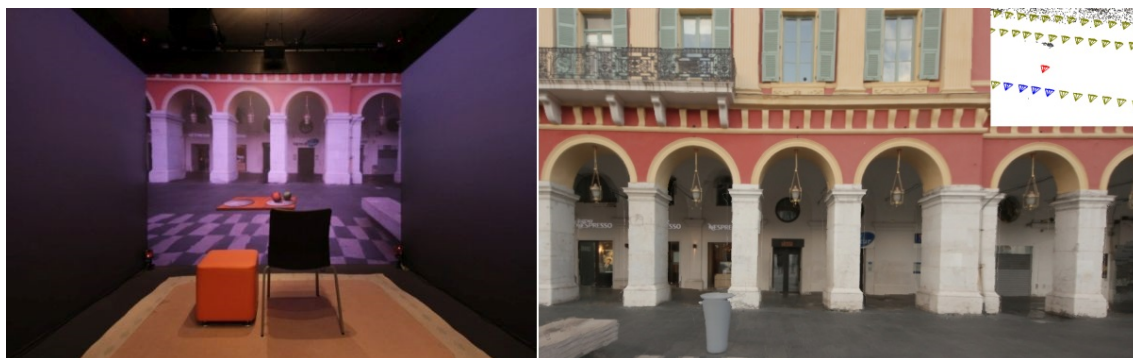


Figure 8. Left: our hardware setup. Right: new point of view reconstructed from input cameras.

#### 6.4.3. *Lightfield Editing*

**Participant:** Adrien Bousseau.

Lightfields capture multiple nearby views of a scene and are consolidating themselves as the successors of conventional photographs. As the field grows and evolves, the need for tools to process and manipulate lightfields arises. However, traditional image manipulation software such as Adobe Photoshop are designed to handle single views and their interfaces cannot cope with multiple views coherently. We conducted a thorough



study to evaluate different lightfield editing interfaces, tools and workflows from a user perspective. We additionally investigate the potential benefits of using depth information when editing, and the limitations imposed by imperfect depth reconstruction using current techniques. We perform two different experiments, collecting both objective and subjective data from a varied number of point-based editing tasks of increasing complexity: In the first experiment, we rely on perfect depth from synthetic lightfields, and focus on simple edits. This allows us to gain basic insight on lightfield editing, and to design a more advanced editing interface. This is then used in the second experiment, employing real lightfields with imperfect reconstructed depth, and covering more advanced editing tasks. Our study shows that users can edit lightfields with our tested interface and tools, even in the presence of imperfect depth. They follow different workflows depending on the task at hand, mostly relying on a combination of different depth cues. Last, we confirm our findings by asking a set of artists to freely edit both real and synthetic lightfields.

This work is a collaboration with Adrian Jarabo, Belen Masia and Diego Gutierrez from Universidad de Zaragoza and Fabio Pellacini from Sapienza Universita di Roma. This work was published at ACM Transactions on Graphics 2014 (Proc. SIGGRAPH) [14].

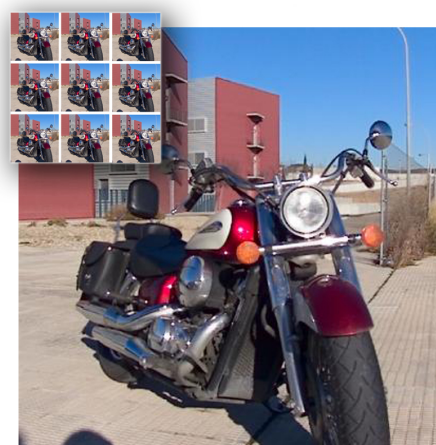


Figure 9. A lightfields represents multiple nearby views of a scene. We conducted a study to evaluate how people edit such data.

#### 6.4.4. True2Form: 3D Curve Networks from 2D Sketches via Selective Regularization

**Participant:** Adrien Bousseau.

True2Form is a sketch-based modeling system that reconstructs 3D curves from typical design sketches. Our approach to infer 3D form from 2D drawings is a novel mathematical framework of insights derived from perception and design literature. We note that designers favor viewpoints that maximally reveal 3D shape information, and strategically sketch descriptive curves that convey intrinsic shape properties, such as curvature, symmetry, or parallelism. Studies indicate that viewers apply these properties selectively to envision a globally consistent 3D shape. We mimic this selective regularization algorithmically, by progressively detecting and enforcing applicable properties, accounting for their global impact on an evolving 3D curve network. Balancing regularity enforcement against sketch fidelity at each step allows us to correct for inaccuracy inherent in free-hand sketching. We perceptually validate our approach by showing agreement between our algorithm and viewers in selecting applicable regularities. We further evaluate our solution by: reconstructing a range of 3D models from diversely sourced sketches; comparisons to prior art; and visual comparison to both ground-truth and 3D reconstructions by designers.

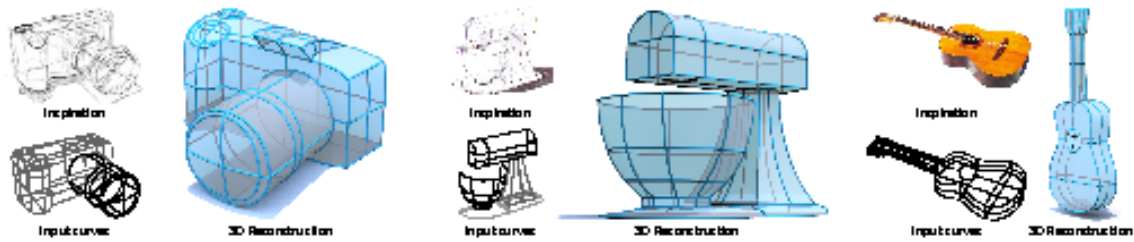


Figure 10. Our single-view modeling system allows us to reconstruct 3D models by tracing curves over existing sketches and photographs.

This work is a collaboration with James McCrae and Karan Singh from the University of Toronto and Xu Baoxuan, Will Chang and Alla Sheffer from the University of British Columbia. The paper was published at ACM Transactions on Graphics 2014 (Proc. SIGGRAPH) [17].

#### 6.4.5. *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. 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, Inria Sophia-Antipolis. The manuscript has been accepted for publication with minor revisions at ACM Transactions on Graphics (TOG).

#### 6.4.6. *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. This 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.



This work is a collaboration with Jean-Dominique Favreau and Florent Lafarge from Titane project team, Inria Sophia-Antipolis and is under submission for the CVPR conference.

#### **6.4.7. *Wrap It! Computer-Assisted Design and Fabrication of Wire Wrapped Jewelry***

**Participants:** Adrien Bousseau, Emmanuel Iarussi.

We developed an interactive tool to assist the process of creating and crafting wire wrapped pieces of jewelry. In a first step, we guide the user in conceiving designs which are suitable to be fabricated with metal wire. In a second step, we assist fabrication by taking inspiration from jigs-based techniques, frequently used by craftsmen as a way to guide and support the wrapping process. Given a vector drawing composed of curves to be fabricated, it is crucial to first decompose it into segments that can be constructed with metal wire. Literature on jewelry-making provides a wide range of examples to perform this task, but they are hard to generalize to any input design. Based on the observation of these examples, we distill and generalize a set of design principles behind the finished pieces of jewelry. Relying on those principles, we propose an algorithm that generates a decomposition of the input where each piece is a single component of wire, that can be wrapped and gathered with the others. In addition, we also automate the design of custom physical jigs for fabrication of the jewelry piece. A jig consists of a board with holes on it, arranged in a regular grid structure. By placing a set of pins (of different radius) on the jig, the craftsman builds a support structure that guides the wrapping process. The wire is bended and twisted around those pins to create the shape. Given the input design curves and the available jig parameters (size, number and radius of the pins), we propose an algorithm to automatically generate an arrangement of pins in order to better approximate the input curve with wire. Finally, users can follow automatically-generated step-by-step instructions to place the pins in the jig board and fabricate the end piece of jewelry.

This ongoing 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.

#### **6.4.8. *Studying how novice designers communicate with sketches and prototypes***

**Participant:** Adrien Bousseau.

We performed a user study to better understand how novice designers communicate a concept during the different phases of its development. Our study was conducted as a one-day design contest where participants had to propose a concept, present it to a jury, describe it to an engineer and finally fabricate a prototype with the help of another participant. We collected sketches and videos for all steps of this exercise in order to evaluate how the concept evolves and how it is described to different audiences. We hope that our findings will inform the development of better computer-assisted design tools for novices.

This is an ongoing work in collaboration with Wendy McKay, Theophanis Tsandilas and Lora Oehlberg from the InSitu project team - Inria Saclay, in the context of the ANR DRAO project.

#### **6.4.9. *Vectorising Bitmaps into Semi-Transparent Gradient Layers***

**Participants:** Christian Richardt, Adrien Bousseau, George Drettakis.

Vector artists create complex artworks by stacking simple layers. We demonstrate the benefit of this strategy for image vectorisation, and present an interactive approach for decomposing bitmap drawings and studio photographs into opaque and semi-transparent vector layers. Semi-transparent layers are especially challenging to extract, since they require the inversion of the non-linear compositing equation. We make this problem tractable by exploiting the parametric nature of vector gradients, jointly separating and vectorising semi-transparent regions. Specifically, we constrain the foreground colours to vary according to linear or radial parametric gradients, restricting the number of unknowns and allowing our system to efficiently solve for an editable semi-transparent foreground.



Figure 11. Our interactive vectorisation technique lets users vectorise an input bitmap (a) into a stack of opaque and semi-transparent vector layers composed of linear or radial colour gradients (b). Users can manipulate the resulting layers using standard tools to quickly produce new looks (c). Semi-transparent layers are outlined for visualisation; these edges are not part of our result.

We propose a progressive workflow, where the user successively selects a semi-transparent or opaque region in the bitmap, which our algorithm separates as a foreground vector gradient and a background bitmap layer. The user can choose to decompose the background further or vectorise it as an opaque layer. The resulting layered vector representation allows a variety of edits, as illustrated in Figure 11, such as modifying the shape of highlights, adding texture to an object or changing its diffuse colour. Our approach facilitates the creation of such layered vector graphics from bitmaps, and we thus see our method as a valuable tool for professional artists and novice users alike.

This work is a collaboration with Jorge Lopez-Moreno, now a postdoc at the University of Madrid, and Maneesh Agrawala from the University of California, Berkeley in the context of the CRISP Associated Team. The paper was presented at the Eurographics Symposium on Rendering (EGSR) 2014, and is published in a special issue of the journal Computer Graphics Forum [16].

## TITANE Project-Team

# 6. New Results

## 6.1. Analysis

### 6.1.1. Parametric Object Detection in Large Scenes

**Participant:** Florent Lafarge.

Point processes are a natural extension of Markov Random Fields (MRF), designed to handle parametric objects. They have shown efficiency and competitiveness for tackling object extraction problems in vision. Simulating these stochastic models is however a difficult task. The performances of the existing samplers are limited in terms of computation time and convergence stability, especially on large scenes. We propose a new sampling procedure based on a Monte Carlo formalism [8]. Our algorithm exploits the Markovian property of point processes to perform the sampling in parallel. This procedure is embedded into a data-driven mechanism so that the points are distributed in the scene in function of spatial information extracted from the input data. The performances of the sampler are analyzed through a set of experiments on various object detection problems from large scenes, including comparisons to the existing algorithms. The sampler is also tested as optimization algorithm for MRF-based labeling problems.

## 6.2. Reconstruction

### 6.2.1. Indoor Scene Reconstruction

**Participants:** Sven Oesau, Florent Lafarge, Pierre Alliez.

*In collaboration with EADS ASTRIUM*

We contributed a method for automatic reconstruction of permanent structures of indoor scenes, such as walls, floors and ceilings, from raw point clouds acquired by laser scanners [6]. Our approach employs graph-cut to solve an inside/outside labeling of a space decomposition. To allow for an accurate reconstruction the space decomposition is aligned with permanent structures. A Hough Transform is applied for extracting the wall directions while allowing a flexible reconstruction of scenes. The graph-cut formulation takes into account data consistency through an inside/outside prediction for the cells of the space decomposition by stochastic ray casting, while favoring low geometric complexity of the model. Our algorithm produces watertight reconstructed models of multi-level buildings and complex scenes.

### 6.2.2. State of the Art in Surface Reconstruction from Point Clouds

**Participant:** Pierre Alliez.

*In collaboration with Matthew Berger, Andrea Tagliasacchi, Lee Seversky, Joshua Levine, Andrei Sharf and Claudio Silva.*

The area of surface reconstruction has seen substantial progress in the past two decades. The traditional problem addressed by surface reconstruction is to recover the digital representation of a physical shape that has been scanned, where the scanned data contains a wide variety of defects. While much of the earlier work has been focused on reconstructing a piece-wise smooth representation of the original shape, recent work has taken on more specialized priors to address significantly challenging data imperfections, where the reconstruction can take on different representations – not necessarily the explicit geometry. This state-of-the-art report surveys the field of surface reconstruction, providing a categorization with respect to priors, data imperfections, and reconstruction output. By considering a holistic view of surface reconstruction, this report provides a detailed characterization of the field, highlights similarities between diverse reconstruction techniques, and provides directions for future work in surface reconstruction [11].

### 6.2.3. Robust Shape Reconstruction and Optimal Transportation

**Participants:** Simon Giraudot, Pierre Alliez.

*In collaboration with David Cohen-Steiner.*

We describe a framework for robust shape reconstruction from raw point sets, based on optimal transportation between measures, where the input point sets are seen as distribution of masses. In addition to robustness to defect-laden point sets, hampered with noise and outliers, our approach can reconstruct smooth closed shapes as well as piecewise smooth shapes with boundaries [10].

## 6.3. Approximation

### 6.3.1. Zometool Shape Approximation

**Participants:** Henrik Zimmer, Florent Lafarge, Pierre Alliez.

*In collaboration with Leif Kobbelt.*

We contributed an algorithm that approximates 2-manifold surfaces with Zometool models while preserving their topology. Zometool is a popular hands-on mathematical modeling system used in teaching, research and for recreational model assemblies at home. This construction system relies on a single node type with a small, fixed set of directions and only nine different edge types in its basic form. While being naturally well suited for modeling symmetries, various polytopes or visualizing molecular structures, the inherent discreteness of the system poses difficult constraints on any algorithmic approach to support the modeling of freeform shapes. We contribute a set of local, topology preserving Zome mesh modification operators enabling the efficient exploration of the space of 2-manifold Zome models around a given input shape. Starting from a rough initial approximation, the operators are iteratively selected within a stochastic framework guided by an energy functional measuring the quality of the approximation. We demonstrate our approach on a number of designs and also describe parameters which are used to explore different complexities and enable coarse approximations [15].

### 6.3.2. CGALmesh: a Generic Framework for Delaunay Mesh Generation

**Participants:** Pierre Alliez, Clement Jamin.

*In collaboration with Jean-Daniel Boissonnat and Mariette Yvinec.*

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. 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 [5].

### 6.3.3. Level-of-Detail Quad Meshing

**Participant:** David Bommès.

*In collaboration with Hans-Christian Ebke and Leif Kobbelt from RWTH Aachen.*

The most effective and popular tools for obtaining feature aligned quad meshes from triangular input meshes are based on cross field guided parametrization. These methods are incarnations of a conceptual three-step pipeline: (1) cross field computation, (2) field-guided surface parametrization, (3) quad mesh extraction. While in most meshing scenarios the user prescribes a desired target quad size or edge length, this information is typically taken into account from step 2 onwards only, but not in the cross field computation step. This turns into a problem in the presence of small scale geometric or topological features or noise in the input mesh: closely placed singularities are induced in the cross field, which are not properly reproducible by vertices in a quad mesh with the pre-scribed edge length, causing severe distortions or even failure of the meshing algorithm. We reformulate the construction of cross fields as well as field-guided parameterizations in a scale-aware manner which effectively suppresses densely spaced features and noise of geometric as well as topological kind. Dominant large-scale features are adequately preserved in the output by relying on the unaltered input mesh as the computational domain [4].

#### **6.3.4. Mesh Watermarking based on a Constrained Optimization Framework**

**Participants:** Xavier Rolland-Nevière, Pierre Alliez.

*In collaboration with Technicolor and Gwenael Doerr.*

A watermarking strategy for triangle surface meshes consists in modifying the vertex positions along the radial directions, in order to adjust the distribution of radial distances and thereby encode the desired payload. To guarantee that watermark embedding does not alter the center of mass, prior work formulated this task as a quadratic programming problem. We contribute a generalization of this formulation with: (i) integral reference primitives, (ii) arbitrary relocation directions to alter the vertex positions, and (iii) alternate distortion metrics to minimize the perceptual impact of the embedding process. These variants are evaluated against a range of attacks and we report both improved robustness performances, in particular for simplification attacks, and improved control over the embedding distortion [9].

#### **6.3.5. Robust 3D Watermarking**

**Participants:** Xavier Rolland-Nevière, Pierre Alliez.

*In collaboration with Technicolor, thesis co-advised by Pierre Alliez and Gwenael Doerr.*

3D models are valuable assets widely used in the industry and likely to face piracy issues. This dissertation deals with robust mesh watermarking that is used for traitor-tracing. Following a review of state-of-the-art 3D watermarking systems, the robustness of several content adaptation transforms are evaluated. An embedding domain robust against pose is investigated, with a thickness estimation based on a robust distance function to a point cloud constructed from some mesh diameters. A benchmark showcases the performance of this domain that provides a basis for robust watermarking in 3D animations. For static meshes, modulating the radial distances is an efficient approach to watermarking. It has been formulated as a quadratic programming problem minimizing the geometric distortion while embedding the payload in the radial distances. This formulation is leveraged to create a robust watermarking framework, with the integration of the spread-transform, integral reference primitives, arbitrarily selected relocation directions and alternate metrics to minimize the distortion perceived. Benchmarking results showcase the benefits of these add-ons w.r.t the fidelity vs. robustness watermarking trade-off. The watermark security is then investigated with two obfuscation mechanisms and a series of attacks that highlight the remaining limitations. A resynchronization approach is finally integrated to deal with cropping attacks. The resynchronization embeds landmarks in a configuration that conveys synchronization information that will be lost after cropping. During the decoding, this information is blindly retrieved and significant robustness improvements are achieved [2].

#### **6.3.6. Spread Transform and Roughness-based Shaping to Improve 3D Watermarking based on Quadratic Programming**

**Participants:** Xavier Rolland-Nevière, Pierre Alliez.

*In collaboration with Technicolor and Gwenael Doerr.*

Modulating the distances between the vertices and the center of mass of a triangular mesh is a popular approach to watermark 3D objects. Prior work has formulated this approach as a quadratic programming problem which minimizes the geometric distortion while embedding the watermark payload in the histogram of distances. To enhance this framework, we introduce two watermarking components, namely the spread transform and perceptual shaping based on roughness information. Benchmarking results showcase the benefits of these add-ons with respect to the fidelity-robustness trade-off [13].



## ALPAGE Project-Team

## 6. New Results

### 6.1. Highlights of the Year

**Benoit Crabbé is a Junior Member of the Institut Universitaire de France (IUF)** since October 2014. Two out of the five academic staff at Alpage are now member of the IUF, Laurence Danlos being a Senior Member since October 2013.

### 6.2. Automatic text normalisation

**Participants:** Benoît Sagot, Marion Baranes.

Since the emergence of the web, one of the goals of natural language processing (NLP) tools has been analysing raw noisy text documents such as blogs, review sites or social networks. These texts commonly contain misspellings, redundant punctuation, smileys, etc. Consequently they require specific preprocessing before being used in different NLP applications. That is why, we worked at Alpage on the development of a new corpora and the implementation of an automatic system for normalisation of such texts:

- **Corpus crap** In 2014, a large-scale extension of the number of normalisation rules used by the MElt part-of-speech tagger for processing noisy computer-generated content has been achieved. This work was carried out in the context of and based on corpora developed within the CoMeRe project, funded by the Institut de Linguistique Française and lead by Thierry Chanier [14].
- **Normalisation system** We have implemented a modular system which follows SxPipe [109]. This system detects if an unknown word to a reference lexicon corresponds to a non-word error (and is not a neologisme or a borrowing). Then, it attempts to normalize non-word errors and grammatical errors. In 2014, we focused on these two latter tasks. First, we have implemented a system which suggests one or several normalization candidates for these non-word errors. As described in [17], to do that, we use an analogy-based approach for acquiring normalisation rules and use them in the same way as lexical spelling correction rules. Secondly, we propose to normalize grammatical errors. To do that, we check for each word if it has common homophones. If this is the case, we consider these homophones as possible candidates for normalization. Finally, we filter all these candidates in order to keep only the one which is the most probable. This filtration is done using a probabilistic model based on a  $n$ -gram system. Moreover, the implementation of this system of normalisation motivated a side task. We developed an unsupervised method for acquiring pairs of lexical entries belonging to the same morphological family, i.e., derivationally related words, starting from a purely inflectional lexicon. This work, detailed in [16], allows us to create new linguistic resources for English, French, German and Spanish which contains derivational relations.

### 6.3. The impact of morphosyntactic processing on post-OCR error correction

**Participants:** Kata Gábor, Benoît Sagot, Pierre Magistry.

State of the art optical character recognition (OCR) software currently achieve an error rate of around 1 to 10% depending on the age and the layout of the text. To our knowledge, very little work has been done to exploit linguistic analysis for post-OCR error correction. Within the PACTE project we are conducting research on reducing the OCR error rate by using contextual information and linguistic processing.

In 2014 we continued our investigations on how named entity recognition can benefit OCR error detection by applying context-aware error correction rules directly to the OCR output. Several grammars have been created or improved to adress OCR problems occurring within different types of named entities. As a result, the SxPipe-PACTE toolchain was created to correct named entities in a noisy input [45], [31].

While the symbolic error correction method works with a very high precision, its limitation lies in its relatively low coverage. In order to deal with the errors occurring outside the recognized entities, we studied the possibility of using lattice-based part of speech tagging to select the best correction hypothesis in context. Different methods were investigated to generate correction hypotheses, using word alignment software or by observing frequently occurring error types. The initial results confirm that a significant number of the remaining OCR errors can be corrected via lattice-based tagging, as long as the noise introduced by correction hypotheses is controlled.

## 6.4. Linear-time discriminant syntactico-semantic parsing

**Participants:** Benoit Crabbé, Maximin Coavoux, Djamé Seddah.

In this module we study efficient and accurate models of statistical phrase structure parsing. We focus on linear time lexicalized parsing algorithms (shift reduce, left corner) with approximations entailing linear time processing. The existing prototype involves a global discriminant parsing model of the large margin family (Perceptron, Mira, SVM avatars) able to parse user defined structured input tokens [23]. 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 participation to the SPRML 2014 shared task on parsing morphologically rich languages has been a first step towards testing our model in a multilingual setting where we were among the state of the art systems and state of the art on some languages such as Polish. To our knowledge the parser is one of the fastest existing multilingual parser worldwide (4000 – 8000 tokens/sec.). In order to ease model design for multilingual settings, we currently study efficient feature selection procedures for automating model adaptation to new languages.

The ongoing investigation aims to integrate continuous semantic representations into the model such as word embeddings in order to leverage data sparsity and estimation issues recurrent in lexicalized parsing. To this end we study neural-network-based architectures for structured phrase structure parsing.

## 6.5. Playing with DyALog-based parsers

**Participant:** Éric Villemonte de La Clergerie.

Éric de la Clergerie has continued the development of DYALOG-SR, a transition-based dependency parser running on top of DYALOG and initiated in 2013 to participate to SPMRL'2013. Thanks to DYALOG's tabulation functionalities, this parser implements a dynamic programming algorithm to explore larger search space through the use of beams.

In order to participate to SemEval'14 Task on "broad coverage semantic dependency parsing", DYALOG-SR was extended to handle non-connected dependency graphs rather than standard dependency trees. This was achieved by considering a richer set of transitions, besides the usual Shift and Reduce transitions. However, while working, this extended set of transitions was not ensuring the expected gains when using beams. The issue was finally solved after long investigations, with the identification of multiple causes. One of them was related to the fact that transition paths of various lengths may lead to a final state. In consequence, a noop transition was added to compensate on shorter paths.

A second axe of work was a thorough use of DYALOG-SR over the French TreeBank (FTB) to compare its performances to those published for other parsers. By enriching its set of features and improving the update strategy of the perceptron-based statistical model of DYALOG-SR, we were able to reach state-of-the-art results.

However, the best results were obtained by coupling DYALOG-SR with FRMG, our large-coverage French grammar (derived from a meta-grammar). The results from FRMG were used as features to guide the statistical DYALOG-SR parser. This innovative step proved to provide us with the best results published so far for the FTB (over 90% of Labeled Attachment Score [LAS] over the test part of the FTB) [41].

The improvements of FRMG was pursued in 2014, at the level of the underlying meta-grammar (to extend its coverage over 96% on the FTB) but also by adapting the statistical models developed for DYALOG-SR (in replacement of older and slower SQLite-based models).

## 6.6. Multiword expressions and statistical parsing

**Participants:** Sarah Beniamine, Marie-Hélène Candito, Benoît Sagot, Djamé Seddah.

Multi-word expressions recognition (MWE recognition) and syntactic parsing are two tasks that have been extensively investigated. Yet, systems combining both tasks have been rather rare. In particular, works on parsing have tended to use training and test data with gold MWEs (generally with each MWE) merged into one token. In 2013, Djamé Seddah led the organization of the first shared task on statistical parsing Morphologically Rich Languages (SPMRL) [127], hosted by the fourth SPMRL workshop. The primary goal of this shared task was to bring forward work on parsing morphologically ambiguous input in both dependency and constituency parsing, and to show the state of the art for MRLs. The shared task proposed a data set for 9 languages. The French part of this data set is particular, in that it uses a representation combining MWEs and syntax, which allows to investigate techniques for performing parsing and MWE recognition. A first system was proposed for the dependency parsing track of the Shared Task, in collaboration with Matthieu Constant (LIGM, Université Marne-la-Vallée) [74]. This work investigates pipeline and joint architecture for both tasks. In 2014, Marie Candito and Matthieu Constant continued that line of work [2], focusing on using an alternative representation of syntactically regular MWEs, which captures their syntactic internal structure. The objective of such representation was two fold. First, it is well-known that the MWE status is not clear-cut, and that MWE status can hold due to syntactic and/or semantic criteria. In particular, syntactically regular MWEs exhibit various degrees of semantic non-compositionality. For such MWEs, an atomic representation fails to capture internal partial semantic composition, and also fails to take advantage of the internal syntactic regularity. Indeed, one hypothesis of this work was that augmenting the regularity of the syntactic representations could help parsing. The results of this work is that while this hypothesis could not be verified, the resulting system has comparable performance to that of previous works on this dataset, but it has the advantage of predicting both syntactic dependencies and the internal structure of MWEs, a crucial feature to capture the various degrees of semantic compositionality of MWEs.

In the same time, Sarah Beniamine and Benoît Sagot also investigated the use of internal regular structures for MWEs, yet for *syntagmatic* syntactic parsing. The objective is to guide a parser with predicted MWEs, while keeping a regular syntactic representation.

## 6.7. Graph-based approaches for deep-syntactic and semantic parsing

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

With most state-of-the-art statistical parsers routinely crossing a ninety percent performance plateau in capturing tree structures, the question of *what next* crucially arises. Most of the structures used to train current parsing models are degraded versions of a more informative data set: the Wall Street journal section of the Penn treebank ([91]) which is often stripped from its richer set of annotations (i.e. traces and functional labels are removed), while, for reasons of efficiency and availability, projective dependency trees are often given preference over richer graph structures [96], [107]. This led to the emergence of *surface* syntax-based parsers [70], [97], [100] whose output cannot by itself be used to extract full-fledged predicate argument-structures. For example, control verb constructions, it-cleft structures, argument sharing in ellipsis coordination, etc. are among the phenomena requiring a graph to be properly accounted for. The dichotomy between what can usually be parsed with high accuracy and what lies in the deeper syntactic description has initiated a line of research devoted to closing the gap between surface syntax and richer structures.

At Alpage, we built our work on the widely known transition-based parsing approach [95], which is state-of-the-art to parse surfacic syntactic trees [141]. Shift-reduce transition-based parsers essentially rely on *configurations* formed of a stack and a buffer, with stack transitions used to move from a configuration to the next one, until reaching a final configuration.

## 6.8. English Broad-coverage Semantic Dependency Parsing

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

We successfully tested our graph-based approach described in Section 6.7 on a shared task on broad-coverage semantic dependency parsing part of the International Workshop on Semantic Evaluation (SemEval 2014, [99]). We were given three resources, which constitute parallel semantic annotations over the same common text (the Penn Treebank (PTB), [91]). The first one is part of the tectogrammatical layer of the Prague Czech-English Dependency Treebank, the second one is the reduction of the Minimal Recursion Semantics, available through the HPSG annotation of the PTB, into bi-lexical dependencies [82]. Finally, the third one is the predicate-argument structures extracted from the Enju Parser [131]. The shared task consisted of two tracks: a closed one where we needed to use these three resources only and an open one, where we could use whatever we needed to produce the best semantic representations.

At Alpage, we developed two semantic parsers: The first one is based on a previous work on DAG parsing [107] and the second one on the FRMG surfacic syntactic parser [133]. We use two parsers to assess the validity of our approach. The top performing models we submitted used a mix of syntactic features (tree fragments from a constituent syntactic parser [100], dependencies from a syntactic parser [58], elementary spinal trees using a spine grammar [126], etc.) to improve our results. Our intuition is that syntax and semantic are not independent of each other and using syntax could improve semantic parsing. Our systems performs well and were able to compete with the top performers. Those systems, as well as the software needed to parse these new data sets, are already available.

## 6.9. Development of syntactic and deep-syntactic treebanks: Extending our Coverage

**Participants:** Djamé Seddah, Marie-Hélène Candito, Corentin Ribeyre, Benoît Sagot, Éric Villemonte de La Clergerie.

Taking its roots in the teams that initiated the first syntactically annotated the French Treebank, the first metagrammar compiler and one of the best wide coverage grammars, Alpage has a strong tendency to focus on creating pioneer resources that serve both to extend our linguistics knowledge and to nurture accurate parsing models. Recently, we focused on extending the lexical coverage of our parsers using semi-supervised techniques (see above) built on edited texts. In order to evaluate these models, we built the first free out-domain treebank for French (the Sequoia treebank, [69]) covering various domains such as Wikipedia, Europarl and bio medical texts on which we established the state-of-the-art. Exploring other kind of texts (speech, user generated content), we faced however various issues inherently tied to the nature of these productions. Syntactic divergences from the norm are actually prominent and are a severe bottleneck for any data driven parsing model. Simply because a structure not present in a training set cannot be reproduced. This analysis naturally occurred as a side effect of our experiments in parsing social media texts. Actually, the first version of the French Social Media Bank (FSMB) was conceived as a stress test for our tool chains (tokenization, tagging, parsing). Our recent experiments showed that to reach a decent performance plateau, we need to include some of the target data into our training set. Focusing on processing direct questions and social media texts, we built two treebanks of about 2,500 sentences each: one devoted to questions and one built to extend the FSMB<sup>0</sup>. These initiatives are funded by the Labex EFL.

- The French Social Media Bank 2.0: We are about to release the second part of the FSMB, 2600 sentences from Twitter, Facebook and other sources, with an extended annotation scheme able to describe more precisely the various phenomena at stakes in the social media text streams. To do so we extended our pre-processing chain (included and available in the MeLT tagger) to include a much more robust normalizer and tokenizer than the one we used to build the first version of the FSMB. The building phase being over, publications on this topics are on preparation.

<sup>0</sup>Let us note that the ever evolving nature of user generated content makes this a necessity.

- The French Question Bank: The building of a treebank made solely of questions comes from the simple fact that in both the FTB and the Sequoia treebank, there's only 150 direct questions. Making the parsing of such constructions extremely difficult for our data driven parsers. Following our now classical methodology, we selected more than 3200 sentences coming from governmental sources, from the TREC resources – allowing to have a strong set of aligned sentences with the English resources – and from social media sources as well. In the case of the TREC part, those are the questions used by [85], which allows some potentially interesting cross-language experiments. Unlike in the English Question Bank, phrasal-movement are annotated with functional paths and not traces. This allows to maintain a strong compatibility with the FTB annotation scheme. Our Question bank is the only resources of its kind for any other languages than English.

Both resources are available in constituency and dependency. The later being still verified for the FSMB 2.0.

Note that we just started another annotation campaign aiming at adding a deep syntax layer to these two data sets, following the Deep Sequoia as presented above. These resources will prove invaluable to building a robust data driven syntax to semantic interface.

In the same time, Alpage collaborated with the Nancy-based Inria team Sémagramme in the domain of deep syntax analysis. Deep Syntax is intended as an intermediary level of representation, at the interface between syntax and semantics, which partly abstracts away from syntactic variation, and aims at providing the canonical grammatical functions of predicates. This means for instance neutralizing diathesis alternation and making explicit argument sharing, such as occurring for infinitival verbs. The advantage of a deep syntactic representation is to provide a more regular representation to serve as basis for semantic analysis. Note though it is computationally more complex, as we switch from surface syntactic trees to deep syntactic graphs, since shared arguments are made explicit.

We collaboratively defined a deep syntactic representation scheme for French and built a gold deep syntactic treebank [21], [43]. More precisely, each team used an automatic surface-to-deep syntax converter module, applied it on the Sequoia corpus (already annotated for surface syntax), and manually corrected it. Remaining differences were collaboratively adjudicated. The surface-to-deep syntax converter tool used by Alpage is built around the OGRE Graph Rewriting Engine built by Corentin Ribeyre [105].

The Deep Sequoia Treebank is too small to train a deep syntactic analyzer directly. In order to obtain more annotated data, we further used the surface-to-deep syntax converter to obtain predicted (non validated) deep syntactic representations for the French Treebank [36], which is much bigger than the Sequoia treebank (more than 18.000 sentences compared to 3,000 sentences). We performed an evaluation of a small subset of the resulting deep syntactic graphs. The high level of performance we obtained (more than 98% of F-score in labeled dependencies recovery task) which suggests that the deep syntax version of the French Treebank can be used as pseudo-gold data to train deep syntactic parsers, or to extract syntactic lexicons augmented with quantitative information.

## 6.10. Towards a French FrameNet

**Participants:** Marie-Hélène Candito, Marianne Djemaa, Benoît Sagot.

The ASFALDA project <sup>0</sup> 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. It is a three-year project which started in October 2012, with the objective of building 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 [57], <sup>0</sup> 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 :

<sup>0</sup><https://sites.google.com/site/anrasfalda/>

<sup>0</sup><https://framenet.icsi.berkeley.edu/>

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 2014, we first finished the work on the lexicon, which was started in 2013 [19]. The step 2 (semantic annotations on top of syntactic representations) is ongoing :

- We wrote the annotation guide. In particular Marianne Djemaa focused on how to annotate phenomenon known to exhibit syntax/semantic divergences [42].
- We designed the annotation workflow and built an automatic pre-annotator, which proposes candidate semantic annotations that must be disambiguated manually.
- We started in July 2014 to manage six annotators, who were hired to perform the manual annotation phase.

## 6.11. 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 [93] but semantic resources for French are scarce. Moreover, existing resources offer fine grained sense distinctions 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 [40]
- 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 [12]

## 6.12. Development of Verb $\ni$ net

**Participants:** Laurence Danlos, Quentin Pradet.

VerbNet is an English lexical resources for verbs, which is internationally known and widely used in numerous NLP applications [89]. Verb $\ni$ 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 verbes 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, [35], [28]. This work is currently under progress in collaboration with Takuya Nakamura (Institut Gaspard Monge) and the resource should be freely available in 2015.

## 6.13. Development of FDTB1

**Participants:** Laurence Danlos, Margot Colinet, Jacques Steinlin.



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 [106], 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 practical 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 dépêché 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 dépêché 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 [22], whose results have been used to enhance Lefff, [44].

The FDTB corpus contains 18 535 sentences and FDTB1 identifies 9 833 discourse connectives. This resource is freely available.

## 6.14. Discourse Parsing

**Participants:** Laurence Danlos, Chloé Braud.

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 more 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 work, Chloé Braud develops a discourse relation classifier, carrying experiments on French and English. Focusing on the problem on implicit relation identification, this work tries to tackle the lack of manually annotated data, a discourse specific difficulty, by exploiting the similarities between explicit and implicit relations. In 2014, this work lead to systems based on domain adaptation methods [18], [13], demonstrating improvements on the French corpus Annodis [56].

## 6.15. Multilingual and cross-lingual terminology extraction

**Participants:** Valérie Hanoka, Benoît Sagot.

Language diversity spans more than 7000 languages. Among them, 24 macrolanguages<sup>0</sup> have at least 50 million first-language speakers. Traditional terminology techniques, which are mostly based on language-dependent linguistic tools (part of speech tagging, phrase chunking) requires a considerable effort to be developed for a new language. This effort is likely to be even more critical if the term extraction is to be based on noisy text (i.e. displaying linguistic creativity, spelling errors and ungrammatical sentences). In this context, the need has arisen to examine the issue of a less language-specific method for term extraction.

To that end, our approach take advantage of existing language typologies in order to alleviate for the lack of language-dependent linguistic processing. We based our reflexions and experiments on a sample of 7 typologically different language: Arabic, Chinese, English, French, German, Polish and Turkish.

<sup>0</sup>A macrolanguage is defined as "multiple, closely related individual languages that are deemed in some usage contexts to be a single language" in the ISO 639-3 standard.

As a starting point, we considered the minimal textual preprocessing (character normalization, segmentation) needed to allow for a comprehensive multilingual approach to automatic term extraction. In order to gain further insight on the influence of the morphology for term extraction, we examined the impact of the deletion of selected morphological information on words of morphologically rich languages.

For the different settings, models based on Conditional Random Fields (CRF) have been trained on existing gold data. We proposed an adapted version of the evaluation algorithm of [94] able to issue terminological scores for all the language of our sample. The scores thus obtained allowed to identify the best experimental setting for each language tested.

The results were surprising in two ways: First, the cross-lingual<sup>0</sup> application of models works well (the best cross-lingual models' accuracies range from 0.8% to 0.97%). Secondly, the languages which makes the overall best cross-lingual models are those who have the richest morphology (i.e: Turkish).

Finally, we developed and used a multilingual translation graph [32] to extend the multilingual terminology obtained using two methods: those presented in [83] and a more formal one, based on a simulated annealing clustering algorithm.

## 6.16. Word order variation in Old French

**Participants:** Benoit Crabbé, Alexandra Simonenko, Benoît Sagot.

As participant of the strand *Experimental Grammar* of the Labex EFL project *Empirical Foundations of Linguistics*<sup>0</sup> 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.<sup>0</sup>

The quantitative results will be compared with what is known from corpus studies on the relative ordering of subject and complement in Old French [90]. 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 [64]. This comparative perspective is expected to provide new insights on French language evolution.

## 6.17. Cross linguistic factors governing word order

**Participants:** Benoit Crabbé, Kristina Gulordava.

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 [63].

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, we explore to which extent a computational corpus-based analysis can provide new evidence not only for empirical, but also for theoretical linguistic research.

## 6.18. Anaphoricity detection and coreference resolution

**Participant:** Emmanuel Lassalle.

<sup>0</sup>A model trained on data of one language and applied to data of another language

<sup>0</sup>[www.labex-efl.org](http://www.labex-efl.org)

<sup>0</sup>SRCMF corpus: <http://srcmf.org/>; MCVF: <http://www.voies.uottawa.ca>

Resolving coreference in a text, that is, partitioning mentions (noun phrases, verbs, etc) into referential entities, is a challenging task in NLP leading to many different approaches. Anaphoricity detection, on the other hand, consists in deciding whether a mention is anaphoric (aka discourse-old) or non-anaphoric (discourse-new). This task is strongly related to coreference resolution and has been mainly addressed as a preliminary task to solve, leading to pipeline architectures.

A first line of work compares several methods for learning latent structures encoding coreference clusters that optionally take into account very accurate constraints on mention pairs. We study the relationship between standard decoding strategies used with pairwise models and those used with structured learning of latent structures, providing both topological and empirical comparisons. We also show that further gains can be obtained by the addition of pairwise constraints. Our experiments on the CoNLL-2012 dataset show that our best system obtains state-of-the-art results, and significant gains compared to standard locally-trained models.

Our second line of work introduces a new structured model for learning anaphoricity detection and coreference resolution in a jointly. 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. This model is refined by the use of pairwise constraints, and 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.

This work has been achieved in collaboration with Pascal Denis, a former Alpage member, now at Inria Lille-Nord-Europe (EPI Magnet).

## MULTISPEECH Team

# 6. New Results

## 6.1. Highlights of the Year

The version 2 of our source separation toolbox FASST [65] has been downloaded more than 300 times since its release in January 2014.

## 6.2. Explicit modeling of speech production and perception

**Participants:** Yves Laprie, Slim Ouni, Vincent Colotte, Anne Bonneau, Agnès Piquard-Kipffer, Martine Cadot [Univ. Lorraine], Antoine Liutkus, Emmanuel Vincent, Odile Mella, Benjamin Elie, Camille Fauth, Julie Busset, Andrea Bandini, Guillaume Gris, Simon Meoni.

### 6.2.1. Articulatory modeling

#### 6.2.1.1. Acquisition of articulatory data

Acquisition of articulatory data plays a central role in the construction of articulatory models and investigation of articulatory gestures. In cooperation with the IADI laboratory (Nancy hospital) we thus conducted a series of preliminary experiments intended to acquire cine-MRI data. Images of the film are reconstructed thanks to the cine-GRICS algorithm developed at IADI [56].

The second research track concerns ultrasound (US) imaging which presents the interest of offering a good temporal resolution without any health hazard and at a reasonable price. However, it cannot be used alone because there is no reference coordinate system and no spatial calibration. We thus used a multimodal acquisition system developed by the Magrit team, which uses electromagnetography sensors to locate the US probe, and the method used to calibrate the US modality. We experimented this system to investigate the most appropriate acquisition protocol for Magnetic Resonance Imaging [37].

We also use an articulograph to acquire articulatory data. Within the framework of the EQUIPEX OR-TOLANG, we acquired this year an AG501, a 24-channel articulograph. This system is the most advanced electromagnetography acquisition system. It has been used for two articulatory studies: (1) investigating the effects of posture and noise on speech production [48] and (2) studying the pauses in spontaneous speech from an articulatory point of view. We also conducted an exploratory study on retrieving the 3D shape of the palate from electromagnetography tracings (the work of Simon Meoni, a master student in Cognitive Sciences).

#### 6.2.1.2. Acoustic-to-articulatory inversion

Our previous works about acoustic-to-articulatory inversion relied on the exploration of a vast articulatory codebook covering the whole articulatory space that could be reached by a speaker. This solution presents the main drawback of requiring the construction a codebook for each speaker. We thus developed a multimodal approach to estimate the area function and the length of the vocal tract of oral vowels. The method is based on an iterative technique consisting in deforming an initial area function so that the output acoustic vector matches a specified target. The chosen acoustic vector is the formant frequency pattern. In order to regularize the ill-defined problem, several constraints are added to the algorithm. First, the lip termination area is estimated via a facial capture software. Then, the area function is constrained so that it does not get too far from a neutral position, and so that it does not change too quickly from a temporal frame to the next, when dealing with dynamic inversion. The method proves to be efficient for approximating the area function and the length of the vocal tract for oral French vowels, both in static and dynamic configurations.

### 6.2.1.3. Articulatory models

The development of articulatory models is a crucial aspect of articulatory synthesis since this determines the success of synthesis. The previous model was developed for X-ray images. This means that the laryngeal part of the model associates the larynx with the piriform sinuses event if these two structures are not in the same sagittal plane. The new model separates the two structures if needed. Additionally, the larynx and the epiglottis are controlled independently which corresponds to the anatomical truth. Previous attempts to modeling epiglottis used principal component analysis applied to the contours drawn on X-ray images. Unfortunately the width of the epiglottis varies from one image to the other and PCA thus learns a spurious “inflating” component. The new model uses the epiglottis centerline plus a constant width which prevents this error.

The second major improvement concerns the use of virtual targets in the construction of the articulatory model. Virtual targets are used to separate the contribution of the tongue contour from those of the palate. The objective is to render the articulation of consonants more correctly since they require a contact between the tongue and the palate at a very precise point [38].

These two improvements of the articulatory model were used in the articulatory copy synthesis experiments [11].

The construction of models was also tackled from a data mining point of view. A robust data mining approach was designed to automatically extract complex statistically significant connections between data (e.g. interactions between more than two variables). This work could be used for data other than X-ray images [54].

### 6.2.2. Expressive acoustic-visual synthesis

Right now, we are investigating the state-of-the-art of the field of expressive speech and how to acquire efficiently expressive speech corpus. As a first step, we are also investigating visual acquisition techniques to track facial expression. This is the work of the visiting PhD student Andrea Bandini (from University of Bologna). Another step toward expressive speech synthesis is to have an expressive face model. In this context, the expressivity is mainly based on the dynamics. In fact, when the human facial movements are natural and accurately replicated on the 3D model, we can reach a reasonable expressivity. In this context, we are conducting new research toward an expressive talking head. In this context, we acquired a high-resolution 3D model of a human speaker and we are developing methods to animate the model using motion capture data. This was the work of the master student Guillaume Gris. We also investigated the advantage of generating visual speech from sequences of 2D Images, when the 3D data is lacking [43].

### 6.2.3. Categorization of sounds and prosody for native and non-native speech

Categorization of sounds and prosody for non-native speech is the object of the ANR+DFG project IFCASL devoted to French and German languages. Within this project, we built a bilingual corpus and started a study about the realization of (final) voicing in both languages. We also gave a training course about non-native phonetic realizations for a Spring School devoted to *Individualized centered approaches to speech processing* [63].

#### 6.2.3.1. Bilingual speech corpus of French and German language learners

We designed a corpus of native and non-native speech for the French-German language pair, with a special emphasis on phonetic and prosodic aspects. To our knowledge there is no suitable corpus, in terms of size and coverage, currently available for this target language pair [9].

We adopted a two step process to create the corpus. Firstly, a bilingual corpus including all sounds of each language and all speech phenomena of potential interest was recorded from a few speakers (14), and analyzed. Its analysis revealed/confirmed: 1) the existence of special strategies due to sentence reading and sentence listening conditions, 2) the importance of recording duration (the recording sessions should not last more than one hour to avoid subjects' fatigue), 3) the frequency and importance of some mispronunciations (voicing problems, erroneous presence (or absence) of /h/ for German (or French) non-native speakers, rhythm ...). Secondly, we specified and collected the final corpus [24], which is focused on the problems revealed by

the preliminary corpus. One hundred speakers (50 French and 50 German speakers), beginners and advanced speakers, recorded 60 sentences in their second language and 30 in their native language, which gave a total amount of about 6000 non-native and 3000 native sentence realizations. The sentences were read in two conditions depending upon whether or not the subjects listen to a reference before producing the sentence. A small text as well as sentences devoted to focus analysis completed the corpus. The data was segmented and labelled at word and phone levels by an automatic alignment algorithm elaborated by our team (cf. 6.4.3.2). The outputs were then manually checked at the levels of phones and words (phonetic transcription) and corrections were made if necessary. In order to check the homogeneity of the corrections made by the seven annotators, phone boundaries were compared with those achieved by a golden annotator on a few sentences using the CoALT tool.

#### 6.2.3.2. *Devoicing of final obstruents by German learners*

We investigated a typical example of L1-L2 interference: the realization of voiced fricatives in final position, where the opposition between voiced and unvoiced consonants is neutralized in German (with a bias towards unvoiced consonants) but not in French. As a consequence, German speakers learning French as a second language often produce unvoiced fricatives in final position instead of the expected voiced consonants. We analysed the production of French voiced fricatives for 40 non-native (beginners and advanced speakers) and 8 native speakers. We measured the ratio of locally unvoiced frames in the consonantal segment and also the ratio of consonantal duration vs. the duration of the preceding vowel. Results showed that the realizations of French fricatives by German speakers varied with speakers, speakers' level and experimental condition (there were two conditions depending on whether or not the subjects listened to a reference before producing the sentence) [23]. As could be expected we observed a continuum between typically voiced and typically unvoiced realizations, and best level speakers tend to produce more typically French realizations. Our next study will concern the perceptual identification of learners' realizations and the link between perceptual answers and acoustic cues values.

### 6.3. Complex statistical modeling of speech

**Participants:** Emmanuel Vincent, Antoine Liutkus, Denis Jouviet, Dominique Fohr, Irina Illina, Joseph Di Martino, Emad Girgis, Arseniy Gorin, Nathan Souviraà-Labastie, Luiza Orosanu, Imran Sheikh, Xabier Jaureguiberry, Baldwin Dumortier.

#### 6.3.1. *Acoustic modeling*

##### 6.3.1.1. *Theory for audio source separation*

Our work on audio source separation was marked by the release of version 2 of our toolbox FASST, which was demonstrated at ICASSP 2014 [65], and by the publication of a review paper about guided audio source separation for *IEEE Signal Processing Magazine* [16]. Audio source separation is an inverse problem, which requires the user to guide the separation process using prior models for the source signals and the mixing filters or for the source spectra and their spatial covariance matrices.

On the topic of the mixing parameters, we studied the impact of sparsity penalties over the mixing filters [8] and deterministic subspace constraints [10] over the spatial covariance matrices.

Modelling the spectra of the sources is a fundamental problem in source separation, that aims at catching their main features while requiring few parameters to estimate. We proposed a new framework called Kernel Additive Modelling (KAM). In contrast to Nonnegative Matrix Factorization approaches (NMF), KAM permits to model sources spectro-temporal evolutions only locally. It generalizes many methods from the state-of-the-art, including REPET (voice/music separation) and HPSS (harmonic/percussive separation) and is the first framework to settle them on principled statistical grounds. This year, we have thus been very active not only in diffusing REPET and its variants to a large audience, notably through the publication of a chapter book on the topic [58], but also by establishing many international collaborations on KAM, leading to the publication of one journal paper in *IEEE TSP* [13] and to two international conference papers [25], [42].



In parallel, we started a new research track on the fusion of multiple source separation techniques. In the specific case when the source spectra are modeled by NMF, the number of components of the NMF is known to have a noticeable influence on separation quality. Many methods have been proposed to select the best order for a given task. To go further, we proposed to use model averaging. As existing techniques do not allow an effective averaging, we introduced a generative model in which the number of components is a random variable and we proposed a modification to conventional variational Bayesian (VB) inference. Initial experiments showed promising results [33], [32].

#### 6.3.1.2. Audio separation based on multiple observations

An interesting scenario for informed audio source separation is when the signals to separate can be observed through deformed references. We proposed a general approach for 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 [46], [66].

A related topic is the removal of interferences from live recordings. In this scenario, there are as many microphones as source signals, but each microphone captures not only its dedicated source, but also some interference from the other ones. We proposed a variant of KAM, called KAM for Interference Removal (KAMIR) that permits to address this scenario. The corresponding study has been achieved in collaboration with New York and Erlangen universities.

#### 6.3.1.3. Separation and dereverberation

In order to complement source separation by dereverberation of the source signals, we devoted some work to the estimation of the reverberation time (RT60). In many situations, the room impulse response (RIR) is not available and the RT60 must be blindly estimated from a speech or music signal. Current methods often implicitly assume that reverberation dominates direct sound, which restricts their applicability to relatively small rooms or distant sound sources. We proposed a blind RT60 estimation method that is independent of the room size and the source distance and showed that the estimation error is significantly reduced even in the case when reverberation dominates [21].

#### 6.3.1.4. Corpora for audio separation

Finally, we pursued our long-lasting efforts on the evaluation of audio source separation by providing more details about the DEMAND dataset, that is the first-ever publicly available dataset of multichannel real-world noise recordings [55]. Furthermore, we have continued our efforts on providing corpora for the evaluation of music source separation methods (notably for music/voice separation) and target at significantly extending the SiSEC corpus in 2015 to several hundreds complete recordings, to be used for the first time at SiSEC 2015.

#### 6.3.1.5. Detailed acoustic modeling

Acoustic models aim at representing the acoustic features that are observed for the sounds of the language, as well as for non-speech events (silence, noise, ...). Currently context-dependent hidden Markov models (CD-HMM) constitute the state of the art for speech recognition. However, for text-speech alignment, simpler context-independent models are used as they provide better performance.

In conventional HMM-based approaches that rely on Gaussian mixture densities (GMM), the Gaussian components are estimated independently for each density. Thus, we have focused recent studies on enriching the acoustic models themselves in view of handling trajectory and speaker consistency in decoding. A new modeling approach was developed that takes advantage of the multiple modeling ideas and involves a sharing of parameters. The idea is to use the multiple modeling approach to partition the acoustic space according to classes (manual classes or automatic classification). Then, for each density, Gaussian components are estimated using the data associated to the classes. These class-based Gaussian components are then pooled to provide the set of Gaussian components of the density. Finally class dependent mixture weights are estimated for each density; such approach allows us to better parameterize GMM-HMM without increasing significantly the number of model parameters. Experiments on French radio broadcast news data demonstrated the improvement of the accuracy with such parameterization compared to models with a similar, or even a larger number of parameters. Another approach has been proposed that combines the structuring of the Gaussian components of the densities with respect to some data classes, with the stranded-based approach

which introduces probabilities for the transitions between the Gaussian components of the densities when moving from one frame to the next. A detailed analysis of stranded GMM was conducted on data containing different types of non-phonetic variability [29]. The combination of stranded GMM with class-structured densities was evaluated on an English connected digits task using adult and child data [27] and for phonetic decoding on a larger French telephone speech database [26]. This approach was later combined with feature normalization [28].

#### 6.3.1.6. Robust acoustic modeling

In the framework of using speech recognition for helping communication with deaf or hard of hearing people, robustness of the acoustic modeling is investigated. Current studies relate to improving robustness with respect to speech signal level and environment noise through multicondition training and enhanced set of acoustic features.

#### 6.3.1.7. Unsupervised acoustic model training

In previous experiments relating to the combination of speech decoder outputs for improving speech recognition performance [4], it was observed that when a forward-based and a backward-based decoder were providing a same word hypothesis, such common word hypothesis is correct in more than 90% of the cases [71]. Hence, we have investigated how such behavior can help for selecting data for unsupervised training of acoustic models. Best performance is achieved when selecting automatically transcribed data (speech segments) that have the same word hypotheses when processed by the Sphinx forward-based and the Julius backward-based transcription systems, and this selection process outperforms confidence measure based selection. Overall, selecting automatically transcribed speech segments that have the same word hypotheses for the two speech transcription systems, and adding this automatically transcribed and selected data to the manually transcribed data leads to significant word error rate reductions on the ESTER2 data (radio broadcast news) when compared to the baseline system trained only on manually transcribed speech data [34].

#### 6.3.1.8. Score normalization

Existing techniques for robust ASR typically compensate distortion on the features or on the model parameters themselves. By contrast, a number of normalization techniques have been defined in the field of speaker verification that operate on the resulting log-likelihood scores. We provided a theoretical motivation for likelihood normalization due to the so-called “hubness” phenomenon and we evaluated the benefit of several normalization techniques on ASR accuracy for the 2nd CHiME Challenge task. We showed that symmetric normalization (S-norm) reduces the relative error rate by 43% alone and by 10% after feature and model compensation [53].

### 6.3.2. Linguistic modeling

#### 6.3.2.1. Out-of-vocabulary proper name retrieval

Recognition of proper names 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. 8.1.4), 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 proper names and increase the ASR vocabulary size. We focus on exploiting the lexical context based on temporal information from diachronic documents. Our assumption is that time is an important feature for capturing name-to-context dependencies. We also studied different metrics for proper name selection in order to limit the vocabulary augmentation: a method based on Mutual Information and a new method based on cosine-similarity measure. Recognition results show a significant reduction of the proper name error rate using augmented vocabulary [30][31].

### 6.3.2.2. Hybrid language modeling

In the framework of using speech recognition for helping communication with deaf or hard of hearing people, the handling of out-of-vocabulary words is a critical aspect. Indeed, the size of the vocabulary is always limited (even if large or very large), and the system is not able to recognize words out of its lexicon. Such words would then be transcribed as sequences of short words which involve similar sounds as the unknown word. However the interpretation of such sequences of small word require a lot of efforts. Hence the idea of combining in a single model a set of words (the most frequent and/or most relevant for the application context) and a set of syllables. With such an approach, unknown words are usually recognized as sequences of syllables which are easier to interpret. By setting different thresholds on the confidence measures associated to the recognized words (or syllables), the most reliable word hypotheses can be identified, and they have correct recognition rates between 70% and 92% [44][45].

### 6.3.2.3. Music language modeling

Similarly to speech, music involves several levels of information, from the acoustic signal up to cognitive quantities such as composer style or key, through mid-level quantities such as a musical score or a sequence of chords. The dependencies between mid-level and lower- or higher-level information can be represented through acoustic models and language models, respectively. We pursued our pioneering work on music language modeling, with a particular focus on the modeling of long-term structure [20]. We also proposed a new Bayesian n-gram topic modeling and estimation technique, which we applied to genre-dependent modeling of chord sequences and to music genre classification [15].

## 6.3.3. Speech generation by statistical methods

### 6.3.3.1. Enhancing pathological voice by voice conversion techniques

Enhancing the pathological voice in order to make it more intelligible would allow persons having this kind of voice to communicate more easily with those around them. In our group we chose to improve the pathological voice by means of voice conversion techniques. Since we began this study, we have succeeded to predict the complete magnitude spectrum. In doing so, we free ourselves from the prediction of the fundamental frequency of speech (F0). Such an interesting result allows us to obtain converted speech of good audio quality. Now in order to obtain perfect conversion, we are trying, with Emad Girgis, a postdoctoral student who began his work in November 2014, to predict the phase spectrum. To achieve this goal, Emad intends to use Deep Neural Networks (DNN). We expect first results in the beginning of 2015.

### 6.3.3.2. Enhancing pathological voice by voice recognition techniques

Another possibility for enhancing the pathological voice is to recognize it. Othman Lachhab, a PhD student, is working on the recognition of the esophageal voice: using high order temporal derivatives combined with an Heteroscedastic Linear Discriminant Analysis (HLDA) he reached an interesting phone recognition rate of 63.59% [36]. Currently Othman, is trying to improve his results by using voice conversion techniques. Using these techniques pathological features are projected in a clean-natural speech feature space, and preliminary results exhibit an increase of 1.70% of the phone recognition rate.

### 6.3.3.3. F0 detection using wavelet transforms

Another possible interesting track for improving voice conversion techniques is to predict the fundamental frequency of speech. For doing so, it is necessary to have a good F0 detector. As part of her thesis, Fadoua Bahja developed many F0 detection algorithms [69] [1]. The latest, using wavelet transform for denoising the cepstrum signal, has been submitted for publication in an international journal.

## 6.4. Uncertainty estimation and exploitation in speech processing

**Participants:** Emmanuel Vincent, Dominique Fohr, Odile Mella, Denis Juvet, Agnès Piquard-Kipffer, Dung Tran.

### 6.4.1. Uncertainty and acoustic modeling

In many real-world conditions, the speech signal is overlapped with noise, including environmental sounds, music, or undesired extra speech. Speech enhancement is useful but insufficient: some distortion remains in the enhanced signal which must be quantified in order not to be propagated to the subsequent feature extraction and decoding stages. The framework of uncertainty decoding assumes that this distortion has a Gaussian distribution and seeks to estimate its covariance matrix [5]. A number of uncertainty estimators and propagators have been proposed for this purpose, which typically operate on diagonal covariance matrices and are based on fixed mathematical approximations or heuristics. We obtained more accurate uncertainty estimates by propagating the full uncertainty covariance matrix and by fusing multiple uncertainty estimators [50], [51]. Overall, we obtained 18% relative 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 and propagator.

In order to motivate further work by the community, we created a new international evaluation campaign on that topic in 2011: the CHiME Speech Separation and Recognition Challenge [2]. After two successful editions in 2011 and 2013, we started working and collecting a new corpus towards the organization of a third edition to be announced in 2015.

### 6.4.2. 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. 8.1.5), our goal is to find the best way for displaying the speech transcription results. 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 (cf. 6.3.2.2); hence such information can be used to adjust the display of the transcription results.

We have adopted a two-step process. Firstly, we conducted a feasibility study with three hard-of-hearing persons including written display tests on print media and interviews. Secondly, we set up an experimental protocol with five hard-of-hearing persons. It included comprehension tests of 40 written sentences recorded by a French native speaker video projected onto a screen. We have also conducted parallel interviews. Their analysis revealed: (1) the interest of the participants in the project; (2) their difficulties to read International Phonetic Alphabet; (3) the importance of knowing the context of communication; (4) the need for aid in case of errors of the speech recognition system by emphasizing the words that are supposed to be well recognized by the system. At this stage of the experimental period, the best display associates writing in a bold spelling the words that are supposed to be correctly recognized, and writing in a normal font using simplified French phonetics the words that are possibly wrongly recognized (according to their confidence measure). The next step will be to set up another experimental protocol in order to compare the current display in three conditions (written sentences vs written sentences with oral and lip reading vs lip reading only).

### 6.4.3. Uncertainty and phonetic segmentation

As described below, phonetic segmentation has been studied this year for spontaneous speech and non-native speech. Moreover, some portions (of about 30 secondes) of various speech documents have been manually annotated (checking and correction of an automatic segmentation). In the future this manually annotated data will be used to analyze the accuracy of the automatic segmentation, and also to elaborate measures that estimate the quality of the segmentation.

#### 6.4.3.1. Alignment with spontaneous speech

Within the ANR ORFEO project (cf. 8.1.2), we addressed the problem of the alignment of spontaneous speech. The ORFEO audio files 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. For selecting the best acoustic models, we compared the alignment outputs obtained with the different acoustic models by using our tool CoALT and the manually annotated portions described above.

We also designed a new automatic grapheme-phoneme tool to generate the potential pronunciations of words and proper names. For what concerns overlapping speech, among the different orthographic transcripts corresponding to the overlapping area, we determined as the main transcript the one that best matches the audio signal, the others are kept in other tiers (in a Praat TextGrid file) with the same time boundaries.

#### *6.4.3.2. Alignment with non-native speech*

Non-native speech alignment with text is one critical step in computer assisted foreign language learning [3]. The alignment is necessary to analyze the learner's utterance, in view of providing some prosody feedback (as for example bad duration of some syllables). However, non-native speech alignment with text is much more complicated than native speech alignment. This is due to the pronunciation deviations observed on non-native speech, as for example the replacement of some target language phonemes by phonemes of the mother tongue, as well as errors in the pronunciations. Non-native speech alignment with text is currently studied in the ANR IFCASL project (see 8.1.3 ).

#### **6.4.4. Uncertainty and prosody**

A statistical analysis was conducted on a large annotated speech corpus to investigate the links between punctuation and automatically detected prosodic structures. The speech data comes from radio broadcast news and TV shows, that were manually annotated during French speech transcription evaluation campaigns. These corpora contain more than 3 million words and almost 350,000 punctuation marks. The detection of the prosodic boundaries and of the prosodic structures is based on an automatic approach that integrates little linguistic knowledge and mainly uses the amplitude and the direction of the F0 slopes, as well as phone durations. A first analysis of the occurrences of the punctuation marks, with respect to various sub-corpora, has highlighted the variability among annotators. Then, a detailed analysis of the prosodic parameters with respect to the punctuation marks, whether followed or not by a pause, and of the links between the automatically detected prosodic structures and the manually annotated punctuation marks was conducted [18].

## PANAMA Project-Team

# 6. New Results

## 6.1. Highlights of the Year

The EUSIPCO 2014 Best Student Paper Award was awarded to our joint paper [32] on dynamic screening for sparse regularization. .

A review paper on audio source separation, rooted in METISS/PANAMA know-how and contributions to this topic over the years, was published in the IEEE Signal Processing Magazine [25].

A new version of the Flexible Audio Source Separation Toolbox (FASST) was released in January 2014 and downloaded 300 times.

BEST PAPERS AWARDS :

[32] **European Signal Processing Conference EUSIPCO 2014**. A. BONNEFOY, V. EMIYA, L. RALAIVOLA, R. GRIBONVAL.

## 6.2. 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 8.2.1.1 ).

### 6.2.1. A new framework for sparse representations: analysis sparse models

**Participants:** Rémi Gribonval, Nancy Bertin, Srdan Kitic, Cagdas Bilen, 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. Within the SMALL FET-Open project, we initiated a research programme dedicated to this analysis model, in the context of a generic missing data problem (e.g., compressed sensing, inpainting, source separation, etc.). We obtained a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. We also considered a number of pursuit algorithms for solving the missing data problem, including an  $\ell^1$ -based and a new greedy method called GAP (Greedy Analysis Pursuit). Our simulations demonstrated the appeal of the analysis model, and the success of the pursuit techniques presented.

These results have been published in conferences and in a journal paper [100]. Other algorithms based on iterative cosparse projections [83] as well as extensions of GAP to deal with noise and structure in the cosparse representation have been developed, with applications to toy MRI reconstruction problems and acoustic source localization and reconstruction from few measurements [101].

Successful applications of the cosparse approach to sound source localization, audio declipping and brain imaging have been developed this year. In particular, we compared the performance of several cosparse recovery algorithms in the context of sound source localization [97] and showed its efficiency in situations where usual methods fail ([37], see paragraph 6.6.3 ). It was also shown to be applicable to the hard declipping problem [49]. Application to EEG brain imaging was also investigated and a paper was published in MLSP14 [28] (see paragraph 6.6.4 ).

### 6.2.2. Theoretical results on sparse representations

**Participants:** Rémi Gribonval, Anthony Bourrier, Pierre Machart, Yann Traonmilin, Gilles Puy.



*Main collaboration: Mike Davies (University of Edinburgh), Patrick Perez (Technicolor R&I France), Tomer Peleg (The Technion)*

**Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems:** The primary challenge in linear inverse problems is to design stable and robust "decoders" to reconstruct high-dimensional vectors from a low-dimensional observation through a linear operator. Sparsity, low-rank, and related assumptions are typically exploited to design decoders whose performance is then bounded based on some measure of deviation from the idealized model, typically using a norm. We characterized the fundamental performance limits that can be expected from an ideal decoder given a general model, i.e., a general subset of "simple" vectors of interest. First, we extended the so-called notion of instance optimality of a decoder to settings where one only wishes to reconstruct some part of the original high-dimensional vector from a low-dimensional observation. This covers practical settings such as medical imaging of a region of interest, or audio source separation when one is only interested in estimating the contribution of a specific instrument to a musical recording. We defined instance optimality relatively to a model much beyond the traditional framework of sparse recovery, and characterized the existence of an instance optimal decoder in terms of joint properties of the model and the considered linear operator [106], [105]. This year, noiseless and noise-robust settings were both considered in the journal paper [16]. We showed somewhat surprisingly that the existence of noise-aware instance optimal decoders for all noise levels implies the existence of a noise-blind decoder. A consequence of our results is that for models that are rich enough to contain an orthonormal basis, the existence of an  $L_2/L_2$  instance optimal decoder is only possible when the linear operator is not substantially dimension-reducing. This covers well-known cases (sparse vectors, low-rank matrices) as well as a number of seemingly new situations (structured sparsity and sparse inverse covariance matrices for instance). We exhibit an operator-dependent norm which, under a model-specific generalization of the Restricted Isometry Property (RIP), always yields a feasible instance optimality and implies instance optimality with certain familiar atomic norms such as the  $\ell^1$  norm. Current work explores the existence of convex decoders for general union of subspaces models under generalized RIP assumptions, as well as conditions ensuring that random low-dimensional projections ensure the RIP even when the projection is from an infinite-dimensional space to a finite-dimensional one. Envisioned applications are in compressive learning (see Section 6.4).

**Connections between sparse approximation and Bayesian estimation:** Penalized least squares regression is often used for signal denoising and inverse problems, and is commonly interpreted in a Bayesian framework as a Maximum A Posteriori (MAP) estimator, the penalty function being the negative logarithm of the prior. For example, the widely used quadratic program (with an  $\ell^1$  penalty) associated to the LASSO / Basis Pursuit Denoising is very often considered as MAP estimation under a Laplacian prior in the context of additive white Gaussian noise (AWGN) reduction.

In 2011 we obtained a result [85] highlighting the fact that, while this is *one* possible Bayesian interpretation, there can be other equally acceptable Bayesian interpretations. Therefore, solving a penalized least squares regression problem with penalty  $\phi(x)$  need not be interpreted as assuming a prior  $C \cdot \exp(-\phi(x))$  and using the MAP estimator. In particular, we showed that for *any* prior  $P_X$ , the minimum mean square error (MMSE) estimator is the solution of a penalized least square problem with some penalty  $\phi(x)$ , which can be interpreted as the MAP estimator with the prior  $C \cdot \exp(-\phi(x))$ . Vice-versa, for *certain* penalties  $\phi(x)$ , the solution of the penalized least squares problem is indeed the MMSE estimator, with a certain prior  $P_X$ . In general  $dP_X(x) \neq C \cdot \exp(-\phi(x))dx$ . In 2013, we extended this result to general inverse problems [88], [86], [87]. This year, we worked on the characterization of such relations beyond the Gaussian noise model, with the objective of understanding whether similar results hold when the quadratic data-fidelity term is replaced with other convex losses.

### 6.2.3. Algorithmic and theoretical results on dictionary learning

**Participants:** Rémi Gribonval, Nancy Bertin, Srđan Kitic, Cagdas Bilen, Luc Le Magoarou, Melanie Ducoffe.

*Main collaboration (theory for dictionary learning): Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (Inria, Paris)), Martin Kleinstueber, Matthias Seibert (TU-Munich),*

*Main collaboration (dictionary learning for gesture recognition): Anatole Lecuyer, Ferran Argelaguet (EPI HYBRID, Rennes)*

**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 literature 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 [89], [109] [6] on this topic, which concentrated on the noiseless case for non-overcomplete dictionaries, this year we obtained new results showing the relevance of an  $\ell^1$  penalized cost function for the locally stable identification of overcomplete incoherent dictionaries, in the presence of noise and outliers [54]. Moreover, we established new sample complexity bounds of dictionary learning and other related matrix factorization schemes (including PCA, NMF, structured sparsity ...) [55], [46], [38].

**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 [50]. A preprint is available [56], a paper will appear at ICASSP 2015, and a journal paper is in preparation.

**Operator learning for cospase representations :** Besides standard dictionary learning, we also considered learning in the context of the cospase 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 cospase analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete analysis operator.

In specific situations, when prior information is available on the operator, it is possible to express it in parametric form and learn this parameter. For instance, in the sound source localization problem, we showed that the unknown speed of sound can be learned jointly in the process of cospase recovery, under mild conditions. This work was presented at iTWIST'14 workshop [48].

**Dictionary learning for gesture modeling** In collaboration with the HYBRID project-team (internship of Melanie Ducoffe), we explored the potential of dictionary learning in the context of motion tracking. Motion tracking technology, especially for commodity hardware, requires robust gesture recognition algorithms to fully exploit the benefits of natural user interfaces. We proposed a gesture recognition algorithm based on the sparse representation of motion data, with a learning phase consisting in learning a dictionary of basic gestures. A paper is in preparation.

### 6.3. Activities on waveform design for telecommunications

**Participant:** Rémi Gribonval.

*Main collaboration: Marwa Chafii, Jacques Palicot, Carlos Bader (Equipe SCEE, Supelec, Rennes)*

Peak to Average Power Ratio (PAPR), Orthogonal Frequency Division Multiplexing (OFDM), Generalized Waveforms for Multi Carrier (GWMC)

In the context of the TEPN (Towards Energy Proportional Networks) Comin Labs project (see Section 8.1.1.2), 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 [17] and several conference communications [27], [33]. The work now concentrates on characterizing waveforms with optimum PAPR.

## 6.4. Emerging activities on compressive sensing, learning and inverse problems

Compressive sensing, compressive learning, acoustic wavefields, audio inpainting,

### 6.4.1. Audio inpainting

**Participants:** Rémi Gribonval, Nancy Bertin, Corentin Guichaoua, Srđan Kitić, Anh Tho Le.

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. It consists in reconstructing a set of missing pixels in an image based on the observation of the remaining pixels. Sparse representations have proved to be particularly appropriate to address this problem. However, inpainting audio data has never been defined as such so far. A series of works about audio inpainting was initiated by the METISS team in the framework of the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL (Sparse Models, Algorithms and Learning for Large-Scale data).

Building upon our previous 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 [60]), new results were obtained this year to address the case of audio declipping with the competitive cosparsity approach. Its promising results, especially when the clipping level is low, were confirmed experimentally by the formulation and use of a new algorithm named Cosparsity Iterative Hard Thresholding, which is a counterpart of the sparse Consistent Iterative Hard Thresholding. These results were presented during the iTwist'14 workshop [49]. Additional experiments were performed (internship of Anh Tho Le) to confirm the results on a larger database and investigate optimal parameters (nature and redundancy of the dictionary, relaxation parameter for the cosparsity level).

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).

### 6.4.2. Blind Calibration of Compressive Sensing systems

**Participants:** Rémi Gribonval, Cagdas Bilen, Gilles Puy.

*Main collaborations: Gilles Chardon, Laurent Daudet (Institut Langevin)*

We consider the problem of calibrating a compressed sensing measurement system under the assumption that the decalibration consists of unknown gains on each measure. We focus on blind calibration, using measures performed on a few unknown (but sparse) signals. A naive formulation of this blind calibration problem, using  $\ell^1$  minimization, is reminiscent of blind source separation and dictionary learning, which are known to be highly non-convex and riddled with local minima. In the considered context, when the gains are real valued and non-negative, we showed that in fact this formulation can be exactly expressed as a convex optimization problem, and can be solved using off-the-shelf algorithms. Numerical simulations demonstrated the effectiveness of the approach even for highly uncalibrated measures, when a sufficient number of (unknown, but sparse) calibrating signals is provided. We observed that the success/failure of the approach seems to obey sharp phase transitions [84]. We extended the framework to phase-only decalibration, using techniques revolving around low-rank matrix recovery [66], [65], [110], [64], and to joint phase and gain decalibration [15].

### 6.4.3. Compressive Gaussian Mixture estimation

**Participants:** Rémi Gribonval, Anthony Bourrier, Nicolas Keriven.

*Main collaborations: Patrick Perez (Technicolor R&I France)*

When fitting a probability model to voluminous data, memory and computational time can become prohibitive. In this paper, 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 [71], [69], [70]. This year, extensions to non-isotropic Gaussians, with new algorithms and preliminary applications to speaker verification have been conducted.

## 6.5. Recent results on tensor decompositions

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 back as far as Jordan who was interested in simultaneously diagonalizing two matrices at a time [96]. 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 [92] 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 exercises 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 [114]. 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 [91], and on the means to achieve such a decomposition, which will become the famous canonical decomposition [74]. In honor to Hitchcock's pioneer 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 [98]. Consequently, a new class of efficient algorithms able to take into account the properties of tensors to be decomposed is needed.

### 6.5.1. CP decomposition of semi-symmetric semi-nonnegative three-way arrays

**Participant:** Laurent Albera.

*Main collaboration (Line search and trust region strategies):* Julie Coloigner (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France)

*Main collaborations (Jacobi-like approaches):* Lu Wang (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France), Huazhong Shu (LIST, China)

We proposed new algorithms for the CP decomposition of semi-nonnegative semi-symmetric three-way tensors. In fact, it consists in fitting the CP model for which two of the three loading matrices are nonnegative and equal. Note that such a problem can also be interpreted as a nonnegative Joint Diagonalization by Congruence (JDC) problem.

#### **Line search and trust region strategies**

We first circumvented the nonnegativity constraint by means of changes of variable into squares, leading to a (polynomial) unconstrained optimization problem. Two optimization strategies, namely line search and trust region, were then studied. Regarding the former, a global plane search scheme was considered. It consists in computing, for a given direction, one or two optimal stepsizes, depending on whether the same stepsize is used in various updating rules. Moreover, we provided a compact matrix form for the derivatives of the objective function. This allows for a direct implementation of several iterative algorithms such as Conjugate Gradient (CG), Levenberg-Marquardt (LM) and Newton-like methods, in matrix programming environments like MATLAB. Note that the computational complexity issue was taken into account in the design phase of the algorithms, and was evaluated for each algorithm, allowing to fairly compare their performance.

Thus, various scenarios have been considered, aiming at testing the influence of i) an additive noise, which can stand for modeling errors, ii) the collinearity between factors, iii) the array rank and iv) the data size. The comparisons between our CG-like, Newton-like and LM-like methods (where semi-nonnegativity and semi-symmetry constraints are exploited), and classical CP algorithms (where no constraints are considered), showed that a better CP decomposition is obtained when these a priori are exploited, especially in the context of high dimensions and high collinearity. Finally, based on our numerical analysis, the algorithms that seem to yield the best tradeoff between accuracy and complexity are our  $CG_{2\text{steps}}$ -like and LM-like algorithms. This work was published in the Elsevier Linear Algebra and Applications journal [19].

Next, we considered an exponential change of variable leading to a different (non-polynomial) unconstrained optimization problem. Then we proposed novel algorithms based on line search strategy with an analytic global plane search procedure requiring new matrix derivations. Their performance was evaluated in terms of estimation accuracy and computational complexity. The classical ELS-ALS [108] and LM [112] algorithms without symmetry and nonnegativity constraints, and the ACDC algorithm [115] where only the semi-symmetry constraint is imposed, were tested as reference methods. Furthermore, the performance was also compared with our algorithms based on a square change of variable. The comparison studies showed that, among these approaches, the best accuracy/complexity trade off was achieved when an exponential change of variable was used through our ELS-ALS-like algorithm. This work was published in the Elsevier Signal Processing journal [18].

#### **Jacobi-like approaches**

The line search (despite the use of global plane search procedures) and trust region strategies may be sensitive to initialization, and generally require a multi-initialization procedure. In order to circumvent this drawback, we considered in this work Jacobi-like approaches, which are known to be less sensitive to initialization. Note that our line search and trust region approaches can then be used to refine the solution obtained by the latter.

More particularly, we formulated the high-dimensional optimization problem into several sequential polynomial and rational subproblems using i) a square change of variables to impose nonnegativity and ii) LU or QR matrix factorization for parameterization. The two equal nonnegative loading matrices are actually written as the Hadamard product of two equal matrices which can be factorized as the product of elementary matrices, each one depending on only one parameter.

The proposed approach reduces the optimization problem to the computation of the two equal nonnegative loading matrices only. The third loading matrix is algebraically derived from the latter. This requires an appropriate parameterization of the set of matrices whose inverse is nonnegative. This work was published in a journal paper [26]. Numerical experiments on simulated matrices emphasize the advantages of the proposed algorithms over classical CP and JDC techniques, especially in the case of degeneracies.

## **6.6. Source separation and localization**

Source separation, sparse representations, tensor decompositions, semi-nonnegative independent component analysis, probabilistic model, source localization

### **6.6.1. A general framework for audio source separation**

**Participants:** Frédéric Bimbot, Rémi Gribonval, Nancy Bertin.

*Main collaboration:* E. Vincent, Y. Salaün (EPI PAROLE, Inria Nancy); A. Ozerov, N.Q.K. Duong (Technicolor R&I France)

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 [102], 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 [81], [95], [94], knowledge on the source positions and room characteristics [82], or a better initialization of parameters thanks to specific source localization techniques [68].

This accumulated progress lead to two main achievements this year, which show the maturity of our work and which will leverage its impact. First, a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released. It will provide the community with an efficient and ergonomic software, making available the tools from past years' research [58]. Second, 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 [25].

### 6.6.2. Towards real-world separation and remixing applications

**Participants:** Nancy Bertin, Frédéric Bimbot, Jules Espiau de Lamaestre, Anaik Olivero, Jérémy Paret, Nathan Souviraà -Labastie.

*Emmanuel Vincent (EPI PAROLE, Inria Nancy)*

While some challenges remain, work from previous years and our review paper on guided source separation [25] 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 is exemplified by the contract with MAIA Studio, which reaches 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 this context, new tools were developed this year. The introduction of manually-tuned parameters in the automated separation process, which modifies the Wiener filtering coefficients obtained from estimation of the mixtures covariance matrices, allows to find a better trade-off between artifacts and interferences. In order to ensure high audio quality for such applications, some user-guided corrections remain necessary even after an automatic pre-separation ; to this end, we developed an improved display (based on cepstrum and automatic contrast adaptation) and semi-automatic selection and suppression tools in the time-frequency domain. Those tools take as few inputs as possible from the user and their result can be ergonomically adjusted from the baseline output to a manually fine-tuned area, in a very small operating time. We also proposed tools to suppress a time-frequency area and replace it by content extracted from its context, reducing the perceptual impact of the suppression.

In some applicative contexts of source separation, several mixtures are available which contain similar instances of a given source. We have designed a general framework for audio source separation guided by multiple audio references, where each audio reference is a 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 general manner. A nonnegative *matrix co-factorization* algorithm is used which allows sharing of information between the considered mixtures. We have experimented our algorithm on music plus voice mixtures with music and/or voice references. Applied on movies and TV series data, the algorithm improves the signal-to-distortion ratio (SDR) of the sources of lowest intensity by 9 to 12 decibels with respect to original mixture [40]

### 6.6.3. Acoustic source localization

**Participant:** Nancy Bertin, Srđan Kitić, Laurent Albera, Nancy Bertin, Rémi Gribonval.



*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)* [68]. 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 (cospase modeling of the acoustic field).

#### **Implicit localization through audio-based control for robotics**

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 this 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 [76].

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. 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 submitted to ICRA 2015. Future work will include real-world experiments with the robot Romeo from Aldebaran Robotics.

#### **Cospase modeling of the acoustic field**

Cospase modeling is particularly attractive when the signals of interest satisfy certain physical laws that naturally drive the choice of an analysis operator, which is the case for the acoustic field, ruled by the wave equation. Unlike usual localization techniques such as beamforming or TDOA-based direction estimation, which generally consider reverberation as an adverse condition, the cospase modeling of sound propagation has also the interest of considering reverberation as a source of additional information for the localization task. Eventually, it can provide a full coordinate localization of the sources, and not only their direction of arrival.

Building upon our previous results on cospase modeling and recovery algorithms for the wave equation [97], we have obtained additional evidence of the interest of this approach. In particular, we have showed that recasting source localization as a cospase inverse problem allows to scale up to 3-dimensional problems which were untractable with the counterpart sparse approach. Moreover, we have confirmed that our model takes indeed advantage of reverberation, by showing that localization remains possible when the sources and the microphones are partly separated by an acoustically opaque obstacle (a situation where TDOA would obviously fail). These two results were published and presented during ICASSP'14 [37]. Recent results also include algorithmic improvements (through the use of the Alternating Direction Method of Multipliers (ADMM) framework), and evidence that, in addition to its scaling capabilities, the sparse analysis computational cost can even *benefit* from an increase in the number of measurements. A journal paper including these new results and presenting them jointly with co-space modeling in the context of brain source localization (see Section 6.6.4) is under preparation.

#### **6.6.4. Brain source localization**

**Participants:** Laurent Albera, Srdan Kitic, Nancy Bertin, Rémi Gribonval.

*Main collaborations (tensor-based approaches): Hanna Becker (GIPSA & LTSI, France), Isabelle Merlet (LTSI, France), Fabrice Wendling (LTSI, France), Pierre Comon (GIPSA, France), Christian Benar (La Timone, Marseille), Martine Gavaret (La Timone, Marseille), Gwénaél Birot (FBML, Genève), Martin Haardt (TUI, Germany)*

*Main collaborations (from tensor to sparse models): Hanna Becker (GIPSA & LTSI, France), Pierre Comon (GIPSA, France), Isabelle Merlet (LTSI, France), Fabrice Wendling (LTSI, France)*

*Main collaborations (a sparsity-based approach): Hanna Becker (Technicolor, France), Pierre Comon (GIPSA, France), Isabelle Merlet (LTSI, France)*

*Main collaborations (a multimodal sparsity-based approach): Thomas Oberlin, Pierre Maurel, Christian Barillot (EPI VISAGES, Rennes, France)*

### **Tensor-based approaches**

The localization of several simultaneously active brain regions having low signal-to-noise ratios is a difficult task. To do this, tensor-based preprocessing can be applied, which consists in constructing a Space-Time-Frequency (STF) or Space-Time-Wave-Vector (STWV) tensor and decomposing it using the CP decomposition. We proposed a new algorithm for the accurate localization of extended sources based on the results of the tensor decomposition. Furthermore, we conducted a detailed study of the tensor-based preprocessing methods, including an analysis of their theoretical foundation, their computational complexity, and their performance for realistic simulated data in comparison to three conventional source localization algorithms, namely sLORETA [104], cortical LORETA (cLORETA) [103], and 4-ExSo-MUSIC [67]. Our objective consisted, on the one hand, in demonstrating the gain in performance that can be achieved by tensor-based preprocessing, and, on the other hand, in pointing out the limits and drawbacks of this method. Finally, we validated the STF and STWV techniques on real epileptic measurements to demonstrate their usefulness for practical applications. This work was published in the Elsevier NeuroImage journal [13].

### **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. A part of this work was presented at ICASSP'14 [30] while the whole part was submitted to IEEE Signal Processing Magazine.

### **A cosparsity-based approach**

Cosparsity modeling is particularly attractive when the signals of interest satisfy certain physical laws that naturally drive the choice of an analysis operator. We showed how to derive a reduced non-singular analysis operator describing EEG signals from Poisson's equation, Kirchhoff's law and some other physical constraints. As a result, we proposed the CoRE (Cosparsity Representation of EEG signals) method to solve the classical brain source imaging problem. Computer simulations demonstrated the numerical performance of the CoRE method in comparison to a dictionary-based sparse approach. This work was partially presented at MLSP'14 [28].

### **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 [30], the VB-SCCD algorithm [79], 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 L1,2-SVB-SCCD, 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 [29] and a journal paper is in preparation.

#### **A multimodal sparsity-based approach**

In the context of the HEMISFER Comin Labs project (see Section 8.1.1.1), in collaboration with the VISAGES team, we investigated brain imaging using simultaneously recorded electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). To this end, we introduced a linear coupling model that links the electrical EEG signal to the hemodynamic response from the blood-oxygen level dependent (BOLD) signal. Both modalities are then symmetrically integrated, to achieve a high resolution in time and space while allowing some robustness against potential decoupling of the BOLD effect. The joint imaging problem is expressed as a linear inverse problem, which is addressed using sparse regularization. The sparsity prior naturally reflects the fact that only few areas of the brain are activated at a certain time, and it is easily implemented through proximal algorithms. At this stage, the significance of the method and its effectiveness have been demonstrated through numerical investigations on a simplified head model and simulated data on a realistic brain model. A conference paper has been submitted and a journal paper is in preparation.

### **6.6.5. 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 (DeL<sup>R</sup>) [78] 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 [93], RobustICA [116] and DeL<sup>R</sup> algorithms, show that the proposed algorithm offers the best trade-off between performance and numerical complexity. This work was submitted for publication in the IEEE Signal Processing Letters journal.

### **6.6.6. 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], [75], [72], radiocommunications [77] and document restoration [113] to cite a few.

For instance in [113], 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 [80] 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 [111]. Moreover, using the JADE algorithm [73] 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 [72] 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 [80], the thickness of each organ, which stands for the mixing coefficient, is real positive. Furthermore, reflectance indices in [113] 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 [90]; the mixing matrix is again nonnegative.

By imposing the nonnegativity of the mixing matrix within the ICA process, we shown 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 [99] 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 6.5.1. Hence we use the new methods described in section 6.5.1 to perform semi-nonnegative ICA. Performance results in audio and biomedical engineering were given in the different papers cited in section 6.5.1.

## 6.7. Audio and speech content processing

Audio segmentation, speech recognition, motif discovery, audio mining

### 6.7.1. Audio motif discovery

**Participants:** Frédéric Bimbot, Nathan Souviraà -Labastie.

*This work was performed in close collaboration with Guillaume Gravier from the Limkmedia project-team.*

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.

We have designed a system to create audio thumbnails of spoken content, i.e., short audio summaries representative of the entire content, without resorting to a lexical representation. As an alternative to searching for relevant words and phrases in a transcript, unsupervised motif discovery is here used to find short, word-like, repeating fragments at the signal level without acoustic models. The output of the word discovery algorithm is exploited via a maximum motif coverage criterion to generate a thumbnail in an extractive manner. A limited number of relevant segments are chosen within the data so as to include the maximum number of motifs while remaining short enough and intelligible.

Evaluation has been performed on broadcast news reports with a panel of human listeners judging the quality of the thumbnails. Results indicate that motif-based thumbnails stand between random thumbnails and ASR-based keywords, however still far behind thumbnails and keywords humanly authored [35].

### 6.7.2. *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 is involved in the design of algorithms for the analysis and monitoring of the acoustic scene around the user, yielding information which can be fused with other sources of information (physiological, contextual, etc...) in order to trigger an alarm when needed and subsequent appropriate measures.

This ongoing work is focused on the development of robust techniques for audio scene analysis, including statistical classification of audio segments into threat vs non-threat categories, and the use of spatial information to determine the location of the user with respect to the potential threat.

## 6.8. Music Content Processing and Music Information Retrieval

Acoustic modeling, non-negative matrix factorisation, music language modeling, music structure

### 6.8.1. *Music structure representation and decomposition*

**Participants:** Frédéric Bimbot, Corentin Guichaoua, Anaik Olivero.

*Main collaboration: E. Vincent (EPI PAROLE, Inria Nancy), E. Deruty (external consultant)*

Interest has been steadily growing in semantic audio and music information retrieval for the description of music structure, i.e. the global organization of music pieces in terms of large-scale structural units. Our group has defined a detailed methodology for the semiotic description of music structure, based on concepts and criteria which are formulated as generically as possible, i.e. without resorting to intrinsic properties of the musical content, but rather on relationships between musical elements resulting in well-identifiable patterns. The essential principles and practices developed during an annotation effort deployed by our research group on audio data, in the context of the Quaero project, has led to the public release of over 380 annotations of pop songs from three different data sets (<http://musicdata.gforge.inria.fr/structureAnnotation.html>) documented by a technical report which includes a few case studies and a concise statistical overview of the annotated data [31]. From the algorithmic point of view, we are currently exploring tree-based representation of music structure where a sequence of musical elements and their relationships are modeled hierarchically as the derivation of a context-free grammar. Parsimony criteria and specific cost functions adapted to music patterns are used to learn the grammar rules and the possibility of distorting the rules is introduced to account for variability across different repetitions of musical segments.

## SEMAGRAMME Project-Team

# 6. New Results

## 6.1. Generation

G-TAG [52], [66] is a Tree Adjoining Grammar (TAG) based formalism which was specifically designed for the task of text generation. Contrary to TAG, the derivation structure becomes primary, as pivot between the conceptual representation and the surface form. This is a shared feature with the encoding of TAG into ACG. Laurence Danlos (Alpage Inria project), Aleksandre Maskharashvili, and Sylvain Pogodalla have shown how to recast the G-TAG formalism into ACG, relying on the reversibility properties of the later [17], [16], [18].

## 6.2. Discourse Grammar

Laurence Danlos (Alpage Inria project), Aleksandre Maskharashvili, and Sylvain Pogodalla have presented a method to interface a sentential grammar and a discourse grammar. It offers both a smooth integration of the two grammars without using an intermediate processing step, and the possibility to build discourse structures that are direct acyclic graphs (DAG) and not only trees. The analysis is based on a Tree-Adjoining Grammar (TAG) approach to discourse: Discourse Synchronous TAG (D-STAG) [50], [51], and uses an encoding of TAG into ACG. This allows for expressing a higher-order semantic interpretation that enables building DAG discourse structures, and for smoothly integrating the sentential and the discourse grammar thanks to the modular capability of ACG. All the examples may be run and tested with the the ACGtk (submitted).

## 6.3. Large Scale Grammatical Resources

Guy Perrier wrote a complete documentation [36] on FRIGRAM<sup>0</sup> a French grammar with a large coverage, written in the formalism of Interaction Grammars [59]. The different chapters of the 257 pages of documentation correspond to the different parts of speech in French. At the end, two chapters are dedicated to two specific phenomena: extraction (relative, interrogative and cleft clauses) and coordination, which is presented in common with punctuation because of their proximity.

## 6.4. Deep Syntax Annotation of the Sequoia French Treebank

Marie Candito, Guy Perrier, Bruno Guillaume, Corentin Ribeyre, Karën Fort, Djamé Seddah and Eric de la Clergerie annotated the Sequoia French Treebank with deep syntax dependencies [14].

The Sequoia French Treebank [47] is a 3.100 sentences treebank covering several domains (news, medical, europarl and fr-wikipedia). It is freely available and has already been annotated with surface dependency representations.

The participants in the project have defined a deep syntactic representation scheme for French, built from the surface annotation scheme of the Sequoia corpus and abstracting away from it [28]. 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 goal is to obtain a freely available corpus, which will be useful for corpus linguistics studies and for training deep analyzers to prepare semantic analysis.

The different steps of the annotation process were conducted in a collaborative way. As the members of the project are located in two different French towns (Paris and Nancy), they decided to produce a complete annotation of the TreeBank in both towns and to collaboratively adjudicate the two results.

---

<sup>0</sup><http://wikilligramme.loria.fr/doku.php/frig:frig>



Each team separately produced an initial annotated version of the mini reference. The final version, resulting from several iterations and adjudications, is available <sup>0</sup>.

## 6.5. Exploitation of the LVF (Lexicon of French Verbs)

Bruno Guillaume, Karèn Fort, Guy Perrier and Paul Bédaride have worked on the LVF [53] ("Lexique des Verbes du Français", Lexicon of French Verbs). This large lexicon was built by two French linguists, Jean Dubois and Françoise Dubois-Charlier and contains detailed linguistic information about 12.308 lemmas of French verbs. The work presented in [21] describes experiments aiming at mapping the LVF to DICOVALENCE [87]. The two resources (LVF and DICOVALENCE) were built by linguists, based on very different theories, which makes a direct mapping nearly impossible. In the current work, we focus on the linguistic examples given in LVF. These examples are not sentences that can be parsed directly; the first part of the work was to express examples as real natural language sentence. It is then possible to use FRILEX, a Natural Language Processing lexicon based on DICOVALENCE to parse corrected examples given with LVF entries. This results in an automatic partial mapping of LVF entries against DICOVALENCE entries.

## 6.6. Game With A Purpose

Crowdsourcing is nowadays a way of constructing linguistic resources which is more and more used. In the crowdsourcing area, one of the way to motivate a large amount of people to contribute to a project is to present it as a game. Games used in this particular way are called GWAPs (Game With A Purpose).

In Natural Language Processing, examples of GWAP are "Phrase detective" where games are asked to resolve anaphora in English texts and "JeuDeMots" where gamers have to give lexical terms related to a term given by the system (the goal is to build a semantic networks of French lexical items).

Karèn Fort and Bruno Guillaume worked on the definition of a GWAP to help construction of syntactically annotated corpora. With a student (Hadrien Chastant), they presented in April, the design of ZombiLingo [20], a GWAP that allows for the dependency syntax annotation of French corpora. The main aspects of this work are to explain: how to deal with the complexity of the task, how to motivate gamers to contribute and how to ensure that a large numbers of gamers will help to produce an high quality linguistic resource.

With another student (Valentin Stern), a first prototype was built. This first version implements only a part of the mechanisms described in the previous work and it is used as a proof-of-concept of a future game. This prototype was presented at the TALN conference in July [27].

## 6.7. Supertagging

Guillaume Bonfante, Bruno Guillaume, Mathieu Porey and Guy Perrier wrote a book chapter [30] "Supertagging with constraints". This chapter makes a survey of the results obtained in previous publications about supertagging based on polarities [46] and based on the companionship principle [45]. The last section of the chapter presents a new application of the companionship principle to the TAG formalism and presents some experimental results.

## 6.8. Modelling Semantic Phenomena

Despite the valuable insights yielded by the classical theories of discourse semantics, there is a wide range of exceptional phenomena that they fail to address, e.g., anaphora under double negation and modality. Concentrating on these two exceptions, Sai Qian, Philippe de Groote and Maxime Amblard provide a corresponding adaptation of TTDL for each case. Briefly speaking, for the problem of double negation, they propose to encapsulate both the affirmative representation and the negative representation of an expression in its semantics. Negation is treated as an operation which switches the positions of the two representations. Thus a second negation will switch the positions again as if no negation had ever occurred. In this way, a double negation can be eliminated and the desired referent accessibility is modelled. As for anaphora under modality,

<sup>0</sup><https://deep-sequoia.inria.fr>

they propose to enrich the TTDL left context with the notion of modal base, which is proposed by Kratzer. The possible world model is integrated in the semantic representation as well. Moreover, they show how the different adaptations could work in an unified framework, [75].

## 6.9. Quantification in event semantics

Yoad Winter (Utrecht University) has given a type-logical account of quantification in event semantics.

It has been observed in the literature that Davidson's event semantics does not combine smoothly with Montague's compositional semantics. The difficulty comes from a possibly bad interaction between event existential closure, on the one hand, and quantification, negation, or conjunction, on the other hand. In a recent publication, Winter and Zwarts provide a solution to this problem. Winter and de Groote elaborate on this solution. In particular, they provide a treatment of quantified adverbial modifiers, which was absent from Winter and Zwarts, [19].

## 6.10. Pragmasemantic with Effects and Handlers

Jiří Maršík and Maxime Amblard have explored the feasibility of theories of side effects of programming languages in the study of natural language semantics and pragmatics [23]. In the approach that we are developing, the denotations we assign to fragments natural language are effectful computations. To demonstrate on an example, if we was to treat dynamics, then instead of changing the type of sentence denotations from  $o$  to  $c \rightarrow o * c$ , where  $c$  is the type of discourse contexts and  $o$  is the type of propositions, we would treat sentence denotations as effectful computations of type  $o$  that study and modify the context using effectful operations. This explicit distinction between 'result' and 'effects' brings to mind Stalnaker's distinction between 'content' and 'context'.

The motivation for this approach is to make it easier to compose multiple pragmasemantic phenomena by being allowed to put their effects aside. So far, a small prototype handling dynamics, presuppositions and some of their interactions is under development.

## 6.11. Mining Texts at discourse level

Linguistic discourse refers to the meaning of large chunks of text, from phrases to whole documents. It could be very useful for guiding attempts at text mining, which focus on document selection, document summarization, or other knowledge extraction goals. Hence the aim of this work is to apply Knowledge Discovery in Databases (KDD) methods to texts annotated with discourse information. Maxime Amblard with Yannick Toussaint (Orpailleur team) and Sara van de Moosdijk (master 2 intern) approach the problem by extracting discourse relations using unsupervised methods, which are then used to construct a knowledge model with Formal Concept Analysis (FCA). Pattern Structures (PS), an advancement in FCA, allow for the modelling of complex data. Our method is applied to a corpus of medical articles compiled from PubMed. This medical data is enhanced with concepts from the UMLS MetaThesaurus combined with the UMLS Semantic Network to serve as an ontology for Pattern Structure classification. The results show that despite having a large amount of noise, the method is promising and could be applied to other domains than the medical domain. We explore the pitfalls and suggest ways in which the process could be improved (Submission under review).

## 6.12. Exploring real datas

Maxime Amblard explored the use of formal framework for modelling transcription of real interviews, in particular one involves in the SLAM project with schizophrenics. Schizophrenia is well-known among mental illnesses for the severity of the thought disorders it involves, and for their widespread and spectacular manifestations: from deviant social behavior to delusion, not to mention affective and sensitive distortions. The goal of our interdisciplinary work is to (i) analyze linguistic troubles in conversational contexts in which one of the speakers is schizophrenic, (ii) construe how the concept of rationality and logicity may apply to them, and (iii) propose a formal representation about this specific manifestation. Maxime Amblard, Sylvain Pogodalla and Karen Fort propose surveys on past results [35], [29].

Maxime Amblard and Karen Fort have studied experiments they led concerning disfluencies in the discourse of schizophrenic patients (in remediation). These experiments are part of a larger study dealing with other levels of linguistic analysis, that could eventually help identifying clues leading to the diagnostic of the disease. This study largely relies on natural language processing tools, which allow for the rapid processing of massive textual data (here, more than 375,000 words). The first phase of the study, which they present confirmed the correlation between schizophrenia and the number of disfluences appearing in the discourse [25]. Moreover they have discussed ethical issues on the corpus with others [26].

### 6.13. Paraconsistency and Inconsistency-Friendly Logics

Paraconsistent logic is a family of formal systems in which the law of contradiction fails. In such systems, from an inconsistent set, *not* everything follows.

Can Baskent has studied such logical systems and their connections to formal linguistics within the framework of game theory. First, he observed how a game theoretical semantics can be given for some paraconsistent logics [43]. The advantage of game semantics is that it simply reflects the parsing tree of logic, and furthermore presents a semantical structure that uses elements from game theory. Such a study also requires an in-depth study of various paraconsistent logics, and their semantical structures [13]. Such a study requires some understanding of point-set topology, and its relation to logic.

Moreover, paraconsistent logics relate to dynamic logics as well. The logical model defines characterises how dynamic epistemic modalities, which are familiar from multi-agent systems, work [13]. This helps us understand how multi-agent interactions in an inconsistent model work in a sound way.

Another interesting way of seeing how inconsistency-friendly logics work is to consider them within the framework of game theory [37]. Game theory, similar to multi-agent systems, studies the intelligent and rational interaction of decision makers/agents. Yet, it suffers from various paradoxes. Such paradoxes are important from a computational semantical point of view. If paraconsistency is the most suitable tool to analyse paradoxes, then game theoretical paradoxes are not exceptions [37].

The technical work always needs to be supplemented by some conceptual work. Granted, paraconsistent logics find their ways in various philosophical and semantical issues, yet their computational analysis usually falls short. In [44], we discussed the connection between paraconsistent logics and Hintikka's interrogative models. These models have been developed by Hintikka, a pioneer of epistemic logic, and have been properly analysed from paraconsistent perspectives. If inquiry and questioning needs to be accounted for computationally, a paraconsistent approach will be an appropriate tool as well. Similarly, [39] discusses paraconsistency and its connection to social software. *Social Software* is a field conceived by Rohit Parikh, and it studies the computational and logical analysis of social protocols and policies. It lies in the intersection of social choice theory and game theory, and is a subset of logic.

Such results have been presented in various talks including, *World Congress of Paraconsistency* in Kolkata and *Logic and the Foundations of Game and Decision Theory* in Bergen, and warmly received.

## E-MOTION Project-Team

## 5. New Results

### 5.1. Highlights of the Year

- C. Laugier, E. Mazer and K. Mekhnacha have been finalists for the Eurobotics Technology Award 2014. Title “Bayesian perception & Decision: from theory to industrial applications”. March 2014.
- A. Nègre, L. Rummelhard, M. Perrollaz and C. Laugier had applied for a petenent "Procédé d'analyse d'une scene dybnamique, module d'analyse et programme d'ordinateur associés".

### 5.2. A new formulation of the Bayesian Occupancy Filter: a hybrid sampling based framework

**Participants:** Lukas Rummelhard, Amaury Nègre, Christian Laugier.

The Bayesian Occupancy Filter (BOF) is a discretized grid structure based bayesian algorithm, in which the environment is subdivised in cells to which random variables are linked. These random variables represent the state of occupancy and the motion field of the scene, without any notion of object detection and tracking, making the updating part of the filter an evaluation of the distribution of these variables, according to the new data acquisition. In the classic representation of the BOF, the motion field of each cell is represented as a neighborhood grid, the probability of the cell moving from the current one to another of the neighborhood being stocked in an histogram. If this representation is convenient for the update, since the potential antecedents of any cell is exactly determined by the structure, and so the propagation model is easily parallelizable, it also raises determinant issues :

- the structure requires the process rate to be constant, and a priori known.
- in the case of a moving grid, such as an application of car perception, many aliasing problems can appear, not only in the occupation grid, but in the motion fields of cells. A linear interpolation in 4-dimension field to fill each value of the histograms can quickly become unreasonable.
- to be able to match the slowest moves in the scene and the tiniest objects, the resolution of the grid and the motion histogram must be the high. On the other hand, since the system must be able to evaluate the speed of highly dynamic objects (typically, a moving car), the maximum encoded speed is to be high as well. This results in a necessary huge resolution grid, which prevent the system from being used with satisfying results on an embedded device. This huge grid is also mostly empty (most of the motion field histogram for a occupied cell will be empty). On top of that, the perception system being used to represent the direct environment of a moving car, the encoded velocity is a relative velocity, which implies, if we consider the maximal speed of a car to be  $V_{max}$ , to maintain a motion field able to represent speeds from  $-2 * V_{max}$  to  $2 * V_{max}$ . The necessity of such a sized structure is a huge limitation of practical use of the method.

Considering those limitations, a new way to represent the motion field has been developped. To do so, a new formulation of the BOF has been elaborated. This new version allow to introduce in the filter itself a distinction between static and dynamic parts, and so adapt the computation power.

The Hybrid Sampling Bayesian Occupancy Filter (HSBOF) [21] is an evolution of the BOF, in which are introduced additionnal concepts and variables, such as probabilistic classification of the environment between static and dynamic areas, and adaptative motion model structure. The main idea of this new representation is to mix two forms of sampling of the surrounding :

- a uniform sampling, represented as a dense regular grid, for the static objects and the empty areas. In this part, only the occupancy is stored, as the motion model of the static part of the scene is inherent. In practice, the section of the environment includes the vast majority of the scene.

- a non uniform sampling, based on particles drawn in dynamic regions, allowing to focus the computational power on the estimation of their motion. The number of particles used to represent the motion of a particular cell is calculated according to various criterions, such as the confidence in the dynamism of the cell, in its estimated motion, the global needs in the scene, etc. Dynamic regions are resampled at every time step, the amount of particles associated to the different parts of the scene is dynamically calculated.

The motion field in a cell is then represented as a set of samples from the distribution for values which are not null, and a weight given to the static hypothesis. The use of a set of samples to represent the motion field leads to a important decrease of the needed memory space, so do the classification between dynamic objects and static objects or free areas. In the updating process, the antecedent of a cell can be either from the static configuration or from the dynamic configuration, which are both way easier to project in the new reference frame of the moving grid: the static part requires a 2-dimension interpolation to be expressed in the new reference frame, the dynamic part a immediate particle association and a simple rotation of the velocity vectors.

This new version HSBOF is now used in the core of our systems in place of the previous version of the BOF. It presents important improvements in the quality of the estimations, while drastically reducing the memory and computation costs (easily by a 100 factor in term of memory).

### 5.2.1. Probabilistic grid-based collision risk prediction

**Participants:** Lukas Rummelhard, Amaury Nègre, Mathias Perrollaz, Christian Laugier.

We developed a new grid-based approach for collision risk prediction [23], based on the Hybrid-Sampling Bayesian Occupancy Filter framework. The idea is to compute an estimation of the Time To Contact (TTC) for each cell of the grid, instead of reasoning on objects. This strategy avoids to solve the difficult problem of multi-objects detection and tracking and provides a probabilistic estimation of the risk associated to each TTC value.

Using motion sensors embedded in the mobile robot (Inertial Measurement Unit, GPS, Wheel speed and steering sensor, visual odometry, etc.), the displacement of the grid between two updates is estimated. The full description of occupancy and dynamics of the scene given by the HSBOF is then used to assess collision risks in the future and even localize them in the grid. The risk evaluation consists in a short-term prediction of the scene configuration (figure 4 and of the robot position. This way a collision likelihood can be computed over time. Using those likelihoods, computed by cell and particle, an estimation of the risk over a period, and a localization of this risk in the grid are performed.

## 5.3. A new experimental platform for the Technological Research Institute (NanoElec)

**Participants:** Mathias Perrollaz, Nicolas Turro, Jean-François Cuniberto.

Within the framework of the PERFECT projet (founded by the IRT NanoElec), e-Motion has developed a new experimental platform, based on a Renault Zoe electrical vehicule (Fig. 7 ). This developpement takes advantage of the experience developed for creating the previous experimental plateforme (a Lexus LS600H), and go further by integrating more sensors and more functionalities.

The vehicle is equipped with:

- 4 IBEO LUX laser scanners. Each of them scans 4 layers with a field of view of 85 degrees.
- one Velodyne HD64L 3D laser scanner, capable of scanning 64 layers over 360 degrees.
- one trinocular stereo camera, Point Grey Bumblebee XB3, placed behind the windshield.
- 2 Ueye RGB cameras, looking forward and backward the vehicle.
- one XSens IMU/GPS sensor, used for positionning and ego-motion estimation.
- one ITRI 802.11p on-board unit, allowing V2X communication.

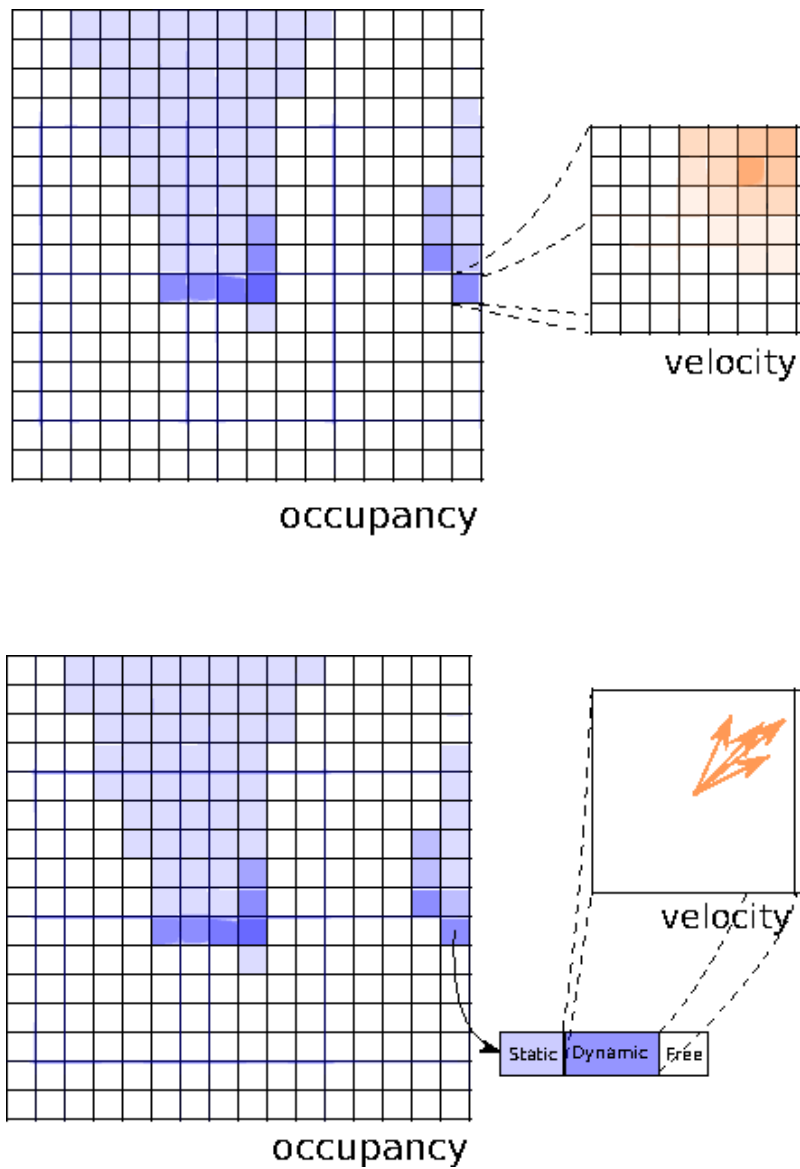


Figure 1.

Data representations in BOF and HSBOF formulation :

- (a) Classic BOF representation : a 2 dimension grid, to each cell are assigned an occupancy value and a velocity histogram,
- (b) Proposed representation : a 2 dimension grid, to each cell are assigned an occupancy value, a static coefficient  $P(V = 0)$  and a set of particles drawn along  $P(V = v|V \neq 0)$



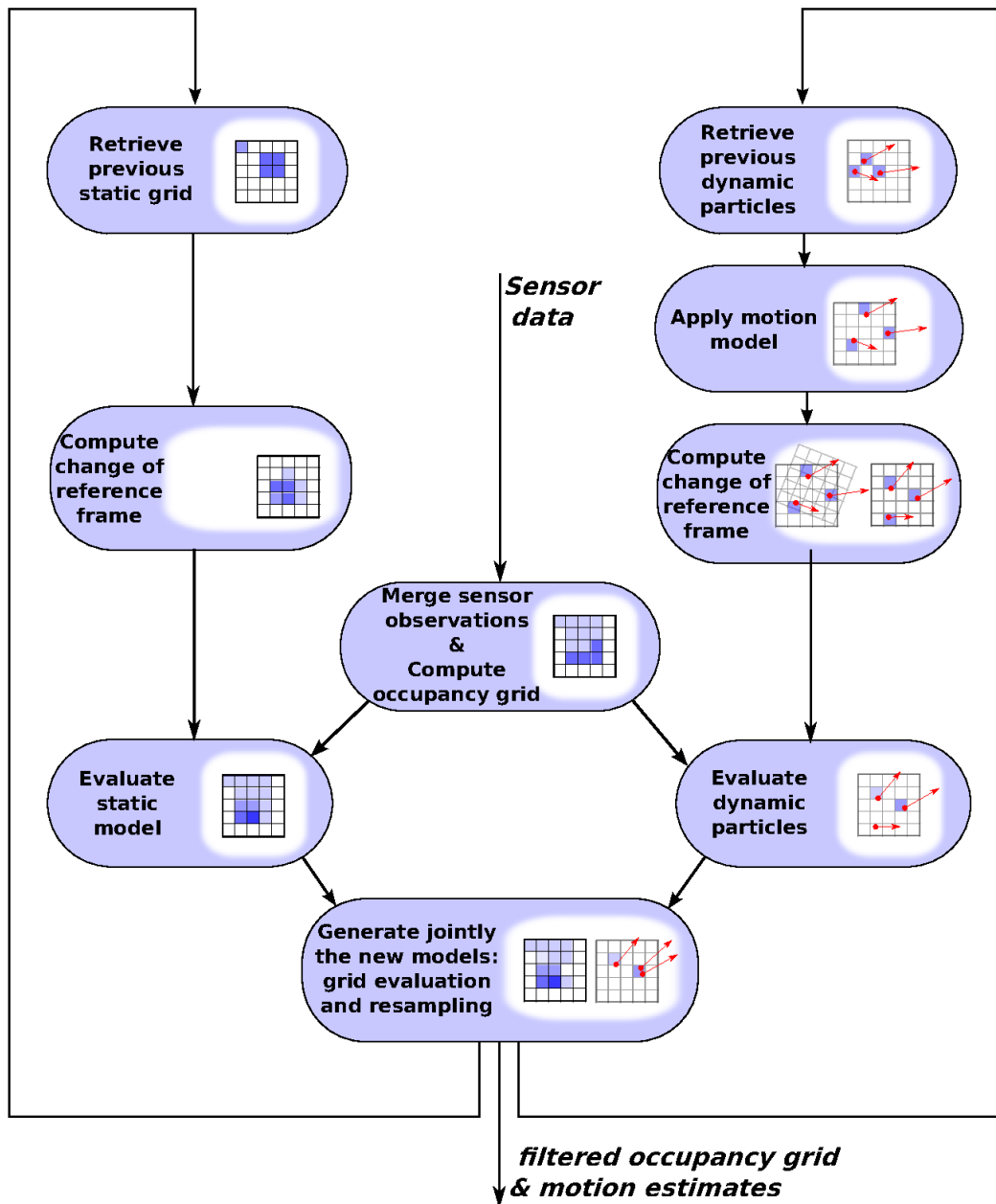


Figure 2. HSBOF algorithm summary. From sensor data instantaneous occupancy grids are successively computed. Those observations are integrated in a Bayesian filter in which coexist and jointly adapt two models, a static grid and a dynamic set of moving particles. The result is obtained by their combination, which provides a filtered occupancy grid as well as inferred motion distributions for cells.

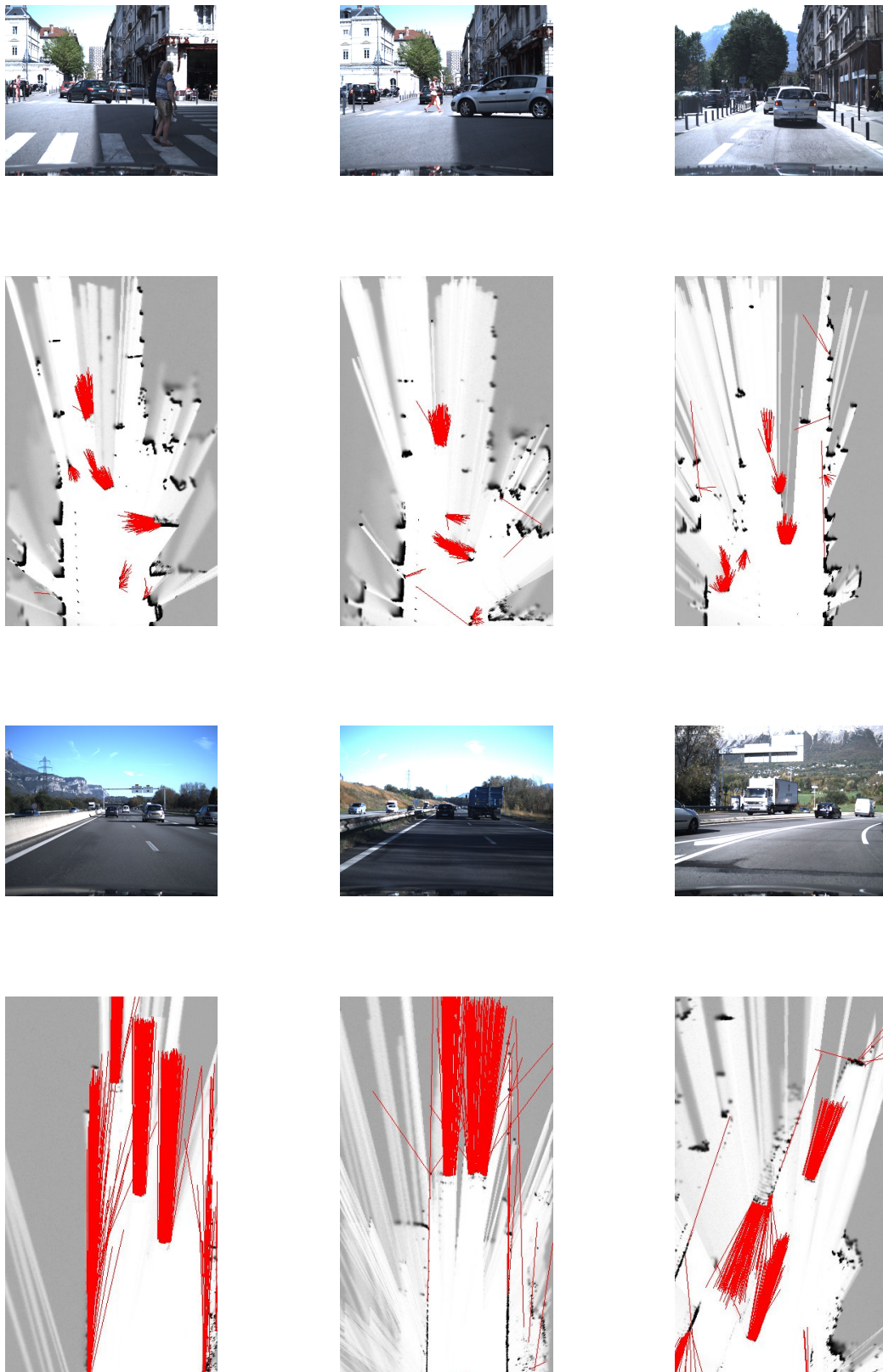


Figure 3. Resulting occupancy grid and velocity field on different urban and highway situations. White cells represent the free space, grey one the unknown space (hidden). Black cells represent the occupied space and red lines represent the average velocity vector for cell with a high dynamic probability.

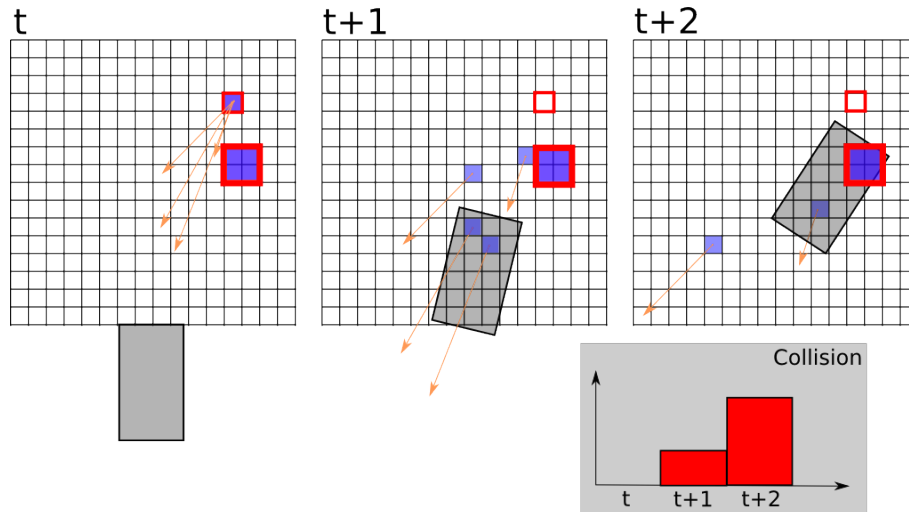


Figure 4. Collision risk estimation over time for a specific cell. The cell position is predicted according to its velocity, along with the mobile robot. This risk profile is computed for every cell, and then used to integrate over time the global collision risk.



(a)



(b)

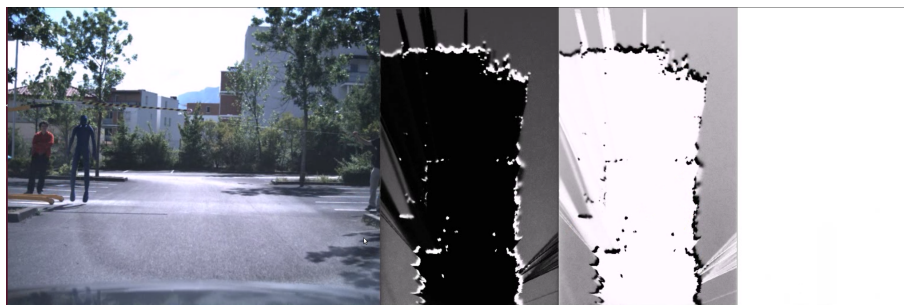
Figure 5. (a) Fake pedestrian used for experiments. (b) The mannequin is attached to a system with a runner, in order to allow lateral displacements.



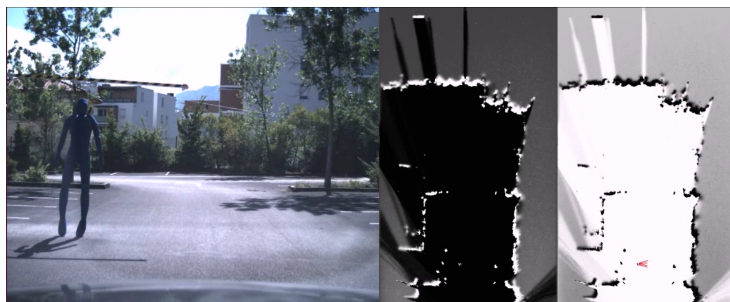
(a)



(b)



(c)



(d)

Figure 6. Results of the system. Each image is a visual capture from the embedded camera, the estimated occupancy grid in front of the car (white for occupied, grey for unknown, black for empty), the estimated motion field (if a case is seen as dynamic, a red motion vector showing the average velocity in the cell is drawn on the map) and finally the estimated risk map for 0.5s. The first sequence (a) (b) presents the appearance of an occluded pedestrian, the second (c) (d) a moving pedestrian heading towards the road.



All the synchronization, display, play/record, and developments capabilities are relying on the ROS middle-ware. The vehicle is fully operational at the end of 2014.

The vehicle is designed for experimenting in both ADAS (Advanced Driver Assistance Systems) and autonomous driving applications. In parallel, V2X communications are installed on the IRT "smart city" environment, so that the vehicle can evolve on this site and interact with it.



Figure 7. The Zoe experimental platform.

## 5.4. Visual localization with Open Street Map

**Participants:** Jean-Alix David, Amaury Nègre.

Given the lack of precision of GPS for localization, it is necessary to implement new ways to improve localization. Here we introduce a new method using a geographic map and a camera to do so. The main point of this method is to combine sensor readings and known data about the environment. We detect lines on the road with the camera, and then compare the extracted lines to the ones stored in the map using ICP (Iterative Closest Point) algorithm.

The used map is OpenStreetMap, it allow to have information on the roads and lanes for example, but there is no information about white marking. So we generated semi-autonomously the lines given roads and number of lanes. Moreover we manually corrected the lines for crossroads using satellite image (see Figure 8 ).

The line extraction is done using ridge detection on a top-down view of the camera image. Moreover we use GPU acceleration to improve performances during image processing (see Figure 9 ).

The OSM generated data and the lines extracted from the camera will then be matched and the transformation between the camera and the absolute map will be compute by using an Iterative Closest Point algorithm. In order to improve the precision, a bayesian filtering approach will also be used to merge the previous results with GPS and Inertial Measurement Unit data.

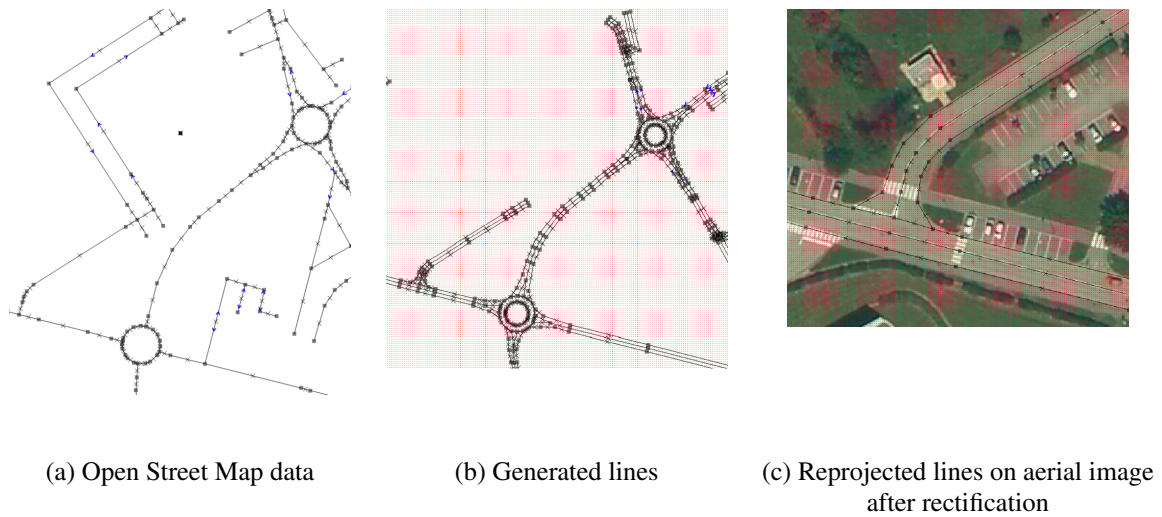


Figure 8. Semi-automatic road line generation from Open Street Map.

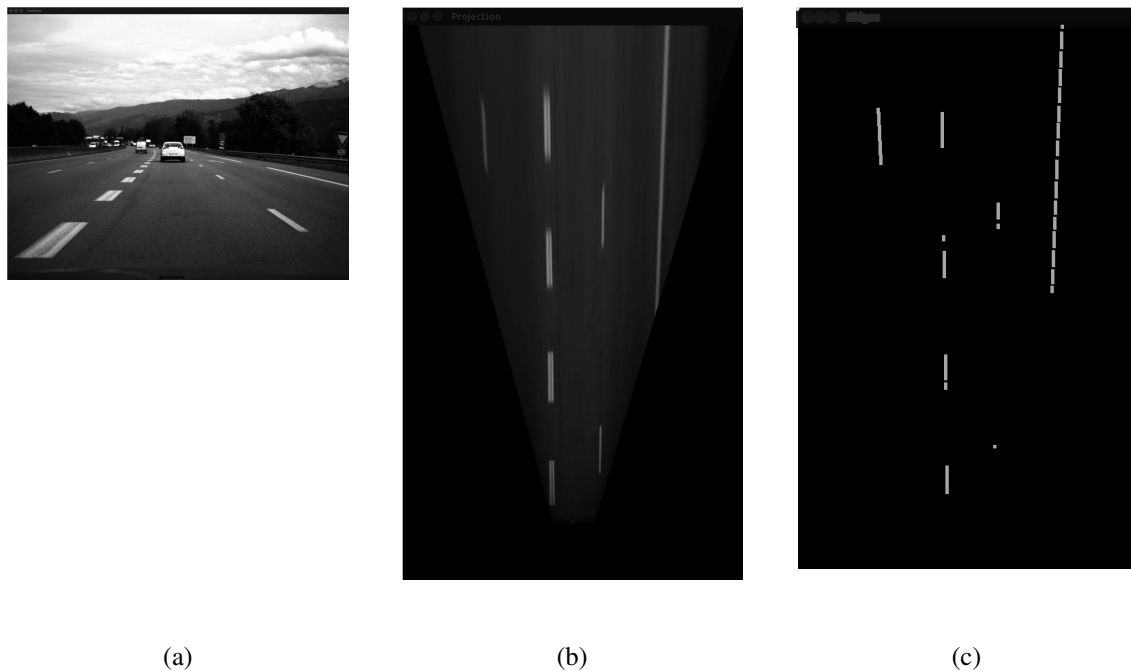


Figure 9. Line detection in camera images: projection in the ground plane (b) ridge extraction (c).



## 5.5. Human Centered Navigation in the physical world

### 5.5.1. Social Mapping

**Participants:** Panagiotis Papadakis, Anne Spalanzani, Christian Laugier.

With robots technology shifting towards entering human populated environments, the need for augmented perceptual robotic skills emerges that complement to human presence. In this integration, perception and adaptation to the implicit human social conventions plays a fundamental role. Toward this goal, we introduce in 2013 a novel methodology to detect and analyse complex spatial interactions of multiple people and encode them in the form of a social map, whose structure is obtained by computing a latent space representation of human proxemic behaviour [32]. In 2014, Panagiotis left to Lagadic-Sophia and we carried on this work by integrating a planning algorithm to validate the perception part on a real robot. This work was published at IROS 2014 [22].

### 5.5.2. Goal oriented risk based navigation in social and dynamic environment

**Participants:** Anne Spalanzani, Procopio Silveira-Stein, Gregoire Vignon, Christian Laugier.

Since 2008 we have proposed a new concept to integrate a probabilistic collision risk function linking planning and navigation methods with the perception and the prediction of the dynamic environments [31]. The likelihood of the obstacles' future trajectory and the probability of occupation are used to compute the risk of collision. A social filter was added to give the robot the ability to move in a social way (see Figure 10 ). In 2014, we obtained an Inria ADT(ADT PN2) to optimize and share the RiskRRT algorithm. This work is under development. We published in [15] a survey on human-aware navigation.

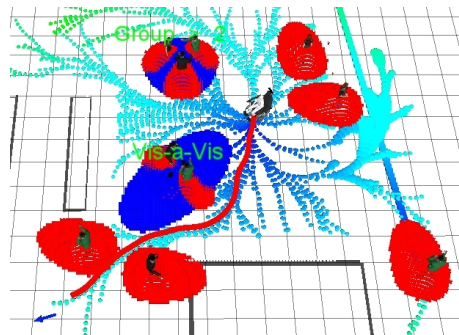


Figure 10. Illustration of the RiskRRT in a social environment

### 5.5.3. Navigation Taking Advantage of Moving Agents

**Participants:** Procopio Silveira-Stein, Anne Spalanzani, Christian Laugier.

In this work, we propose a different form of robotic navigation in dynamic environments, where the robot takes advantage of the motion of pedestrians, in order to improve its own navigation capabilities. The main idea is that, instead of treating persons as dynamic obstacles that should be avoided, they should be treated as special agents with an expert knowledge of navigating in dynamic scenarios. To benefit from the motion of pedestrians, this work proposes that the robot selects and follows them, so it can move along optimal paths, deviate from undetected obstacles, improve navigation in densely populated areas and increase its acceptance by other humans. In 2014, we focused on real experiments (see Figure 11 using the wheelchair and results were published in [16], [25], [26].



Figure 11. Switching navigation method between leader following and independent navigation. In image 1 the robot is engaged in leader following, while in the remaining it uses RiskRRT for the navigation.

#### 5.5.4. Autonomous Wheelchair for Elders Assistance

**Participants:** Arturo Escobedo-Cabello, Gregoire Vignon, Anne Spalanzani, Christian Laugier.

The aging of world's population is bringing the need to provide robotic platforms capable to assist elder people to move [33]. It is necessary that such transportation is reliable, safe and comfortable. People with motor disabilities and elders are expected to benefit from new developments in the field of autonomous navigation robotics.

Autonomously driven wheelchairs are a real need for those patients who lack the strength or skills to drive a normal electric wheelchair. The services provided by this kind of robots can also be used to provide a service of comfort, assisting the user to perform difficult tasks as traversing a door, driving in a narrow corridor etc. In 2014, we combined user intention estimation, a navigation using social convention to perform comfortable trajectories (see Figure 12 . Results were published in the IROS conference [19]. Arturo Escobedo defended his PhD in october 2014.



Figure 12. The Robotic wheelchair assists its user to acheive his task of navigation/.

### 5.6. Human modeling for situation understanding

#### 5.6.1. Situation understanding and risk assessment for intelligent vehicles

**Participants:** Dizan Vasquez, Stéphanie Lefèvre, Suryansh Kumar, Yufeng Yu.

The work on this period has been aimed at establishing a solid theoretical and technological base for our research on situation understanding. A step in this direction was the elaboration of an in-depth survey of the current state of the art on the field, prepared together with the university of Berkeley [12]. In the framework of the same collaboration, we have been working on the introduction of human models in current Advanced Driving Assistance Systems (ADAS). This has led to the development of a novel Lane Keeping Assistance System (LKAS) which is able to learn the driver's driving patterns and use them to predict lane departures as well as to generate controls that mimic the driver's style and are, thus, deemed to be more acceptable. The approach has been evaluated against commercial LKAS using real field data, and the results show that the proposed approach is both more efficient and less intrusive than current approaches. This is, for the best of our knowledge the first use of human models withing LKAS and these results illustrate the strong potential that these models may have in ADAS.

Concerning autonomous navigation, we have focused on human-like motion planning for motion prediction. The main hypothesis is that people behave like planners whose motion optimizes some an unknown cost function. Under this assumption, the main challenge becomes to model that cost function and to learn its parameters from demonstrated behavior. This is called, depending on the community, either *Inverse Reinforcement Learning* (IRL) or *Inverse Optimal Control* (IOC). Now, a problem with IRL is that it requires examples of both desirable and undesirable behavior, which are difficult to obtain with a real platform. Additionally, there is no consistent benchmarking methodology to evaluate different approaches. This has motivated our work in a benchmark comprised of: (a) an evaluation methodology; (b) a simulated experimental platform (Fig. 13 ) based on the Torcs simulator ; and (c) real data gathered with our instrumented Lexus vehicle. The first prototype of this benchmark, developed together with students from Beijing University and IIT Hyderabad, has been presented this year in a vehicular technologies conference .



Figure 13. Experimental platforms: left) Our Torcs-based racing simulator; right) sensor-equipped Lexus vehicle.

### 5.6.2. Socially compliant robot navigation in human environments

**Participant:** Dizan Vasquez.

The models we have applied to intelligent vehicles are also adapted in general to situations where mobile robots share their environment with humans. This has lead us to apply this techniques to the assistive robotics fields, given that it is one of e-motion’s major applications axes. Our first effort in this sense has been to design and develop a robust experimental platform with baseline modules for motion planning, perception and social awareness .

In parallel we started working, in collaboration with the University of Freiburg, on a benchmarking platform for social compliant motion planning, close in spirit to the one proposed for intelligent vehicles. The platform (Fig. 14 ) is described in , it includes several motion planning and feature extraction algorithms as well as a pedestrian simulator based on Helbing’s social force model .

## 5.7. Sensor Fusion for state parameters identification

**Participants:** Agostino Martinelli, Chiara Troiani.

### 5.7.1. General theoretical results

During this year we have focused our research on two distinct domains:

- the visual-inertial structure from motion problem;
- the derivation of analytical solutions for the probability distribution of a Brownian motion that satisfies the unicycle constraint.

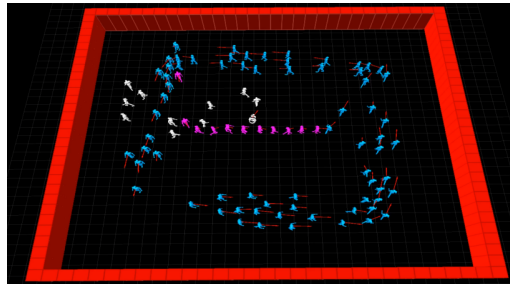


Figure 14. A screenshot of our pedestrian simulator.

The research carried out on the first domain is the follow up of our previous activity. We continued to investigate the observability properties of the visual inertial structure from motion and in particular we have analyzed the case when some of the inertial sensors are missing. This analysis has never been provided before and we started this investigation at the end of last year. During this year we confirmed the validity of our preliminary analysis and we also extended them. The preliminary results were obtained by referring to the case when at least five point features are available and showed that the observability properties of visual inertial structure from motion do not change by removing all the three gyroscopes and one accelerometer. By removing a further accelerometer, if the camera is not extrinsically calibrated, the system loses part of its observability properties. On the other hand, if the camera is extrinsically calibrated, the system maintains the same observability properties as in the standard case. These results have been published on the journal *Foundations and Trends in Robotics* and have also been presented at the last ICRA conference[20].

We recently extended these results by considering the extreme case of a single point feature (i.e., not five). This analysis required to approach an open problem in control theory, called the Unknown Input Observability (UIO). In [20] we proposed a possible method to solve this UIO problem. However, we had to improve this method to deal with this extreme case (i.e., the case of one single point feature). Preliminary results on the extension of this method have been published as a research report [30] and we also plan to present them at the next American Control Conference. By applying this method to our problem, we obtained new interesting results. The new investigation allowed us to conclude that, 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). Our first objective is to validate these new results.

Regarding the second domain mentioned above, we have derived a complete analytical solution for the probability distribution of the configuration of a non-holonomic mobile robot that moves in two spatial dimensions by satisfying the unicycle kinematic constraints. The proposed solution differs from previous solutions since it is obtained by deriving the analytical expression of any-order moment of the probability distribution. To the best of our knowledge, an analytical expression for any-order moment that holds even in the case of arbitrary linear and angular speed, has never been derived before. To compute these moments, a direct integration of the Langevin equation has been carried out and each moment was expressed as a multiple integral of the deterministic motion (i.e., the known motion that would result in absence of noise).

For the special case when the ratio between the linear and angular speed is constant, the multiple integrals can be easily solved and expressed as the real or the imaginary part of suitable analytic functions. As an application of the derived analytical results, we also investigated the diffusivity of the considered Brownian motion for constant and for arbitrary time-dependent linear and angular speed. These results have been published on



the journal of statistical mechanics [13] and also as a research report [29] where we added more specific considerations about the impact of the derived results on mobile robotics.

### 5.7.2. Applications with a Micro Aerial Vehicle

We continued our previous activity about the estimation of the relative motion between two consecutive camera views in order to introduce very efficient algorithms to remove the outliers of the feature-matching process. Thanks to their inherent efficiency, the proposed algorithms are very suitable for computationally-limited robots.

In particular, during this year, we extended the previous results by removing the assumption of planar motions. In this case, to obtain useful results, we had to include one more point feature (i.e., the proposed algorithms only use two feature correspondences and gyroscopic data from IMU measurements to compute the motion hypothesis). By exploiting this 2-point motion parametrization, we proposed two algorithms to remove wrong data associations in the feature matching process for case of a 6DoF motion. We showed that in the case of a monocular camera mounted on a quadrotor vehicle, motion priors from IMU can be used to discard wrong estimations in the framework of a 2-point-RANSAC based approach. The proposed methods have been evaluated on both synthetic and real data and presented at the last ICRA conference [27].

## 5.8. Compiling Probabilistic Programs Onto Reconfigurable Logic Using Stochastic Arithmetic

**Participants:** Emmanuel Mazer, Marvin Faix.

It is of great interest to perform light weight probabilistic inferences for applications such as sensor fusion. Our goal is to design systems to perform these inferences without using a Von Newman machine nor standard floating point arithmetic. By addressing the core of how computations are made, we can explore the tradeoffs between system precision with power consumption and computation time, enabling artificial systems with limited resources, such as mobile and embedded systems, to better operate under uncertainty. Figure 15 illustrates the tool-chain, which starts from the specification of the Bayesian Program in Bayesian programming language, and evaluates it on a reconfigurable device.

This study is part of BAMBI (Bottom-up Approaches to Machines dedicated to Bayesian Inference, [www.bambi-fet.eu](http://www.bambi-fet.eu)): a European collaborative research project relying on the theory of Bayesian inference to understand the natural cognition and aiming at designing bio-inspired computing devices.

A Bayesian machine has probability distributions as inputs and returns a probability distribution as output. It is defined by a joint probability distribution on a set of discrete and finite variables:  $P(M \wedge D \wedge L)$ . Where  $M$ ,  $D$  and  $L$  are themselves conjunctions of variables, for example  $D = D_1 \wedge \dots \wedge D_k$ . We define the soft evidences on the variables  $D_k$  as the probability distribution  $\tilde{P}(D_k)$ . These soft evidences will be the inputs of the Bayesian machine.

So, given the soft evidences  $\tilde{P}(D_k)$  and the joint distribution  $P(M \wedge D \wedge L)$ , the machine will fulfil the specification if it computes:

$$P'(M) = \frac{1}{Z} \sum_{D_1} \tilde{P}(D_1) \dots \sum_{D_k} \tilde{P}(D_k) \sum_L P(M \wedge D \wedge L) \quad (1)$$

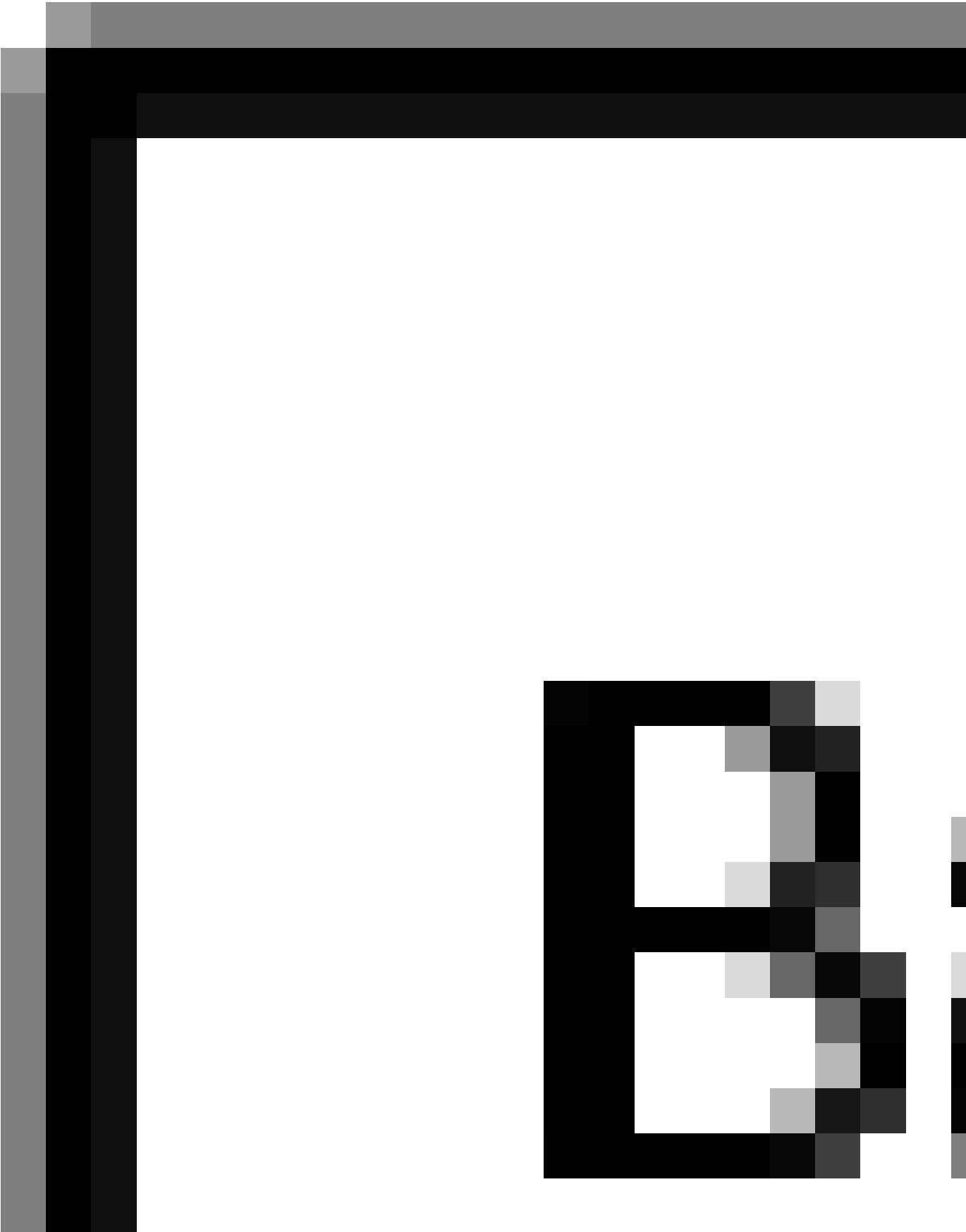
with

$$Z = \sum_M \left( \sum_{D_1} \tilde{P}(D_1) \dots \sum_{D_k} \tilde{P}(D_k) \sum_L P(M \wedge D \wedge L) \right)$$

In other words the machine computes a soft inference based on the joint distribution  $P(M \wedge D \wedge L)$ .

A modified version of the probabilistic language ProBT is used to specify the machine: the joint distribution, the output and the inputs are specified with this language<sup>0</sup>. The next program is an example of a simple specification using the Python bindings of ProBT.





```

#import the ProBT bindings
from pypl import *
#define the variables
dim3 = plIntegerType(0,2)
D1 = plSymbol(D1,dim3)
D2 = plSymbol(D2,dim3)
M= plSymbol(M,dim3)
#define the distribution on M
PM= plProbTable(M,[0.8,0.1,0.1])
#define a conditional distribution on D1
PD1_k_M = plDistributionTable(D1,M)
PD1_k_M.push(plProbTable(D1,[0.5,0.2,0.3]),0)
PD1_k_M.push(plProbTable(D1,[0.5,0.3,0.2]),1)
PD1_k_M.push(plProbTable(D1,[0.4,0.3,0.3]),2)
#define a conditional distribution on D2
PD2_k_M = plDistributionTable(D2,M)
PD2_k_M.push(plProbTable(D2,[0.2,0.6,0.2]),0)
PD2_k_M.push(plProbTable(D2,[0.6,0.3,0.1]),1)
PD2_k_M.push(plProbTable(D2,[0.3,0.6,0.1]),2)
#define the joint distribution
model=plJointDistribution(PM*PD1_k_M*PD2_k_M)
#define the soft evidence variables
model.set_soft_evidence_variables(D1^D2)
#define the output
question=model.ask(M)

```

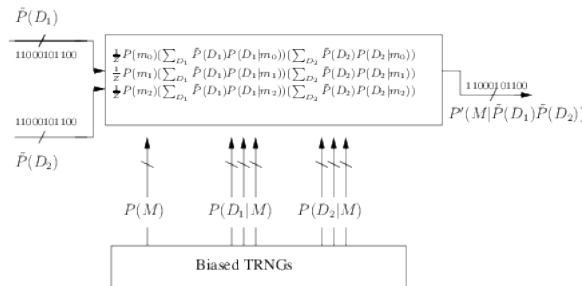


Figure 16. The probabilistic machine corresponding to the given program..

Figure 16 presents the high-level representation of the architecture for the Bayesian Machine. It comprises the main stochastic machine along with the True Random Generators (TRNG), responsible for the generation of the stochastic bit streams for the constants considered in the problem.

<sup>0</sup>A free version of ProBT is available at <http://www.probayes.com/fr/Bayesian-Programming-Book/> and the version with soft evidence will be placed on [www.bambi-fet.eu](http://www.bambi-fet.eu) before the NIPS conference

The proposed tool-chain is working and accepts any ProBT program with discrete variables as entry. The tool-chain generates a VHDL file which is the description of the stochastic circuit and can be implemented on a FPGA. A Cyclone IV FPGA, from Altera has been targeted as supporting device. A machine has been synthesised to demonstrate the applicability and scalability of the proposed tool-chain. ProBT is also used to compute the exact result using standard arithmetic. This allows to evaluate the results given by FPGA with the synthesised VHDL program.

Figure 17 (right) shows the RTL generated by the synthesis tool, where it is possible to identify the connections between the components, corresponding to the circuit in Figure 17 (left). This circuit was implemented using 6 Logic Elements. The circuit was tested with bit streams integrated over  $2^{31}$  to do the conversion from stochastic to binary.

The proposed tool-chain is working and accepts any ProBT program with discrete variables as entry. The tool-chain generates a VHDL file which is the description of the stochastic circuit and can be implemented on a FPGA. A Cyclone IV FPGA, from Altera has been targeted as supporting device. A machine has been synthesised to demonstrate the applicability and scalability of the proposed tool-chain.

ProBT is also used to compute the exact result using standard arithmetic. This allows to evaluate the results given by FPGA with the synthesised VHDL program.

Figure 17 (right) shows the RTL generated by the synthesis tool, where it is possible to identify the connections between the components, corresponding to the circuit in Figure 17 (left). This circuit was implemented using 6 Logic Elements. The circuit was tested with bit streams integrated over  $2^{31}$  to do the conversion from stochastic to binary.

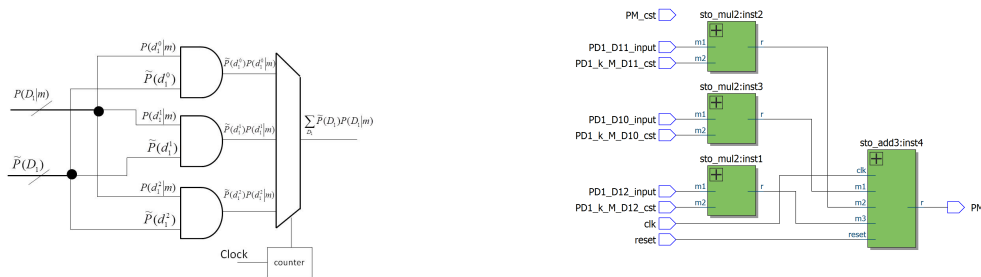


Figure 17. Stochastic circuit computing  $\sum_{D_1} \tilde{P}(D_1)P(D_1|m)$  and the corresponding RTL.

We are now focusing on solving the time dilution problem by introducing memory in the architecture. Then we will make an attempt to build a filter with similar ideas by re-fitting the output into the initial joint distribution.

## FLOWERS Project-Team

## 6. New Results

### 6.1. Highlights of the Year

PY. Oudeyer and M. Lopes, together with J. Gottlieb (Univ. Columbia, NY) organized the first International Symposium on Neurocuriosity symposium on Information Seeking, Curiosity and Attention, pioneering a gathering of world experts on curiosity from developmental psychology, neuroscience, ethology and computational modelling (see <https://openlab-flowers.inria.fr/t/first-interdisciplinary-symposium-on-information-seeking-curiosity-and-attention/21>). This was achieved in the context of associated team Neurocuriosity with the cognitive neuroscience lab of J. Gottlieb at Univ. Columbia, NY, US. The first results investigating predictions of theoretical formal models of curiosity on human exploration were also published [25].

O. Mangin obtained the Best thesis poster from Bordeaux doctoral school of mathematics and computer science, for his PhD thesis "The Emergence of Multimodal Concepts: From Perceptual Motion Primitives to Grounded Acoustic Words" [24].

The team, in collaboration with Inaki Iturrate and Luis Montesano, published major results on calibration-free brain-computer interface methods, where incremental machine learning algorithms are used to remove the phase of calibration for an important family of use contexts [44] [45].

In october 2014, the team announced the release of a new version of the Poppy Project platform, dedicated in particular to the use of tools for creating and programming interactive robots in Education and Art. This platform, which is a result of research on the role of morphology in skill acquisition within ERC project Explorers, was selected as finalist for the Global Fab Awards 2014 (<https://www.fab10.org/en/awards>) which select the best worldwide projects in the Makers ecosystem. It was also presented in major international press and media (<https://www.poppy-project.org/in-the-press/>), in multiple hackatons and demos, in particular at the major international conference LeWeb (<https://www.poppy-project.org/social-life/>), and its video on the web was seen 125k times. Poppy Project was presented at Elysée, during a French Tech event, to François Hollande (<http://www.inria.fr/centre/bordeaux/actualites/poppy-le-robot-humanoide-a-l-elysee>), and in Bordeaux to Axelle Lemaire. Web site: <http://www.poppy-project.org>

The Flowers team made major achievements in diffusing science and technology towards the general public. The team developed the IniRobot pedagogical kit, for the discovery of computer science and robotics in primary schools. The kit was first developed and evaluated in schools, in collaboration with a group of teachers, and then began to be largely disseminated and used in september 2014 to schools in Talence, Bordeaux, Lormont, and Lille. A dedicated web site has been created, allowing all users and contributors to share their experiences with the kit: <https://dm1r.inria.fr/c/kits-pedagogiques/inirobot>. PY. Oudeyer was invited to give a TedX talk (<https://www.youtube.com/watch?v=AP8i435ztwE>, video viewed by more than 9000 people), and was interviewed and invited to talk about our research on major media channels (e.g. Le Monde, Les Echos, France Inter, see <http://www.pyoudeyer.com/press/>).

## 6.2. Robotic And Computational Models Of Human Development

### 6.2.1. Computational Models Of Information-Seeking, Curiosity And Attention

**Participants:** Manuel Lopes, Pierre-Yves Oudeyer [correspondant], Jacqueline Gottlieb, Adrien Baranes, Pierre Rouanet, Brice Miard, Jonathan Grizou.

#### 6.2.1.1. *The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration*

Devising efficient strategies for exploration in large open-ended spaces is one of the most difficult computational problems of intelligent organisms. Because the available rewards are ambiguous or unknown during the exploratory phase, subjects must act in intrinsically motivated fashion. However, a vast majority of behavioral and neural studies to date have focused on decision making in reward-based tasks, and the rules guiding intrinsically motivated exploration remain largely unknown. To examine this question we developed a paradigm for systematically testing the choices of human observers in a free play context. Adult subjects played a series of short computer games of variable difficulty, and freely choose which game they wished to sample without external guidance or physical rewards. Subjects performed the task in three distinct conditions where they sampled from a small or a large choice set (7 vs. 64 possible levels of difficulty), and where they did or did not have the possibility to sample new games at a constant level of difficulty. We show that despite the absence of external constraints, the subjects spontaneously adopted a structured exploration strategy whereby they (1) started with easier games and progressed to more difficult games, (2) sampled the entire choice set including extremely difficult games that could not be learnt, (3) repeated moderately and high difficulty games much more frequently than was predicted by chance, and (4) had higher repetition rates and chose higher speeds if they could generate new sequences at a constant level of difficulty. The results suggest that intrinsically motivated exploration is shaped by several factors including task difficulty, novelty and the size of the choice set, and these come into play to serve two internal goals—maximize the subjects' knowledge of the available tasks (exploring the limits of the task set), and maximize their competence (performance and skills) across the task set. This was published in [25].

#### 6.2.1.2. *A new experimental setup to study the structure of curiosity-driven exploration in humans*

We started evaluating several games that test how humans explore a space of motor tasks of different complexities. Our objective is to observe what exploratory behaviors do people use when learning a new skill. The main hypothesis we are testing is that skills that provide a larger learning progress will be favored and so we will see a progression from the simpler to the more complex skills. Surely there are individual differences and the causes and impact of those differences is a very important research topic. The Abstract Games we created allows us to create several dimensions of complexity for the games. In this task, there are several abstract forms that appear in the screen and the user is able to control them using its own body (tracked using a Kinect sensor), see Fig. 2. The relation between the degrees of freedom and the forms/colors/sizes of the shapes is arbitrary and the user must explore its body to be able to control its behavior. This was published in [58].

### 6.3. Life-Long Robot Learning And Development Of Motor And Social Skills

#### 6.3.1. *Exploration and learning of sensorimotor policies*

##### 6.3.1.1. *Non-linear regression algorithms for motor skill acquisition: a comparison*

**Participants:** Thibaut Munzer [correspondant], Freek Stulp, Olivier Sigaud.

Endowing robots with the capability to learn is an important goal for the robotics research community. One part of this research is focused on learning skills, where usually two learning paradigms are used sequentially. First, a robot learns a motor primitive by demonstration (or imitation). Then, it improves this motor primitive with respect to some externally defined criterion. We realized a study on how the representation used in the demonstration learning step can influence the performance of the policy improvement step. We provide a conceptual survey of different demonstration learning algorithms and perform an empirical comparison of their performance when combined with a subsequent policy improvement step. These study have been published at the JFPDA conference [61].

During this work, we have discovered that many (batch) regression algorithms (amongst others, locally weighted (projection) regression, Gaussian mixture regression, radial basis function networks, and Gaussian process regression) use only one of two underlying model representations to represent a function: a weighted sum of basis function, or a mixture of linear models. Furthermore, we show that the former is a special case of the latter. This insights provides a deep understanding of the relationship between these algorithms, that, despite being derived from very different principles, use a function representation that can be captured within one unified model. A review article on this topic has been submitted to Neural Networks.

#### 6.3.1.2. *Simultaneous On-line Discovery and Improvement of Robotic Skill Options*

**Participants:** Freek Stulp [correspondant], Laura Herlant, Antoine Hoarau, Gennaro Raiola.

The regularity of everyday tasks enables us to reuse existing solutions for task variations. For instance, most door-handles require the same basic skill (reach, grasp, turn, pull), but small adaptations of the basic skill are required to adapt to the variations that exist (e.g. levers vs. knobs). In a joint project with Laura Herlant of Carnegie Mellon University, we developed the algorithm “Simultaneous On-line Discovery and Improvement of Robotic Skills” (SODIRS) that is able to autonomously discover and optimize skill options for such task variations. We formalize the problem in a reinforcement learning context, and use the  $PI^{BB}$  algorithm to continually optimize skills with respect to a cost function. SODIRS discovers new subskills, or “skill options”, by clustering the costs of trials, and determining whether perceptual features are able to predict which cluster a trial will belong to. This enables SODIRS to build a decision tree, in which the leaves contain skill options for task variations. We demonstrate SODIRS’ performance in simulation, as well as on a Meka humanoid robot performing the ball-in-cup task. This work has led to a publication at IROS [64].

#### 6.3.1.3. *Simultaneous On-line Discovery and Improvement of Robotic Skill Options*

**Participants:** Freek Stulp [correspondant], Nicolas Alberto Torres, Michael Mistry.

Freek Stulp supervised the Master’s thesis project of Nicolas Torres Alberto from the Telecom Physique Strasbourg, which led to a publication at Humanoids’ 14 [65]. The project focused on improving autonomy in learning inverse dynamics models for computed torque control. In computed torque control, robot dynamics are predicted by dynamic models. This enables more compliant control, as the gains of the feedback term can be lowered, because the task of compensating for robot dynamics is delegated from the feedback to the feedforward term. Previous work has shown that Gaussian process regression is an effective method for learning computed torque control, by setting the feedforward torques to the mean of the Gaussian process. We extend this work by also exploiting the variance predicted by the Gaussian process, by lowering the gains if the variance is low. This enables an automatic adaptation of the gains to the uncertainty in the computed torque model, and leads to more compliant low-gain control as the robot learns more accurate models over time. On a simulated 7-DOF robot manipulator, we demonstrate how accurate tracking is achieved, despite the gains being lowered over time. This is a first step towards life-long learning robots, that continuously and autonomously adapt their control parameters (feedforward *and* feedback) over extended periods of time.

#### 6.3.1.4. *Learning manipulation of flexible tools*

**Participants:** Clément Moulin-Frier [correspondant], Marie-Morgane Paumard, Pierre Rouanet.

Clément Moulin-Frier and Pierre Rouanet supervised the internship of Marie-Morgane Paumard from the *Ecole Normale Supérieure de Cachan*, at the Bachelor level. The internship has been realized from May to August 2014. Her report is entitled *Learning the manipulation of flexible tools in developmental robotics: a fishing robot* and is available at this address: [https://flowers.inria.fr/clement\\_mf/files/Paumard\\_RapportDeStage.pdf](https://flowers.inria.fr/clement_mf/files/Paumard_RapportDeStage.pdf).

Learning how to manipulate flexible tools is an harsh issue in robotics, since there is generally no analytical model of the system dynamics available. Learning algorithms are therefore a pivotal tool to control such systems. Marie-Morgane conceived an experiment on the manipulation of a fishing rod by a 2-arm robot equipped with a movement generation and perceptual systems. She studied how an optimization algorithm allows the robot to reach particular position of the hook on the floor. Then, she analyzed the distribution of effects (i.e. final fishhook position) in different contexts as well as optimization performances for particular goals.



#### 6.3.1.5. *Learning how to reach various goals by autonomous interaction with the environment: unification and comparison of exploration strategies*

**Participants:** Clément Moulin-Frier [correspondant], Pierre-Yves Oudeyer.

In the field of developmental robotics, we are particularly interested in the exploration strategies which can drive an agent to learn how to reach a wide variety of goals. We unified and compared such strategies, recently shown to be efficient to learn complex non-linear redundant sensorimotor mappings. They combine two main principles. The first one concerns the space in which the learning agent chooses points to explore (motor space vs. goal space). Previous works (Rolf et al., 2010; Baranes and Oudeyer, 2012) have shown that learning redundant inverse models could be achieved more efficiently if exploration was driven by goal babbling, triggering reaching, rather than direct motor babbling. Goal babbling is especially efficient to learn highly redundant mappings (e.g the inverse kinematics of a arm). At each time step, the agent chooses a goal in a goal space (e.g uniformly), uses the current knowledge of an inverse model to infer a motor command to reach that goal, observes the corresponding consequence and updates its inverse model according to this new experience. This exploration strategy allows the agent to cover the goal space more efficiently, avoiding to waste time in redundant parts of the sensorimotor space (e.g executing many motor commands that actually reach the same goal). The second principle comes from the field of active learning, where exploration strategies are conceived as an optimization process. Samples in the input space (i.e motor space) are collected in order to minimize a given property of the learning process, e.g the uncertainty (Cohn et al., 1996) or the prediction error (Thrun, 1995) of the model. This allows the agent to focus on parts of the sensorimotor space in which exploration is supposed to improve the quality of the model. In [59], we have shown how an integrating probabilistic framework allows to model several recent algorithmic architectures for exploration based on these two principles, and compare the efficiency of various exploration strategies to learn how to uniformly cover a goal space. This was published in [59].

#### 6.3.1.6. *Reusing Motor Commands to Learn Object Interaction*

**Participants:** Fabien Benureau [correspondant], Pierre-Yves Oudeyer.

We have proposed the Reuse algorithm, that exploit data produced during the exploration of an first environment to efficiently bootstrap the exploration of second, different but related environment. The effect of the Reuse algorithm is to produce a high diversity of effects early during exploration. The algorithm only constrains the environments to share the same motor space, and makes no assumptions about learning algorithms or sensory modalities. We have illustrated our algorithm on a 6-joints robotic arm interacting with a virtual object, and showed that our algorithm is robust to dissimilar environments, and significantly improves the early exploration of similar ones. This was published in [34].

#### 6.3.1.7. *Socially Guided Intrinsic Motivation for Robot Learning of Motor Skills*

**Participants:** Mai Nguyen [correspondant], Pierre-Yves Oudeyer.

We have presented a technical approach to robot learning of motor skills which combines active intrinsically motivated learning with imitation learning. Our architecture, called SGIM-D, allows efficient learning of high-dimensional continuous sensorimotor inverse models in robots, and in particular learns distributions of parameterised motor policies that solve a corresponding distribution of parameterised goals/tasks. This is made possible by the technical integration of imitation learning techniques within an algorithm for learning inverse models that relies on active goal babbling. In an experiment where a robot arm has to learn to use a flexible fishing line , we have illustrated that SGIM-D efficiently combines the advantages of social learning and intrinsic motivation and benefits from human demonstration properties to learn how to produce varied outcomes in the environment, while developing more precise control policies in large spaces. This was published in [28].

#### 6.3.1.8. *A social learning formalism for learners trying to figure out what a teacher wants them to do*

**Participants:** Thomas Cederborg [correspondant], Pierre-Yves Oudeyer.

We have elaborated a theoretical foundation for approaching the problem of how a learner can infer what a teacher wants it to do through strongly ambiguous interaction or observation. This groups the interpretation of a broad range of information sources under the same theoretical framework. A teacher's motion demonstration, eye gaze during a reproduction attempt, pushes of good/bad buttons and speech comment are all treated as specific instances of the same general class of information sources. These sources all provide (partially and ambiguously) information about what the teacher wants the learner to do, and all need to be interpreted concurrently. We introduce a formalism to address this challenge, which allows us to consider various strands of previous research as different related facets of a single generalized problem. In turn, this allows us to identify important new avenues for research. To sketch these new directions, several learning setups were introduced, and algorithmic structures are introduced to illustrate some of the practical problems that must be overcome. This was published in [26].

### 6.3.2. Task learning from social guidance

#### 6.3.2.1. Inverse Reinforcement Learning in Relational Domains

**Participants:** Thibaut Munzer [correspondant], Bilal Piot, Mathieu Geist, Olivier Pietquin, Manuel Lopes.

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.

## 6.4. Autonomous And Social Perceptual Learning

### 6.4.1. Unsupervised and online non-stationary obstacle discovery and modelling using a laser range finder

**Participants:** Guillaume Duceux, David Filliat [correspondant].

Recognizing objects is an important capability for assistance robots, but most methods rely on vision and a heavy training procedures to be able to recognize some objects. Using laser range finders has shown its efficiency to perform mapping and navigation for mobile robots. However, most of existing methods assume a mostly static world and filter away dynamic aspects while those dynamic aspects are often caused by non-stationary objects which may be important for the robot task. We propose an approach that makes it possible to detect, learn and recognize these objects through a multi-view model, using only a planar laser range finder. We show using a supervised approach that despite the limited information provided by the sensor, it is possible to recognize efficiently up to 22 different object, with a low computing cost while taking advantage of the large field of view of the sensor. We also propose an online, incremental and unsupervised approach that make it possible to continuously discover and learn all kind of dynamic elements encountered by the robot including people and objects. These results have been published at the IROS conference [40].

### 6.4.2. Task oriented representations by discriminative modulation of a generative learning method

**Participants:** Mathieu Lefort, Alexander Geppert [correspondant].

PROPRE (which stands for PROjection - PREdiction) is a generic and modular unsupervised neural learning paradigm that extracts meaningful concepts of multiple data flows based on predictability across stimuli. It consists on the combination of three modules. First, a topological projection of each data flow on a self-organizing map. Second, a decentralized prediction of each projection activity from each other map activities. Third, a predictability measure that quantifies the prediction error. This measure is used to modulate the projection learning so that to favor the mapping of predictable stimuli across data flows. This model was applied to the visual supervised classification of the pedestrian orientation. The modulation of the visual representation learning by the predictability measure (quantifying the ability to detect the orientation of the pedestrian) improves significantly classification performances of the system independently of the predictability measure used [55]. Moreover, PROPRE provides a combination of interesting functional properties, such as online and incremental learning [56].

#### **6.4.3. Learning of multimodal representations based on the self-evaluation of their predictability power**

**Participants:** Mathieu Lefort, Thomas Kopinski, Alexander Gepperth [correspondant].

PROPRE paradigm (see section 6.4.2 ) was also applied to the classification of gestures caught from two time-of-flight (ToF) cameras. In this context, the predictability measure acts as a self-evaluation module that biases the learned representations towards stimuli correlated across modalities, i.e. related to the ability of one camera to predict the other one. We show in [57] that this unsupervised multimodal representations learning improves the gesture recognition performance, compared to isolated camera representations learning, even not as much as supervised one.

#### **6.4.4. Resource-efficient online learning of classification and regression tasks**

**Participants:** Mathieu Lefort, Thomas Kopinski, Thomas Hecht, Alexander Gepperth [correspondant].

This activity investigates the coupling of generative and discriminative learning (SOM and regression) to achieve incremental learning that stays resource-efficient when the number of input and output dimensions is high. On the one hand, we apply this technique to sensory classification problems where input dimensionalities can exceed 10000 in the presence of multiple categories. On the other hand, we target the learning of forward and inverse regression models for robotics, possibly combining proprioceptive with sensory information which again leads to high data dimensionality. A special kind of regression task we consider in this context is optimal integration of sensory information, where the most likely underlying value must be inferred from several noisy sensor readings. In contrast to popular approaches like XCF or LWPR, our approach achieves efficiency by avoiding a precise partitioning of the input space, relying on a dimensionality-reduced topological projection of the input space instead. While this achieves slightly inferior results on standard benchmarks, we can treat high-dimensional incremental learning problems that are inaccessible to other algorithms, and especially to LWPR. This activity has resulted in two submissions to ESANN 2015 and one to IEEE Transactions on Autonomous Mental Development.

#### **6.4.5. Indoor semantic mapping on a mobile robot**

**Participants:** Louis-Charles Caron [correspondant], Alexander Gepperth, David Filliat.

Semantic mapping is the act of storing high-level information in a persistent map of the environment. The semantic information considered here is the identity of objects encountered by a mobile robot in an indoor environment [35]. The robot runs a SLAM algorithm and builds a map using a laser range finder. The semantic information is collected by analysing the point cloud provided by an RGB-D camera mounted on the robot. The choice of features used to describe the objects, the type of fusion and the recognition algorithm influence the overall capacity of the algorithm. Shape features perform very well, but are blind to changes in color. The fusion of different types of features can reduce the recognition rates on some objects but increases the overall figure. This increase is more significant as the number of objects to recognize gets larger [36]. After running the object recognition algorithm, the identity of the objects is stored alongside the map. The stored information influences future recognition attempts on objects that were already seen by the robot to improve the recognition process. A 3-d map along with a snapshot and the identity of each object seen is displayed to a user.

## 6.5. Robot Design And Morphological Computation

### 6.5.1. Rapid morphological exploration with the Poppy humanoid platform.

**Participants:** Matthieu Lapeyre [correspondant], Steve N’Guyen, Alexandre Le Falher, Pierre-Yves Oudeyer.

In the paper [53], we discuss the motivation and challenges raised by the desire to consider the morphology as an experimental variable on real robotic platforms as well as allowing reproducibility and diffusion of research results in the scientific community. In this context, we present an alternative design and production methodology that we have applied to the conception of Poppy humanoid, the first complete 3D printed open-source and open-hardware humanoid robot. Robust and accessible, it allows exploring quickly and easily the fabrication, the programming and the experimentation of various robotic morphologies. Both hardware and software are open-source, and a web platform allows interdisciplinary contributions, sharing and collaborations. Finally we conduct an experiment to explore the impact of four different foot morphologies on the robot’s dynamic when it makes a footstep. We show that such experimentation can easily be achieved and shared in couple of days at almost no cost.

## 6.6. Educational Technologies

### 6.6.1. 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 than 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. [69], [37], [38], [39]

### 6.6.2. Education and the Poppy project

**Participants:** Matthieu Lapeyre [correspondant], Pierre-Yves Oudeyer, Didier Roy.

The Poppy platform was initially designed for research purposes and even more specifically for studying biped locomotion and human-robot interaction. However, it has been designed with open science goals in mind, both to share our research and create tools for researchers. As we are convinced of the need for multidisciplinary contributions in order to improve the state of the art in the robotics field, we decided right from the beginning to use and create modern and easy-to-use tools. This choice has strongly affected the way we designed our platform. Indeed, being simple to use, easily reproducible and hackable, modular, 3D printable and as plug ’n play as possible lead to the development of hardware (Poppy) and software (pypot) tools that can be also used by non-expert people.

Thus Poppy meets a growing societal need: education and training in technologies combining computer science, electronics and mechanics, as well as a training tool for the emergent revolutionary 3D printing process. Since October 2013 (open source release), we have been contacted by several Fablabs, universities, engineering schools and even high schools. We have had the opportunity to meet with educational teams and it appears they are looking for new motivational tools for group projects.

In this context, the Poppy platform appears well suited. Indeed, it integrates advanced and yet easily accessible techniques (3D printing, Arduino, Python) in an embodiment that motivates students and the wider public. With its openness, design and rather low-cost, Poppy is highly hackable and provides a unique context for learning and experimenting with these technologies in a Do-It-Yourself (DIY) way.

The paper [54] describes the use of the Poppy platform as a tool for scientific researches as well as educational and artistic applications.

Several experiments with Poppy in middle and high schools, science museums and Fablabs in France and abroad are already underway and will be discussed in the upcoming **Partnerships and Cooperations** sections.

### 6.6.3. Expression of emotions with Poppy Humanoid

**Participants:** Fabien Benureau [correspondant], Matthieu Lapeyre.

Two students in 3rd year of the Cognitive Science major at the University of Bordeaux led a TER project this year using Poppy under the supervision of Fabien Benureau, exploring how the attitude towards robots influences how humans recognise the emotion they try to express. Poppy having no facial expression — or face — yet, the students expressed the five expressions they selected (anger, surprise, joy, sadness, disgust) with body movements alone. They videotaped the sequences of movements (videos are available here [http://python.sm.u-bordeaux2.fr/ter/2014/sc/desprez-zerdoumi/?page\\_id=289](http://python.sm.u-bordeaux2.fr/ter/2014/sc/desprez-zerdoumi/?page_id=289)) and created an experiment asking volunteers to guess which emotion was displayed. The form also included the Negative Attitude towards Robots Scale (NARS), to investigate the possible correlation between fear of robot and the ability to identify their emotional attitude. The results showed no correlation between the two, although it was admitted that the experiment would have to be improved and ran again before any conclusion could be made.

## 6.7. Interactive Learning and user adaptation

### 6.7.1. Interactive learning from unlabeled instructions

**Participants:** Grizou Jonathan [correspondant], Itturate Inaki, Montesano Luis, Pierre-Yves Oudeyer, Manuel Lopes.

Interactive learning deals with the problem of learning and solving tasks using human instructions. It is common in human-robot interaction, tutoring systems, and in human-computer interfaces such as brain-computer ones. In most cases, learning these tasks is possible because the signals are predefined or an ad-hoc calibration procedure allows to map signals to specific meanings. In this work, we addressed the problem of simultaneously solving a task under human feedback and learning the associated meanings of the feedback signals. This has important practical application since the user can start controlling a device from scratch, without the need of an expert to define the meaning of signals or carrying out a calibration phase. We proposed an algorithm that simultaneously assign meanings to signals while solving a sequential task under the assumption that both, human and machine, share the same a priori on the possible instruction meanings and the possible tasks. This work was published in a conference paper [45] and a journal paper will be submitted in January 2015.

We communicated about this work to the human-robot interaction (HRI) community. A robot equipped with our algorithm would be able to interact with a human without knowing in advance the specific communicative signals used by the human. This work was published in the HRI Pioneer workshop [46].

This work was presented during the thesis defense of Jonathan Grizou entitled: Learning from Unlabeled Interaction Frames, on October 24, 2014. The video, slides, and thesis manuscript can be found at: <http://jgrizou.com/projects/thesis-defense/>



### 6.7.2. Calibration-Free BCI Based Control

**Participants:** Grizou Jonathan [correspondant], Itturate Inaki, Montesano Luis, Pierre-Yves Oudeyer, Manuel Lopes.

We applied previous work on interactive learning from unlabeled instructions [45] to Brain-Machine Interaction problem, leading to a Calibration-Free brain computer interfaces. So far in such brain-computer interfaces (BCI), an explicit calibration phase was required to build a decoder that translates raw electroencephalography signals from the brain of each user into meaningful instructions. Our method removes the calibration phase, and allows a user to control a device to solve a sequential task. We performed experiments where four users use BCI to control an agent on a virtual world to reach a target without any previous calibration process. Our approach is promising for the deployments of BCI applications out of the labs. This work was published in a conference paper [44] and a journal paper will be submitted in January 2015.

This work was presented during the thesis defense of Jonathan Grizou entitled: Learning from Unlabeled Interaction Frames, on October 24, 2014. The video, slides, and thesis manuscript can be found at: <http://jgrizou.com/projects/thesis-defense/>

## 6.8. Studying the Co-Construction of Interaction Protocols in Collaborative Tasks with Humans

### 6.8.1. Experimental Setups for User Study of Alignment in Asymmetric Interactions

**Participants:** Anna-Lisa Vollmer [correspondant], Jonathan Grizou, Manuel Lopes, Katharina Rohlfing, Pierre-Yves Oudeyer.

In interaction, humans align and effortlessly create common ground in communication, allowing efficient collaboration in widely diverse contexts. Robots are still far away from being able to adapt in such a flexible manner with non-expert humans to complete collaborative tasks. Challenges include the capability to understand unknown feedback or guidance signals, to make sense of what they refer to depending on their timing and context, and to agree on how to organize the interaction into roles and turns.

As a first step in approaching this issue, we investigated the processes used by humans to negotiate a protocol of interaction when they do not already share one. We developed a new experimental setup, where two humans have to collaborate to solve a task. The channels of communication they can use are constrained and force them to invent and agree on a shared interaction protocol in order to solve the task. These constraints allow us to analyze how a communication protocol is progressively established through the interplay and history of individual actions.

We consider a remote construction task, where one user (user A) knows what to build but do not have access to the construction site while its partner (user B) is at the site but do not know what to do. By constraining the communicative channel between the two partners, we study how, and if, they will agree on a similar set of signals to convey information and what type of information they tend to produce. The experimental setup consist of box with button, a video recording system and two screens. User A can send signals to user B by pressing buttons (fig. 14 ). Signals are displayed on a screen (fig. 14 ) at user B side. User A is not aware of what is displayed on user B screen, neither user B is aware of the relation between button presses and screen events. The video of user B construction scene is streamed to a screen at user B side. The task consist of bulding arbitrary construction (fig. 14 ) using colored toy bricks (fig. 14 ).

The various data recolted during these interaction sequences (fig. 15 ) allow us to study the Co-Construction of Interaction Protocols. This work was published in a conference paper [66].

## 6.9. Other

### 6.9.1. A Framework for Proactive Assistance

**Participants:** Alexandre Armand, David Filliat [correspondant].



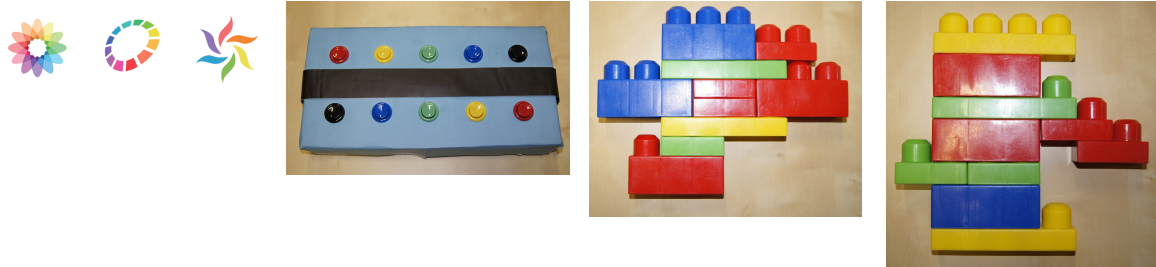


Figure 14. Three examples of sign displayed on the learner screen; The box and the button use as an interface for the teacher to communicate with the learner; Examples of construction presented to the teacher.

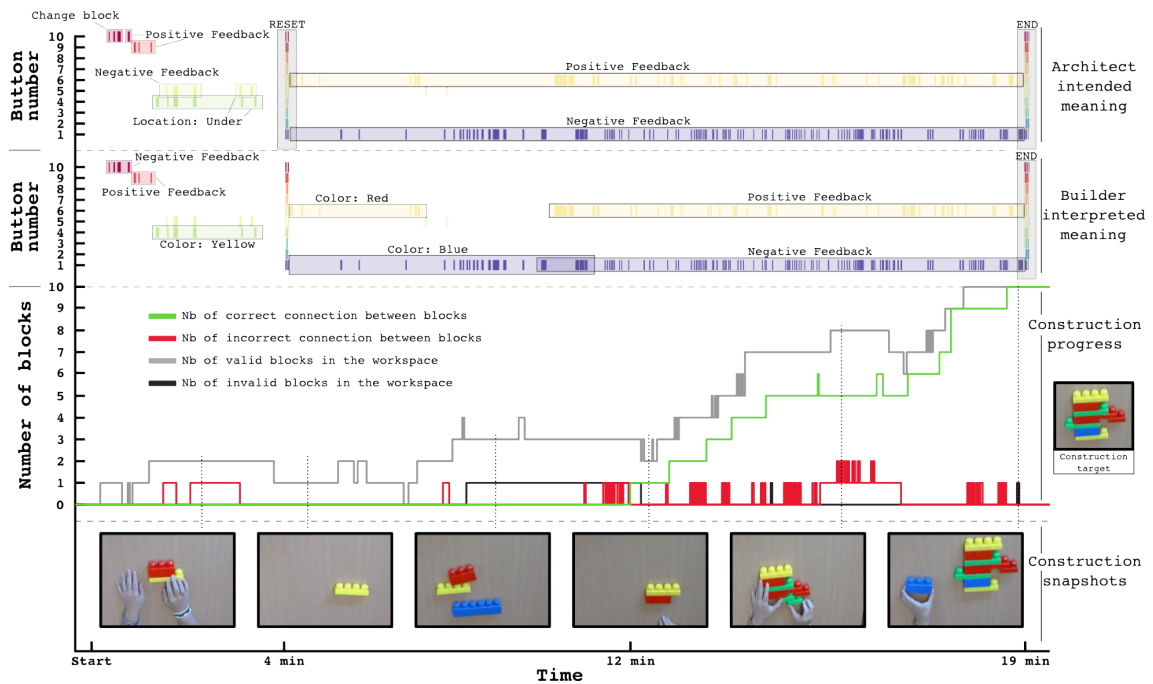


Figure 15. Timeline for one experiment of an architect and a builder collaborating towards building the construction target (right hand side). The top and middle part show the timeline of button presses associated with the intended meaning from the architect (top) and the understood meaning from the builder (middle).

We worked in collaboration with Renault on the problems of adapting driving assistance systems by learning individual drivers behaviours and of integrating more advanced perception in these systems. Advanced Driving Assistance Systems usually provide assistance to drivers only once a high risk situation has been detected. Indeed, it is difficult for an embedded system to understand driving situations, and to predict early enough that it is to become uncomfortable or dangerous. Most of ADAS work assume that interactions between road entities do not exist (or are limited), and that all drivers react in the same manner in similar conditions. We propose a framework that enables to fill these gaps. On one hand, an ontology which is a conceptual description of entities present in driving spaces is used to understand how all the perceived entities interact together with the subject vehicle, and govern its behavior. On the other hand, a dynamic Bayesian Network enables to estimate the driver situation awareness with regard to the perceived objects, based on the ontology inferences, map information, driver actuation and learned driving style. This work was published in a workshop [33] and a conference paper [32].

## HEPHAISTOS Team

## 6. New Results

### 6.1. Highlights of the Year

Yves Papegay received a "Wolfram Innovator Award" in December 2014

### 6.2. Robotics

#### 6.2.1. Cable-driven parallel robots (CDPR)

##### 6.2.1.1. 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 leg lengths. This forward kinematic problem (FK) is usually very complex and admits several solutions. For parallel robot with rigid legs we have established the important property (P) that the FK may be solved in real-time i.e. being given the leg lengths  $\rho$  and platform pose  $\mathbf{X}$  it is possible to determine the single pose  $\mathbf{X}_1$  that can be reached from  $\mathbf{X}$  if the leg lengths has been changed to  $\rho + \Delta\rho$  provided that  $\Delta\rho$  satisfies some properties. For CDPR with sagging cables determining all the FK solutions is more complex but we have proposed the first algorithm to solve it the for a full scale model of sagging cables [24]. For CDPR with non sagging cables the problem is also very complex because we cannot make any assumption on the number of cables under tension i.e for a CDPR with  $m$  cables we have to solve all the FK problems for all possible set of cables under tension from 1 to  $m$  and as soon as this number is lower than 6 the system of equations is much larger than for classical robots. We have however been able to propose an interval analysis based algorithm that allow one to get all the solutions [18]. But we have also shown that for non sagging cables the property (P) does not hold. Indeed it requires that the system of equations that governs the FK remains the same at all time. But for CDPR this system depend on the set of cables under tension (which is called the *cable configuration CC*) and it may change when the cable lengths change from  $\rho$  to  $\rho + \Delta\rho$ , even for redundant CDPR [23]. If the CC changes at some point the pose solution of the FK together with the cable tensions will differ from the one that is obtained when assuming no change in the CC. This has a drastic effect on control as we have now a system whose state equations may change over time but also on design as in the new CC the cable tensions may be quite different from the expected one. Hence property (P) will hold if and only if we are able to show that there will not be any change in the CC during the change of the cable lengths and therefore it is crucial to detect CC changes. But this require to fully simulate the discrete-time control laws together with the behavior of the coiling system. We have been able to implement a simulation tool that tracks a trajectory for the robot for arbitrary control laws and coiling system model [22], [25]. The principle of the algorithm is to determine if on a time interval  $[t, t + \delta t]$  the solution of the FK with the current CC satisfies (P) by using Kantorovitch theorem. If this is not the case  $\Delta t$  is divided by 2 and the process is repeated. We then check if there is a time  $t_1$  in  $[t, t + \delta t]$  for which the tension of a cable in the CC may become 0. If there is no such  $t_1$  for any cable in the current CC, then it will be the CC at time  $t + \delta t$  and we may compute the pose and cable tensions at any time in  $[t, t + \delta t]$ . If there is a least one such  $t_1$  (and there may be several  $t_1^i, t_1^j, \dots$  as we consider each cable in the CC) we order these times by increasing values and check sequentially if a cable tension become negative with the current CC at time  $(t_1^i + t_1^{i+1})/2$ . If yes we determine what can the CC at this time by looking at all possible CC. As soon as the new CC at time  $t_1^i$  has been determinated the simulation can go on. Implementing this algorithm has been difficult mainly for numerical reasons: the accuracy of the calculation may sometime exceed the floating point accuracy and we have to resort to symbolic computation and extended arithmetics. Our tests have shown that indeed CC changes may occur on trajectories: on a typical trajectory up to 10 different CC will appear with 5 or 6 cables under tension. These results have been confirmed experimentally on a prototype at LIRMM.

### 6.2.1.2. Tool for Agencement Analysis and Synthesis of CDPRs

**Participants:** Laurent Blanchet, Jean-Pierre Merlet [correspondant].

HEPHAISTOS has been working on tools to design the layout and geometry of CDPRs, while accounting for numerical errors as well as practical errors – actual position of the winches, of the attachments on the platform, errors of the controllers, of the cables, etc. Within this work, collision analysis plays an important role. Indeed the concept of cable robot aims to increase the workspace that is restricted for robots having rigid legs but interferences may reduce this workspace. Two types of interference analysis approaches exist: intersection of numerically-mapped boundaries (InB) and distance between features (DbF). The two sets of interference types that can be analysed using these approaches are distinct but overlapping. The first approach greatly benefits from Inria’s computational geometry research and particularly from the AABB tree algorithms implemented in CGAL. Algorithms and implementation based on those were developed, along with several new algorithm and implementation to extend the scope of intersection types, and thus, of interference types. Algorithms to improve efficiency of given intersection types were also developed. We have already used the second approach, DbF, to develop algorithms for leg interference of parallel robots that are very efficient for non deformable cables but now well adapted for sagging cables. An interference detection algorithm has been developed and implemented for a restricted scope of applications [10], and research is on-going for a more generic case.

### 6.2.1.3. Visual-servoing of a parallel cable-driven robot

**Participants:** Rémy Ramadour, Jean-Pierre Merlet [correspondant], François Chaumette [correspondant].

The last two years, we studied how visual servoing could improve accuracy, controllability and performance of cable-driven parallel robots [13]. Previous works on this domain showed very interesting results but some issues remained to be investigated, such as :

- ratio accuracy/workspace : cable-driven parallel robots are known to allow a large reachable workspace, but also to have complex geometric and dynamic models which affect the accuracy. Using visual-servoing in a closed-loop scheme, we were able to enhance the accuracy by a factor of ten, allowing to manipulate daily-life objects in a whole living room.
- image-based joint-space control : in order to reach a desired pose, the usual method involves several computing and evaluations of both the Jacobian matrix of the manipulator and the interaction matrix linking visual features to the displacements of the end-effector. We designed a control scheme, based on an iterative updating using the Broyden update law, in order to link the visual features directly to the joint coordinates. This scheme is less sensitive to model uncertainties and require much less computing.
- stability of the command law : classical control laws ignore cable configuration effects that change the pose of the platform. We have proposed a counter-intuitive strategy: the robot MARIONET-ASSIST we are using has a specific geometry that allow to predict which cables set may be under tension for a given trajectory i.e. we are able to split the trajectory in parts for which we know all possible cables configurations. Among them we select the one that optimize an accuracy criteria and we enforce it by forcing the cables not part of the configuration to be slack by adding a sufficient amount of length to their nominal values. It allowed to enhance both the stability and the accuracy of a vision-based control scheme [26].

We also used interval analysis in order to guaranty every step of the process, in order to provide safety and reliability of our methods, as the robots that we use were initially deployed in the context of assistive technologies.

Finally, simulations and experiments on prototypes were conducted and presented in order to validate the mentioned results. However, the prototype that we used presents a very particular configuration (all wires are connected to the same point on the end-effector, allowing only translational movements), further works may be required in order to test our methods for a wider variety of cable-driven parallel robots.

### 6.2.1.4. 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 aim to design and realize a CDPR of large size as a proof of concept in additive manufacturing of building based on ultra-high performance concrete.

Challenges are modeling and control to get enough accuracy.

### 6.2.2. Assistance robotics

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 [20], [21], [19].

#### 6.2.2.1. 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  $\tau$  of the arm and the maximal force  $F$  that can be exerted by the hand, which are related by the equation

$$\tau = \mathbf{J}^T F \quad (2)$$

where  $\mathbf{J}$  is a matrix which depends only upon the configuration of the arm. This equations constitutes an underconstrained linear system. In biomechanics the torque  $\tau$  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  $\mathbf{J}^T$  to calculate  $F$ . But there are several uncertainties that are neglected when using this method: the measurement errors on  $\tau$  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 ( $\tau$  is an interval vector  $\tau_m$ ,  $\mathbf{J}^T$  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  $\tau_v$  for the  $\tau$  that is guaranteed to include the real  $\tau$ . Furthermore  $\tau$  must be included in the intersection  $\tau_i$  of  $\tau_v$  and  $\tau_m$  while  $F$  must be included in the intersection  $F_i$  of  $F_m$  and  $F_s$ . If  $\tau_i$  is strictly included in  $\tau_m$ , 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  $\tau_v$ . If one of these situation occurs we repeat the process until no significant improvement of  $F_s$  or  $\tau_v$  is obtained. In a second step we consider that the uncertainties that lead to uncertainties in the matrix  $\mathbf{J}^T$  are constrained as we have to satisfy  $\tau_v = \mathbf{J}^T F_s$ . 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  $\mathbf{J}^T$  that is included in the initial one. In turn this may allow to obtain better  $\tau_v$  and  $F_s$ . The process stops when no improvement has been obtained for  $F_s$ ,  $\tau_v$  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.

#### 6.2.2.2. Walking analysis

**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 analyse the walking of the user. We have used this property for performing until mid 2013 a large scale experiment: 23 young adults and 25 elderly people (> 69 years) were asked to walk along with two reference trajectories with the help of the walker. The objective of this research is to develop walking quality index and examine if the walker may be used to monitor the health state of elderly people at home. We compared and statistically analyzed the walking patterns of the two groups of people. The results show that it is possible to obtain new indicators by using the walker measurements [9],[14]. Next step will be to perform a similar analysis for a sit-to-stand (STS) exercise and to test our approach in two rehabilitation centers, MATIA in Spain (in the framework of the RAPP project) and Centre Héliomarin de Vallauris.

A start-up plan was proposed in November 2014 to transfer the walking analysis technology of Hephaistos with the ANG walker. In order to study the feasibility of our plan, we have interviewed Patrick Nenert (Kiné, Centre Hélio-Marin), Françoise Dubourgeois (DR, EPHAD) and Sophie Morgenstern (Métropole NCA, Living Lab Paillon 2020) about their impression of the walker and the possibility of the future collaboration with them. Several contact with local actors of the silver economy sector have already been established : Livinglab Paillon2020 (Nice), CIU-santé, as well as with research lab for collaboration on future projects (Lapcos, I3M, Gredeg).

#### 6.2.2.3. *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.

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 plan to organize a national forum on this topic with Nathalie Nevejans from University of Douai.

## 6.3. Miscellaneous results

### 6.3.1. *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.

Technology demonstrated by our prototype has been transferred to our industrial partner in 2012 when 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.

In 2014, we have worked again 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

# 6. New Results

## 6.1. Visual servoing

### 6.1.1. *Photometric moment-based visual servoing*

**Participants:** Manikandan Bakthavatchalam, François Chaumette.

The goal of this work is to determine an adequate set of visual features to control the six degrees of freedom of a dynamic system. Thanks to a collaboration with Omar Tahri from Le2I in Le Creusot, we have been able to improve the results obtained previously with shifted moments for increasing the stability domain of visual servoing [24].

### 6.1.2. *Histogram-based visual servoing*

**Participants:** Quentin Bateau, Eric Marchand.

Classically visual servoing considered the regulation in the image of a set of visual features (usually geometric features). Recently direct visual servoing schemes, such as photometric visual servoing, have been introduced in order to consider the image as a whole and thus avoid the extraction and the tracking of such geometric features. In this preliminary work, we propose a method to extend direct visual servoing approaches by using a global descriptor, namely intensity histograms, on the whole or multiple sub-sets of the images in order to achieve the control of a 6 degrees of freedom (DoF) robot.

### 6.1.3. *Predictive visual servoing*

**Participants:** Nicolas Cazy, Paolo Robuffo Giordano, François Chaumette.

This study is devoted to the application of predictive control to visual servoing. In a first step, we have developed and compared several predictive models that can be useful when some visual features leave the camera field of view or are lost because of occlusions [25].

### 6.1.4. *Visual servoing of cable-driven parallel robot*

**Participant:** François Chaumette.

This study is realized in collaboration with Rémy Ramadour and Jean-Pierre Merlet from Coprin group at Inria Sophia Antipolis. Its goal is to adapt visual servoing techniques for cable-driven parallel robot in order to achieve accurate manipulation tasks [46]. This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6).

### 6.1.5. *Nanomanipulation*

**Participants:** Le Cui, Eric Marchand.

We began a work, within the ANR P2N Nanorobust project (see Section 8.2.1), on the development of micro- and nano-manipulation within SEM (Scanning Electron Microscope). Our goal is to provide visual servoing techniques for positioning and manipulation tasks with a nanometer precision. This year, we focused on the characterisation of the projection model of a SEM along with the approach required for its calibration [26]. We then address the problem of 6 dof control using photometric feature under an optical microscope [27]. Finally, we focused on the definition of control law able to control the motion along the Z axes with a SEM microscope. Indeed, considering that a SEM is subject to parallel projection model, motion along this axis is not observable. We then address this problem using defocus information. An autofocus process has also been studied.

### 6.1.6. 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 closed loop control schemes to audio sensors. It is made in collaboration with Nancy Bertin from Panama group at 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.

## 6.2. Visual navigation of mobile robots

### 6.2.1. Automous navigation of wheelchairs

**Participants:** Vishnu Karakkat Narayanan, François Pasteau, Marie Babel, François Chaumette.

The goal of this work is to design an autonomous navigation framework of a wheelchair by means of a single camera and visual servoing. We first focused on a corridor following task where no prior knowledge of the environment is required. The servoing process matches the non-holonomic constraints of the wheelchair and relies on two visual features, namely the vanishing point location and the orientation of the median line formed by the straight lines related to the bottom of the walls. This overcomes the initialization issue typically raised in the literature [9]. The control scheme has been implemented onto a robotized wheelchair and results show that it can follow a corridor with an accuracy of  $\pm 3$  cm [16].

We then focused on a door passing task [33]. This doorway passing (and corridor turning) task employs the position of a single doorpost in the image as an input to a Lyapunov-based control scheme which allows the wheelchair to take up a desired trajectory about the doorpost. This trajectory avoids collision with the wall and guarantees that the wheelchair positions itself in front of the doorway regardless of its initial position. Results in simulation demonstrate the convergence and robustness of both control schemes. Experiments conducted on a wheelchair indicate the validity of applying the proposed low-level control system [16].

This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6 ) as well as of the Apath project (see Section 8.1.1 ).

### 6.2.2. Semi-autonomous control of a wheelchair for navigation assistance along corridors

**Participants:** Vishnu Karakkat Narayanan, Marie Babel, François Pasteau, Alexandre Krupa.

This study concerns a semi-autonomous control approach that we designed for safe wheelchair navigation. The control relies on the combination of primary tasks of wall avoidance as well as door passing performed by a dedicated visual servoing framework and a manual steering task. A smooth transition from manual driving to assisted navigation is obtained thanks to a gradual visual servoing activation method that guarantees the continuity of the control law. The visual servoing task is then progressively activated, when the wheelchair gets closer to the walls or doorposts, in order to avoid collisions [43]. Experimental results clearly show the ability of the approach to provide an efficient solution for wall avoiding and doorway passing purposes [58]. This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6 ) as well as of the Apath project (see Section 8.1.1 ). Tests with disabled patients in the rehabilitation center Pôle Saint Héliier (Rennes) are under progress and first results prove the ability of our system to smoothly correct the trajectory of the wheelchair in case of hazardous situations.

Current research works are oriented towards multimodal sensor-based servoing, as well as haptic feedback that leads to an intuitive assistive wheelchair navigation. This work is realized in collaboration with Maud Marchal (Hybrid team). In addition, we are currently working with e-Motion team to design a vision-based human-aware semi-autonomous navigation system.

### 6.2.3. Social Spacing and human-robot interaction

**Participants:** Panagiotis Papadakis, Patrick Rives.

A novel probabilistic framework was introduced capable of instantiating diverse models of social spacing and accounting for distinctive dimensions in human-robot interaction, namely, perception capacity and certainty [42]. We have concretely shown how our method allows smooth adaptation in the situation awareness of a robot within common human-robot interaction examples and further showed its utility at the level of path planning by adapting trajectories to social sensitivity levels.

This approach is currently extended to take into account human activity cues in order to enhance robot mapping and navigation and in particular in filtering noisy human detections, detecting passages such as doors and staircases, inferring space occupancy and allowing navigation within unexplored areas.

#### 6.2.4. *Target tracking*

**Participants:** Ivan Markovic, François Chaumette.

This study was realized in the scope of the FP7 Regpot Across project (see Section 8.3.1.2) during the three-month visit of Ivan Markovic, Ph.D. student at the University of Zagreb. It consisted in developing a pedestrian visual tracking from an omni-directional fish-eye camera and a visual servoing control scheme so that a mobile robot is able to follow the pedestrian [37]. This study has been validated on our Pioneer robot (see Section 5.4.2).

#### 6.2.5. *Obstacle avoidance*

**Participants:** Suman Raj Bista, Fabien Spindler, François Chaumette.

This study was realized in collaboration with Andrea Cherubini who is now Assistant Prof. at Université de Montpellier. It is concerned with our long term researches about visual navigation from a visual memory without any accurate 3D localization [4]. In order to deal with obstacle avoidance while preserving the visibility in the visual memory, we have proposed a control scheme based on tentacles for fusing the data provided by a pan-tilt camera and a laser range sensor [11]. A new study devoted to indoors navigation from segments has started recently.

### 6.3. Visual tracking and state estimation

#### 6.3.1. *3D model-based tracking*

**Participant:** Eric Marchand.

This study focused on the issue of estimating the complete 3D pose of the camera with respect to a potentially textureless object, through model-based tracking. We proposed to robustly combine complementary geometrical and color edge-based features in the minimization process, and to integrate a multiple-hypotheses framework in the geometrical edge-based registration phase [45]. This method will be tested in the scope of the FP7 RemoveDebris project [36].

#### 6.3.2. *Pose estimation through plane tracking*

**Participants:** Aurélien Yol, Eric Marchand.

We proposed a method for localizing an Unmanned Aerial Vehicle (UAV) using georeferenced aerial images. Here we provide a multiple usage localization algorithm based on vision only. To ensure robustness, we choose to use the Mutual Information (MI) within a dense tracking process. MI proved to be very robust toward local and global scene variations. However, dense approaches are often related to drift disadvantages. We solve this problem by using georeferenced images. The localization algorithm has been demonstrated through the localization of a hexarotor UAV fitted with a downward looking camera during real flight tests [53].

#### 6.3.3. *3D tracking of deformable objects*

**Participants:** Bertrand Delabarre, Eric Marchand.

We consider the problem of dense non-rigid visual tracking robust towards global illumination perturbations of the observed scene. The similarity function is based on the sum of conditional variance (SCV). With respect to most approaches that minimize the sum of squared differences, which is poorly robust towards illumination variations in the scene, the choice of SCV as our registration function allows the approach to be naturally robust towards global perturbations. Moreover, a thin-plate spline warping function is considered in order to take into account deformations of the observed template [28].

#### 6.3.4. 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 [18], we showed how to design an online active SfM scheme characterized by an error transient response equivalent to that of a reference linear second-order system with desired poles. Indeed, in a nonlinear context, the observability properties of the states under consideration are not (in general) time-invariant but may depend on the current state and on the current inputs applied to the system. It is then possible to simultaneously act on the estimation gains and system inputs (i.e., the camera velocity for SfM) in order to optimize the observation process and impose a desired transient response to the estimation error. The theory has a general validity and can be applied to many different contexts such as when dealing with point features [18], solid objects like spheres or cylinders [51], or planar regions [47]. Furthermore, the active SfM scheme can also be embedded within a classical visual servoing law exploiting the redundancy of the camera motion w.r.t. the considered visual task [52].

#### 6.3.5. Robust visual odometry

**Participants:** Tawsif Gokhool, Patrick Rives, Renato José Martins.

Our aim is concentrated around building ego-centric topometric maps represented as a graph of salient keyframe nodes [14]. Additionally, visual odometry from frame to keyframe alignment helps significantly in drift reduction. On the other hand, the sparsity in this kind of graphical representation leads to reduced overlapping between keyframes which can degrade localisation robustness. Our chosen spherical 360° field of view (FOV) configuration alleviates the overlapping issue by providing an enriched model of the environment with photometric and geometric information content. Following a multitude of advantages with information fusion, merging of frames in a single representation deals with the problem of data redundancy and sensor noise suppression.

Therefore, the second fold of this work consisted in improving the identified conceptual loopholes above by first proposing a generic uncertainty propagation model as applied to our spherical RGB-D database. Secondly, a probabilistic framework was derived which led to a Mahalanobis inconsistency test incorporating both geometric and photometric uncertainty models [32]. Our framework was further improved by adding up a probabilistic model to filter out dynamic points temporally. Finally, the entire probabilistic framework was applied in order to track the most stable points over time.

### 6.4. 3D Scene Mapping

#### 6.4.1. New RGB-D sensor design for indoor 3D mapping

**Participants:** Eduardo Fernandez Moral, Patrick Rives.

A multi-sensor device has been developed for omnidirectional RGB-D (color+depth) image acquisition (see Fig. 3 .a). This device allows to acquire such omnidirectional images at high frame rates (30 Hz). This approach has advantages over other alternatives used today in terms of accuracy and real-time spherical image construction for indoor environments, which are specially interesting for mobile robotics. This device has important prospective applications as fast 3D-reconstruction or Slam.

A calibration method for such device was developed [31], which takes into account the bias of each sensor independently. The proposed calibration method does not require any specific calibration pattern, taking into account the planar structure from the scene to cope with the fact that there is no overlapping between sensors.

In a first instance, this sensor has been exploited for localization and mapping research with mobile robots. For that, the sensor is mounted on a mobile platform together with a standard computer (see Fig. 3 .a). A method to perform image registration and visual odometry has been developed. This method relies in the matching of planar primitives that can be efficiently obtained from the depth images. This technique performs considerably faster than previous registration approaches like ICP, or dense photoconsistency alignment. These last achieve however a better accuracy than our method, what suggests that our method can be used as an initial step to speed-up those.

Slam is also addressed with this device. A solution to this problem using our omnidirectional RGB-D sensor is being researched. The ongoing experiments have shown some initial results for metric-topological pose-graph Slam, where the map consists of a set of spherical keyframes, which are located in a topological arrangement according to their shared observations.

#### 6.4.2. Compact 3D scene representation

**Participants:** Renato José Martins, Patrick Rives, Tawsif Gokhool.

This work follows in the direction of precise and compact scene representation of large scale environments. The aim is to build a complete geometric and photometric “minimal” model, which is stored within a sparse set of augmented spherical images to asset photo-geometry consistence of the scene from multiple points-of-views. In this direction, an uncertainty model from the full structure combined with those of poses was proposed for point-to-point egocentric fusion. This model allows to reduce sensor noise in a particular keyframe sphere when performing a multi-frame fusion scheme of coherent near information. This first fusion scheme is then improved by exploiting the rigidity/influence of neighboring points representing the surface. For that, an intermediary higher level abstraction of the point cloud is generated by partitioning the input domain into elementary cells, then reducing the number of degrees of freedom and enforcing constraints over the points segmented as being part of the same surface.

The adopted solution is a “weaker” representation of a 3D boundary mesh, based on discontinuous convex planar patches, with the segmentation being done considering the geometry (region growing) or photometry (SLIC superpixels). This synthetic scene built with the planar geometric police proved to well represent the original scene (for both indoor and outdoor real data) with a significant small amount of patches and it is exploited to build robust useful “dynamic” 4D world model, which in turn can be used for assisted/autonomous navigation or virtual reality applications.

#### 6.4.3. Semantic mapping

**Participants:** Romain Drouilly, Patrick Rives, Panagiotis Papadakis.

Autonomous navigation is one of the most challenging problems 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 [56]. The goal of this work is to study how semantics can be used to improve all the steps of navigation process. In a first time, we have developed a new navigation-oriented hybrid metric-topological-semantic model of the world. It captures high-level information and uses it to build extremely compact description of large environments. Then we have used it to design an efficient localization algorithm, able to find a given map content faster than classical methods and allowing human-understandable queries [30]. In a second time, we have studied how semantics can be used to discover unobserved things in the scene. Particularly, we have shown that both statics and dynamic entities, identified by a robot, can inform about the structure of the environment in unobserved areas [29]. We have used this to do “map extrapolation”, that is extending a map beyond robot’s perceptual limits by reasoning on semantics. This approach has been shown to be of great interest in everyday-life environment. Finally, we have proposed a new scheme for trajectory planing, taking into account not only geometric constraints

but also high-level understanding of the world. We have shown the usefulness of this approach to navigate complex environments with highly dynamic areas on both simulated and real-world datasets, well-suited for large outdoor environment navigation.

#### **6.4.4. Augmented reality**

**Participant:** Eric Marchand.

Using Slam methods becomes more and more common in Augmented Reality (AR). To achieve real-time requirement and to cope with scale factor and the lack of absolute positioning issue, we proposed to decouple the localization and the mapping step. This approach has been validated on an Android Smartphone through a collaboration with Orange Labs [38][39]

### **6.5. Medical robotics**

#### **6.5.1. Non-rigid target tracking in ultrasound images based on hierarchical grid interpolation**

**Participants:** Lucas Royer, Jason Chevie, Marie Babel, Alexandre Krupa.

In order to track the motion of a tumour or cyst during needle insertion, we developed a first approach to track a deformable target within a sequence of 2D ultrasound images. It is based on a dedicated hierarchical grid interpolation algorithm (HGI) that is typically used for real-time video compression purposes. This approach provides a continuous motion representation of the target by using a grid of control points that models both their global displacement and local deformations. The motion of each control point is estimated by a hierarchical and multi-resolution local search method in order to minimize the sum of squared difference of the target pixel intensity between successive images. This approach was validated from 2D ultrasound images of real human tissues undergoing rigid and non-rigid deformations [48] and was recently adapted for tracking 3D deformations.

#### **6.5.2. Non-rigid target tracking in ultrasound images based on physically-based model**

**Participants:** Lucas Royer, Alexandre Krupa.

A second approach for automatically tracking deformable target within 2D ultrasound images has been developed [50]. It 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 second approach was validated on simulated and real data, both for rigid and non-rigid motions of soft tissues [49]. It was recently extended for deformable target tracking in 3D ultrasound volumes.

#### **6.5.3. 3D steering of flexible needle by visual servoing**

**Participants:** Alexandre Krupa, Pierre Chatelain.

The objective of this work is to provide robotic assistance during needle insertion procedures such as biopsy or ablation of localized tumor. A method has been developed for steering a beveled-tip flexible needle actuated by a robotic arm in such a way to control the needle curvature in 3D space [34]. It is based on the design of a new duty-cycling control strategy that makes possible to control both the 2 lateral angular velocities and the insertion velocity of the needle tip (3 DOF). An image-based visual servoing approach has then been developed to automatically position the needle tip on a 3D target indicated by the user. It is based on the use of geometrical visual features extracted from 2 images provided by 2 orthogonal cameras observing a translucent gelatin phantom where the needle is inserted. Preliminary results of this automatic targeting task demonstrate the feasibility of this new concept and its robustness to needle kinematic model errors [35]. We recently extended this approach to automatically steer the needle toward a target by an image-based visual servoing that uses geometrical features extracted from images provided by a 3D ultrasound probe.



#### 6.5.4. Optimization of ultrasound image quality by visual servoing

**Participants:** Pierre Chatelain, Alexandre Krupa.

This study focuses on the automatic positioning of a 2D ultrasound probe in such a way to optimize the quality of the acquired ultrasound images. It is based on 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. In collaboration with Nassir Navab we considered this ultrasound confidence map as a new modality and recently designed a visual servoing control law for image quality optimization.

### 6.6. Control of single and multiple Unmanned Aerial Vehicles

#### 6.6.1. State estimation and flight control of quadrotor UAVs

**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 [17], 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.

This work has been realized in collaboration with the Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

#### 6.6.2. Collective control of multiple UAVs

**Participant:** Paolo Robuffo Giordano.

The challenge of coordinating the actions of multiple robots is inspired by the idea that proper coordination of many simple robots can lead to the 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 [22], 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 [54] the rigidity theory has been extended to the case of *bearing measurements*, and directed graphs.

These works were realized in collaboration with the robotics group at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany and with Technion, Israel.

## RITS Team

# 6. New Results

## 6.1. Highlights of the Year

YoGoKo<sup>0</sup>, a startup company of RITS, was founded in 2014 by employees from three research institutes: Mines ParisTech, Telecom Bretagne and Inria. YoGoKo makes use of softwares developed in teams specialized in Internet technologies. RSM (Telecom Bretagne), CAOR (Mines ParisTech) and RITS (Inria) are research teams have been working together since 2006 on innovative communication solutions applied to Intelligent Transportation Systems. They contributed to several collaborative R& D projects related to ITS (CVIS, ITSSv6, GeoNet, DriveC2X, SCORE@F, ...). In 2012, these laboratories engaged together into the development of a common demonstration platform which comprises connected vehicles (fleet of conventional vehicles from Mines ParisTech and fleet of autonomous vehicles from Inria), roadside equipments and cloud-based services. YoGoKo demonstration platform was finally revealed on Feb. 11 th 2014 during the Mobility2.0 event organized by the French Ministry of Transport. This successful demonstration and the extremely warmfull feedback gained at this occasion triggered the launch of YoGoKo as a company. YoGoKo develops innovative communication solutions for fixed and mobile multi-connected devices. The objective is to maintain secure and continuous connectivity with their communication peers, either in their immediate environment or a remote location (control centers or Internet hosts).

## 6.2. Development of a Platform for Arbitration and Sharing Control Applications

**Participants:** David Gonzalez Bautista, Vicente Milanes Montero, Fawzi Nashashibi, Joshué Pérez Rastelli.

RITS have been leading the activities in the framework DESERVE project, related to arbitration and control sharing in automated vehicle. The analysis of existing vehicle control (and arbitration) solutions, considering the driver in the control loop is the main challenge of this work. We consider sharing control techniques and different solutions in the task management. New standard in the taxonomy of autonomous driving, as the SAE J3016, are considered in the arbitration and sharing control design. The aim is to allow the applications to make effective use of the driver model to improve the acceptability of the functions developed, as: Driver Drowsiness and Driver intention.

The arbitration module is defined into the IWI manager of the DESERVE abstraction. This component determines the action to be taken by the driver. The Driver Assistance Systems involve two main decision makers: the driver and the automated systems. This module considers different inputs, as follow: the Trajectory planning, Driver stage, Risk Management. The output determines who should take the control of the vehicle and the level of arbitration (or disposal) of the driver in different situations. This work uses the software tool FEMOT (Fuzzy Embedded MOTor). More detail can be found in [46].

## 6.3. Optimal Energy Consumption for Urban Electric Vehicles

**Participants:** David Gonzalez Bautista, Vicente Milanes Montero, Joshué Pérez Rastelli.

---

<sup>0</sup><http://www.yogoko.fr/>

RITS team is specially supporting two kinds of transport systems: electric mass-produced vehicles and Cybernetic Transport Systems (also electrically propelled) for urban environments. One of the key factors for getting a higher market penetration of such vehicles is their autonomy. Having this in mind, the goal of this research line is to create optimal algorithms for improving electric vehicles' battery life. It covers two specific arenas: 1) determining optimal path planning in terms of energy saving (proposed for 2015); and 2) once the route is determined, generating an adequate speed profile for covering that path. The latter objective has been investigated during 2014. Energetic model of vehicle dynamics have been developed in order to determine the lowest consumption for each of the route segments. It has permitted to develop speed references between segments combined with polynomial transition functions for the whole route to be covered. Additionally, a high-level fuzzy controller has been also designed to make the system robust to low-level failures on reference tracking. Up to 20% of battery savings have been obtained in the first tests with the proposed algorithm, showing the proper performance of the system. Additional work for adding more information from the environment as other road agents or potential unexpected diversion on the road will be also investigated during 2015 for adapting the algorithm to more realistic environments. This work has been also developed in cooperation with MSc students from Simon Bolivar University (Venezuela) and AGMUS University System (Puerto Rico, US).

#### **6.4. Perception and control strategies for autonomous docking for electric freight vehicles**

**Participants:** Joshu e P erez Rastelli, Evangeline Pollard, Vicente Milanes Montero, Fawzi Nashashibi.

The freight transportation is defined as the process of carrying goods and persons from one given point to another. Recently, urban freight transportations have been used as an alternative for the delivery problems of goods in urban environments. The present work is developed in the framework of the Furbot project (FP7), which presents a solution for future urban freight transport with new light-duty architecture with full-electrical vehicles. We focused on the onboard intelligent units, dedicated to improve the perception and control systems onboard the vehicle for the parking/docking process, considering loading and unloading phases of the freight transport procedure. Two lasers were placed on the vehicle in order to localize it with respect to the freight box. A polynomial approach is used for the trajectory planning for a smooth docking maneuver. This proposal was first tested in a 3D simulator, and then validated in a real platform. The results presented in [45] shows the good behavior of our approach, which will be implemented in the FURBOT vehicle at the end of the project.

#### **6.5. Description and technical specification of Cybernetic Transportation Systems: an urban transportation concept**

**Participants:** Joshu e P erez Rastelli, Vicente Milanes Montero, David Gonzalez Bautista, Armand Yvet, Fawzi Nashashibi.

The Cybernetic Transportation Systems (CTS) is an urban transportation concept based on two ideas: the car sharing and the automation of dedicated systems with door-to-door capabilities. In the last decade, many European projects have been developed in this context, where some of the most important are: Cybercars, Cybercars II, CyberMove, CyberC3 and CityMobil, where a first fleet of vehicles were developed by different companies and research centers around Europe, Asia and America. Considering these previous works, the FP7 project Citymobil II is in progress since 2012. Its goal is to solve some of the limitations found so far, including the definition of the legal framework for autonomous vehicles on urban environment. Much of the perception and control software has been improved in the Inria's Cybus. New guidance functionalities were developed, mainly with the introduction of stereovision-based SLAM, and Bezier curve in path planning generation. In this work, automated CTSs involved are used in the different showcases in European cities. This work presents the different improvements, adaptation and instrumentation of the vehicle used. Results show tests in our facilities at Inria-Rocquencourt (France) and the first showcase at Le on (Spain).

## 6.6. Evidential Simultaneous Localization And Mapping to describe intersection

**Participants:** Guillaume Trehard, Evangeline Pollard, Fawzi Nashashibi.

Intersections management remains a tough challenge to tackle before reaching autonomous driving in urban environment. The field of view of the vehicle is often limited by several sensors occlusions, the shapes and priority rules can significantly differ from an intersection to another and road users from pedestrian to public transports have to cross each other in sometimes complex manners.

In this context, mapping the surrounding of the vehicle and being able to estimate its position regarding a global database is crucial.

A solution of *Simultaneous Localization And Mapping* (SLAM) have then been proposed based on a 2D LIDAR sensor [49]. In the rich SLAM literature, the originality of this method lays in the use of Transferable Belief Model (TBM) framework instead of a classic probabilistic one. If this proposition was just a change of mathematical context, TBM led to an explicit management of not-known and conflict information so that its application to SLAM algorithm appeared to be really effective and robust in crowded situations. The proposed solution indeed enables to provide a map of the *static* environment crossed by the vehicle and to detect mobile obstacles in the same process and without additional tracking system.

This *Evidential SLAM* have then been tested with success over different sequences and laser set-ups extracted from the KITTI database [50].

Researches are now focused on the fusion between this SLAM solution and a Global Navigation Satellite System (GNSS) receiver to enable a map-matching on a database such as Open Street Map.

## 6.7. Laser based road obstacle classification

**Participants:** Pierre Merdrignac, Evangeline Pollard, Oyunchimeg Shagdar, Fawzi Nashashibi.

Vehicle and pedestrian collisions often result in fatality to the vulnerable road users, indicating a strong need of technologies to protect such vulnerable road users. Laser sensors have been extensively used for moving obstacles detection and tracking. Laser impacts are produced by reflection on these obstacles which suggest that more information is available for their classification. This year, we introduced the design of a new system for road obstacles classification that is divided in four parts: definition of geometric features, selection of the best features, multi-class *segment* classification based on Support Vector Machines (SVM) and *track* classification from SVM decision values integration. Our study discloses a sorted list of useful features for road obstacle recognition that were used to construct a multi-class SVM. Finally, we tested our system with 2D and 3D laser sequences and shown that it can successfully estimate the class of some road obstacles around the vehicle.

## 6.8. Deformable Parts Model based approach for on-road object detection and classification

**Participants:** Wei-Lin Ku, Evangeline Pollard, Anne Verroust-Blondet.

An important perception problem for driver assistance is the detection of the road obstacles and the recognition of their type (cars, cycles, pedestrians). This year, we tackled the on-road objects detection problem by testing and improving vision-based methods. We proposed and compared several DPM based strategies for on-road object detection and classification, laying emphasis on the problem of detecting smaller/occluded cars and pedestrians. A hybrid approach combining detection from small/large models trained with different clustering method has been introduced to boost the detection performance in both Average Precision and Maximum Recall in every difficulty level. Finally, a geometry reasoning based filtering has been employed to eliminate false alarms while preserving a great deal amount of true positives. Experimental results showed the improvement both in hybrid and geometry reasoning approaches. Most of this work has been done during the internship of Wei-Lin Ku.

## 6.9. Saturated Feedback Control for an Automated Parallel Parking Assist System

**Participants:** Mohamed Marouf, Fawzi Nashashibi, Plamen Petrov.

In 2014, RITS extended its activities in the design and development of specific automated manoeuvres. One particular interesting topic is the parallel parking problem of automatic front-wheel steering vehicles. The problem of stabilizing the vehicle at desired position and orientation is seen as an extension of the tracking problem. A saturated control has been proposed which achieves quick steering of the system near the desired position of the parking spot with desired orientation and can be successfully used in solving parking problems. In addition, in order to obtain larger area of the starting positions of the vehicle with respect to the parking spot for the first reverse maneuver of the parallel parking, an approach using saturated control with two different levels of saturation is proposed. The vehicle can be automatically parked by using one or multiple maneuvers, depending on the size of the parking spot. Simulation results were presented first in [44] to confirm the effectiveness of the proposed control schemes. New results extended to all types of parking lots shapes were recently obtained using this approach. The validation has been performed with real vehicles in the Inria test site.

## 6.10. Vehicle to pedestrian communications

**Participants:** Pierre Merdrignac, Oyunchimeg Shagdar, Evangeline Pollard, Fawzi Nashashibi.

Vehicle and pedestrian collisions often result in fatality and serious injury to the vulnerable road users. While vehicle to vehicle (V2V) communications have taken much attention in the academic and industrial sectors, very limited effort has been made for vehicle to pedestrian communications. Unlike the V2V cases, where antennas are often installed on the vehicle rooftop, pedestrian's handheld device can be carried in such a way e.g. in a bag or in a pocket, which results in poor and unpredictable communications quality. In this work, we seek to an answer to the question whether the Wi-Fi-based V2P communications meet the requirements of the pedestrian safety. This year, we studied the performances of the V2P communications especially for the receive signal strength, packet inter-arrival time, and message delivery ratio. Moreover, in order to demonstrate the feasibility of pedestrian safety supported by the V2P communications, we developed a software tool, V2ProVu, which has the functionalities of Wi-Fi based V2P communications, collision risk calculations, and hazard alarming. This work has been published in [34].

## 6.11. Multicast Communications for Cooperative Vehicular Systems

**Participants:** Ines Ben Jemaa, Oyunchimeg Shagdar, Arnaud de La Fortelle.

Vehicular communications allow emerging new multicast applications such as fleet management and point of interest (POI). Both applications require Internet-to-vehicle multicasting. 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 VANET. We propose a simplified approach that extends Mobile IP and Proxy Mobile IP. This approach aims at optimizing message exchange between vehicles and entities responsible for managing their mobility in Internet. To study the dissemination mechanisms that are suitable for fleet management applications, we propose to revisit traditional multicast routing techniques that rely on a tree structure. For this purpose, we study their application to vehicular networks. In particular, as vehicular networks are known to have changing topology, we study the application of Multicast Adhoc On Demand Vector, MAODV. We propose then Motion-MAODV [35] [16] an improved version of MAODV that aims at enhancing routes built by MAODV in vehicular networks and guarantee longer route lifetime. Finally, to enable geographic dissemination as required by POI applications, we propose the routing protocol Melody that provides a geocast dissemination in urban environments. Through simulations, Melody ensures more reliable and efficient packet delivery to a given geographic area compared to traditional geo-broadcasting schemes in highly dense scenarios.



## 6.12. Visible Light Communication for ITS applications

**Participants:** Mohammad Abu Alhoul, Oyunchimeg Shagdar, Fawzi Nashashibi.

Visible Light Communication (VLC) technology is an efficient supportive communication technology for platooning applications over short inter-vehicle distances. After implementing a complete VLC channel model, which enabling precise calculations of the optical link performance for different intervehicle distances presented in our previous work [1], this year we have studied and proposed tracking-alike method aiming at ensuring the continuity of the Line-of-Sight (LOS) and extending the Field of view (FOV) limitations. This method benefits from the exchanged information about the relative directional position of each member of the platoon, together with front and rear facing directions of each vehicle, which can be very useful data for building a reliable smooth geometrical-based compensation method. The simulation results showed that trajectory influences on the optical incidence and irradiance angles can be compensated efficiently and without deploying any tracking method.

## 6.13. Study on the IEEE 802.11p Channel Congestion Problem

**Participant:** Oyunchimeg Shagdar.

The IEEE 802.11p is a standardized WiFi technology dedicated to V2X communications for especially road safety and efficiency applications. It is expected that vehicles periodically broadcast messages to announce their existences using the IEEE 802.11p frequency channel. However, because the IEEE 802.11p has a limited wireless bandwidth, in dense traffic conditions the V2V communications performances are poor, failing to satisfy the application requirements. In RITS, we study the issue and develop congestion control algorithms. In 2014, we studied the reactive distributed congestion control algorithm, proposed by the European Telecommunications Standards Institute (ETSI), and showed that the algorithm creates unstable resource utilization, which can cause the reactive Distributed Congestion Control (DCC) to perform worse than non-DCC systems. We proposed an asynchronous algorithm, where DCC control is made in such a way that channel resource is used in an asynchronous manner by the different stations. Our results show that the asynchronous DCC approach outperforms both the non-DCC and reactive DCC mechanisms. The work has been reported at a ETSI meeting in December 18, 2014 [54].

## 6.14. Study on V2V Communications and Emergent Behavior of Heuristically-Driven Intelligent Vehicles

**Participant:** Oyunchimeg Shagdar.

The automated cooperative driving applications require efforts on multiple research domains including robotics, artificial intelligence, and communications to build a safe and intelligent collective driving behavior. While some studies show the potentials of the V2V communications for safer and smoother automated driving, it is still not clear if the standardized technologies can meet the strict requirements of the automated driving applications. More importantly, if the decisions for individual vehicles' control are based on the V2V communications, the communications performance must largely affect the "quality" of the collective behavior. Motivated by this, we study the inter-dependencies between communications and collective automated driving behavior. In our study [43] we combine different V2V communication modes with different dynamic path-finding heuristics, over a population of several hundreds of intelligent vehicles, to observe convergence towards stable traffic. The various traffic stability levels are compared in order to exhibit most efficient combinations of communication modes and path-finding heuristics.

## 6.15. Distributed Agreement and String Control in Intelligent Vehicular Networks (IVNs)

**Participant:** Gérard Le Lann.

IVNs are composed of automated (autonomous and communicating) vehicles, ranging from pre-planned platoons to ad hoc vehicular networks (VANETs). Agreement problems in the presence of concurrency and failures are not well investigated yet in IVNs. We have examined a specific class of such problems, those arising in string formations. Regarding string membership (vehicles leaving or joining a string), with few exceptions, safety issues have been addressed so far assuming that (1) no more than 1 insertion operation would be performed at any given time or, (2) every vehicle decides unilaterally, i.e. undertakes a maneuver after having activated some signal, leaving to surrounding vehicles the responsibility of inferring intended maneuvers. Assumption (1) is not realistic. There are numerous risk-prone scenarios where a posteriori reactive approaches (assumption (2)) may fail. Therefore the need for investigating proactive approaches, where vehicles (1) are made aware of intended impending maneuvers, (2) agree on which maneuvers can be safely undertaken, prior to performing physical maneuvers. It follows that a solution to numerous string control problems consists of a pair  $(A, \Phi)$ , where  $A$  stands for a distributed agreement algorithm which achieves global coordination in the presence of failures and concurrency, and  $\Phi$  stands for control laws drawn from control theory and robotics. Algorithm  $A$  is run prior to local activations of  $\Phi$ .

Our work is based on the cohort construct, which serves to formalize the concept of strings. Velocity Agreement is the generic problem selected. At any time, some number of string/cohort members may propose each a new velocity value. In fine, every vehicle computes a unique new velocity  $V$ . Proposed values are propagated via neighbor-to-neighbor (N2N) radio communications. We have devised a solution called the VAgree algorithm. In the presence of up to  $f$  failures (on-board systems, N2N message losses), the following properties shall hold:

- Validity: Decision value  $V = \Psi(\text{proposed values})$ .
- Agreement: No two members decide differently.
- Time-Bounded Termination: VAgree terminates at most  $\theta$  time units.
- Synchronicity: Times at which  $V$  is posted to on-board systems are comprised within a small time interval  $\epsilon$ . Distance traveled during  $\epsilon$  by the member earliest to post  $V$  until the latest member does so is an order of magnitude smaller than vehicle sizes.

The VAgree algorithm is presented in a paper which is under submission.

## 6.16. Standardization and automated vehicles

**Participant:** Michel Parent.

Michel Parent has been active over the last 4 years in this group to introduce automated vehicles in the scope of service robots and he contributed actively in the activities of several working groups (WGs). In WG7, he participated in the writing of the document ISO13482 on the safety of service robots. This document has reached the Final Draft for an International Standard level (FDIS) and is now published in English and French. It has already been used by companies to certify some robotics products, including Robosoft in France for automated vehicles. In WG8, Michel Parent participated in the elaboration of standard procedures for the testing of service robots and in particular for automated vehicles. The document CD18646 « Robots and robotic devices — Performance criteria and related test methods for service robots — Part 1: Locomotion for wheeled robots » is in progress.

## 6.17. Legislation and certification of fully automated road vehicles for urban public transport

**Participant:** Michel Parent.

An important research area of automated road vehicles and one of the focuses of the CityMobil2 Project is to look at the legislation and certification of fully automated road vehicles for urban public transport (the cybercars). This part of the research was done essentially by Michel Parent in 2014 and gave birth to several CityMobil2 deliverables.

One of the tasks was to identify the current legislation in France and the organizations involved in the changes for this legislation. Several meetings were therefore organized at the French level with key persons from the Ministry of Transport, the Ministry of Interior (responsible for the road legislation) and their services (in particular the SRMTG in charge of certifying the guided transport systems). These meetings were essential in obtaining the authorization to operate the cybercars for the demonstration in La Rochelle. At the European level, a meeting was organized in May 2014 with representatives of 12 of the European countries (mostly those involved directly with CityMobil2 or with automated vehicles R&D).

Another task was to propose a certification methodology for automated road transport systems. For this task, a careful analysis of the test site in La Rochelle was conducted and led to a number of use cases. Key elements were defined to perform the risk analysis. Many hazards were identified but the most important ones are the behavior of pedestrian and cyclists. For the analysis, two key variables were defined: the minimum mobile object detection distance (MMODD) and the maximum mobile object speed (MMOS). For each use case, a combination of these 2 variable lead to a maximum vehicle speed in order to reach an acceptable risk evaluated as a combination of severity and probability.

In order to verify the proper behavior of the vehicle itself (lane tracking, obstacle avoidance, comfort,...), a number of standard tests have also been defined and are now proposed at the International level (ISO standards).

## 6.18. Belief propagation inference for traffic prediction

**Participants:** Cyril Furtlehner, Jean-Marc Lasgouttes.

This work [60] 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 under review for a Journal and has been largely revised in the process.

## 6.19. Sparse covariance inverse estimate for Gaussian Markov Random Field

**Participants:** Cyril Furtlehner, Jean-Marc Lasgouttes.

We investigate the problem of Gaussian Markov random field selection under a non-analytic constraint: the estimated models must be compatible with a fast inference algorithm, namely the Gaussian belief propagation algorithm. To address this question, we introduce the  $\star$ -IPS framework, based on iterative proportional scaling, which incrementally selects candidate links in a greedy manner. Besides its intrinsic sparsity-inducing ability, this algorithm is flexible enough to incorporate various spectral constraints, like e.g. walk summability, and topological constraints, like short loops avoidance. Experimental tests on various datasets, including traffic data from San Francisco Bay Area, indicate that this approach can deliver, with reasonable computational cost, a broad range of efficient inference models, which are not accessible through penalization with traditional sparsity-inducing norms.

This work has been presented at ECML/PKDD 2014 [40]. The code for  $\star$ -IPS has been made available at <https://who.rocq.inria.fr/Jean-Marc.Lasgouttes/star-ips/>.

## 6.20. Herding behavior in a social game

**Participants:** Guy Fayolle, Jean-Marc Lasgouttes.

The system *Ma Micro Planète* belongs to the so-called *Massively Multi-Player online Role Playing game* (MMORPG), its main goal being to incite users to have a sustainable mobility. Two objectives have been pursued.

- Construct an experimental platform to collect data in order to prompt actors of the mobility to share information (open data system).
- See how various mechanisms of a game having an additive effect could modify the transportation requests.

At the heart of the game are community-driven *points of interest* (POIs), or *sites*, which have a score that depends on the players activity. The aim of this work is to understand the dynamics of the underlying stochastic process. We analyze in detail its stationary regime in the thermodynamic limit, when the number of players tends to infinity. In particular, for some classes of input sequences and selection policies, we provide necessary and sufficient conditions for the existence of a complete meanfield-like measure, showing off an interesting *condensation* phenomenon.

The work has been published this year in *Queueing Systems* [20].

## 6.21. Properties of random walks in orthants

**Participant:** Guy Fayolle.

We pursued works initiated these last years in several directions.

### 6.21.1. *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-Petersbourg, 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 [55].

### 6.21.2. *About a possible analytic approach for walks with arbitrary big jumps in $\mathbb{Z}_+^2$*

The article [21], achieved in collaboration with K. Raschel (CNRS and University F. Rabelais, Tours) considers random walks with arbitrary big jumps. For that class of models, we announce a possible extension of the analytic approach proposed in [3], initially valid for walks with small steps in the quarter plane. New technical challenges arise, most of them being tackled in the framework of generalized boundary value problems on compact Riemann surfaces.

### 6.21.3. *Correction of papers*

Guy Fayolle found important errors in several articles dealing with models involving random walks in  $\mathbb{Z}_+^2$ . This is the object of the letter to the editors [19]. The concerned authors have provided new correct versions of their studies.

### 6.21.4. *Communication networks with harvesting energy supply*

In collaboration with S. Foss (Heriot-Watt University, Edinburgh), we started to analyze stability and performance of a number of models of parallel queues with multiple access and individual energy supplies. Energy limitation in general decreases the stability region, but also may increase it for specific parameter regions. The most difficult and intriguing cases arise when the input rates of requests and of energy items are close. Preliminary models of physical interest involve random walks in  $\mathbb{Z}_+^4$ .

## 6.22. 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.

The aim of the global optimization algorithm is to pursue the minimization of the mean squared travel time encountered by each user. Our setting can be seen as a resource allocation problem, known as the “Transportation Problem” in Operational Research literature. It is solvable using several algorithms, among which the simplex algorithm. Unfortunately, these algorithms are not usable here for two reasons:

- The allocation of slots to the users is done online, 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.

We therefore present here a simplified approach, which should be fast enough to scale for large systems. The principle of the algorithm 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). Since the implications can be intricate, we only consider a first order effect.

Our work in the Mobility 2.0 project has been briefly described in [\[37\]](#).

## AYIN Team

# 5. New Results

## 5.1. Highlights of the Year

- Yuliya Tarabalka was nominated CR1 since 1 January 2015.
- Josiane Zerubia was elected for a duration of 6 years at the board of directors of the French Society of Photogrammetry and Remote Sensing (SFPT, <http://www.sfpt.fr/>).
- Josiane Zerubia was invited by Technion to give a plenary talk at SIMA'14 in Ein Gedi, Israel organized for the 60th birthday of Prof. Alfred Bruckstein in May, <http://www.cs.technion.ac.il/SIMA14/>.

## 5.2. Markov Random Fields

### 5.2.1. Fusion of multitemporal and multiresolution remote sensing data and application to natural disasters

**Participants:** Ihsen Hedhli, Josiane Zerubia [contact].

*This work was carried out in collaboration with Prof. Gabriele Moser and Prof. Sebastiano Serpico from DITEN departement, University of Genoa, Italy.*

Multitemporal data, Multiresolution data, Supervised classification, Hierarchical Markov random fields.

The capabilities to monitor the Earth surface, and especially urban and built-up areas, from environmental disasters such as floods or earthquakes, and to assess the ground impact and damage of such events play primary roles from multiple social, economic, and human viewpoints. In this framework, accurate and time-efficient classification methods are especially important tools to support rapid and reliable assessment of the ground changes and damages induced by a disaster, in particular when an extensive area has been affected. Given the huge amount and variety of data available currently from last-generation very-high resolution (VHR) satellite missions, (such as Pléiades, COSMO-SkyMed, or WorldView-2), the main difficulty is to develop a classifier that can take benefit of multiband, multiresolution, multirate, and possibly multisensor input imagery. In such a context, Markov random field (MRF) models are widely used to solve classification problems as they permit one to integrate contextual information into the classification scheme. Due to their non-causal nature, these models generally lead to iterative inference algorithms that are computationally demanding (e.g., optimization via simulated annealing), thereby justifying the choice of a hierarchical structure, with good methodological and application-oriented properties such as: (i) the causality in scale, under Markovian assumption, which allows the use of a non-iterative algorithm with acceptable computational time and (ii) the possibility to incorporate images acquired at multiple resolutions in the hierarchy for multiresolution and multisensor fusion purposes [10]. In the proposed method, multirate and multiresolution fusion is based on explicit statistical modeling through hierarchical Markov random field modeling. The model allows both input data collected at multiple resolutions and additional multiscale features derived through wavelets to be fused. The proposed approach consists of a supervised Bayesian classifier that combines: (i) a joint class-conditional statistical model for pixelwise information and (ii) a hierarchical MRF for spatio-temporal and multiresolution contextual information. Step (i) deals, first, with the modeling of the marginal statistics of the spectral channels acquired at each resolution and conditioned to each class. Step (ii) consists in the integration of this statistical modeling in a hierarchical Markov random field for each date. An especially novel element of the proposed approach is the use of multiple quad-trees in cascade (see Figure 1), each associated with each new available image at different dates, with the aim to characterize the temporal correlations associated with distinct images in the input time series and to support the joint analysis of multitemporal, multiresolution, and possibly multisensor imagery. The transition probabilities between scales and between different dates determine the hierarchical MRF since they formalize the causality of the statistical interactions involved [11].



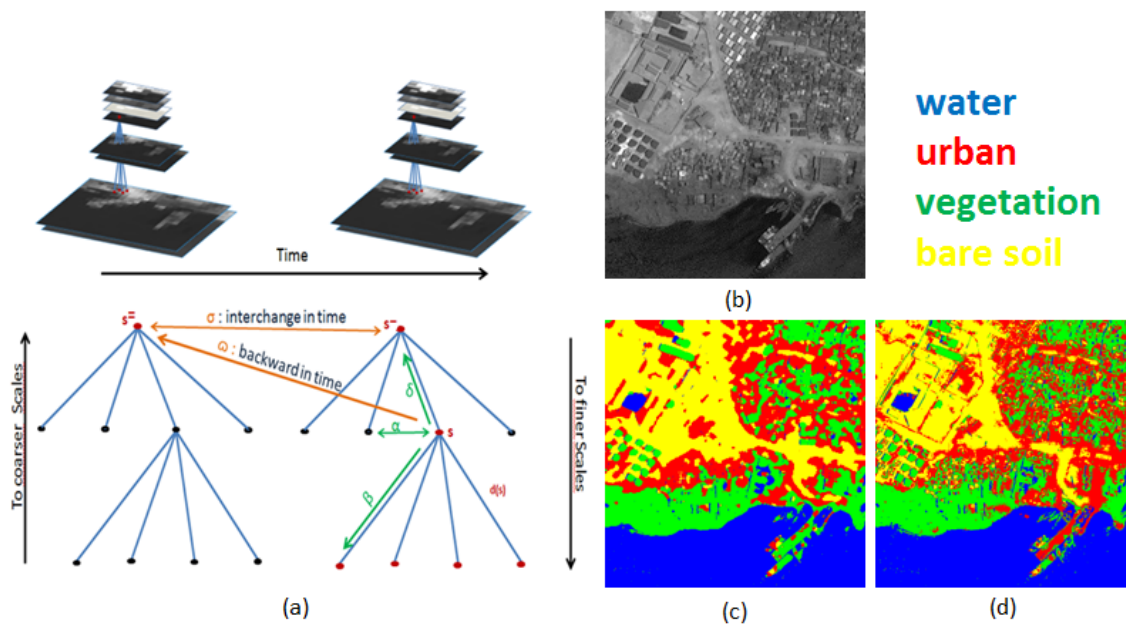


Figure 1. a) Multitemporal hierarchical structure; b) Panchromatic image of Port au Prince (Pléiades, ©CNES distribution Airbus DS, 2013); c) Classification map using single date hierarchical structure; d) Classification map obtained through the proposed multitemporal method.

### 5.2.2. A multi-layer Markov model for change detection in temporally separated aerial image pairs

**Participant:** Josiane Zerubia [contact].

This work was carried out in collaboration with Prof. Zoltan Kato from Institute of Informatics, University of Szeged, Hungary [<http://www.inf.u-szeged.hu/~kato/>], and Praveer Singh from Institut Mines-Telecom.

Multilayer Markov Random Fields (MRF), Histogram of Gradients (HOG), change detection, graph-cut optimization, aerial/satellite images.

In the proposed approach developed last year, we have tried to include both texture as well as pixel level information to build a three layer Markov model using the *Histogram of Oriented Gradients (HOG)* and the *Gray Level Difference* features on the topmost and bottommost layer respectively. Using a ground truth (GT) mask defined manually by an expert for each of the image pairs in the data set (obtained from the Hungarian Institute of Geodesy, Cartography and Remote Sensing), we employ a supervised technique to mark the initial set of pixels / sites as foreground or background. On the basis of the *HOG difference* and the *Gray level difference* feature vector corresponding to all the pixels in the image pair, a probability density function is fitted individually for the binary label set comprising of foreground and background labels using the GT. The probabilistic estimate is calculated using one training image pair for each data set. Using this probabilistic measure, a negative log likelihood is computed for each pixel (for both the features as well as the binary label set) which is then passed to the energy function of the proposed 3-layer MRF model. The final segmentation is obtained by minimizing the energy using a graph-cut algorithm, and subsequently a final foreground and background labelling is obtained over the combined layer. Figure 2, shows aerial image pairs, one of them captured in 1984 by FOMI, Hungary (a) and the other one by GoogleEarth in 2007 (c). (b) is the ground truth and (d) is a combination of the hierarchical MRF based change detection (in red), ground truth (in green) and changes detected correctly (in yellow). This year, we have made a comparison of this method with two other multilayer MRFs for change detection developed at MTA-SZTAKI in Budapest, Hungary.

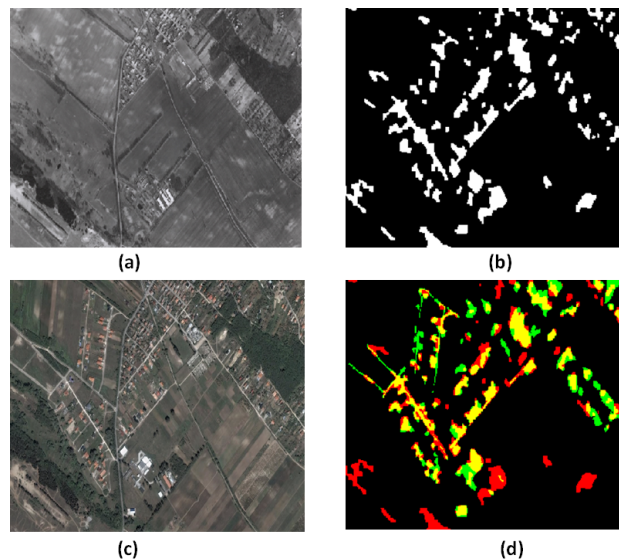


Figure 2. Change detection in an aerial image pair using a hierarchical MRF. a) Aerial image captured in 1984 by ©FOMI; b) Ground truth; c) Aerial image captured by ©GoogleEarth in 2007; d) Combination of the hierarchical MRF based change detection (in red), ground truth (in green), and changes detected correctly (in yellow).

### 5.2.3. Graph-cut-based model for spectral-spatial classification of hyperspectral images

**Participant:** Yuliya Tarabalka [contact].

*This work has been done in collaboration with Aakanksha Rana (Institut Mines-Telecom/EURECOM).*

Hyperspectral images, graph cut, multi-label alpha expansion, contextual information, energy minimization

The very high spatial and spectral resolution of the last generation of remote sensors provides rich information about every pixel in an image scene, hence opening new perspectives in classification, but also presenting the challenge of analysing high data volumes. While pixel-wise classification methods analyze each pixel independently, classification results can be significantly improved by including spatial information in a classifier.

In this work, we proposed a spectral-spatial method for hyperspectral image classification based on a graph cut [15]. The classification task is expressed as an energy minimization problem on the spatio-temporal graph of image pixels, and is solved by using the graph-cut  $\alpha$ -expansion approach. The energy to optimize is computed as a sum of data and interaction energy terms, respectively. The data energy term is computed using the outputs of the probabilistic support vector machines classification. The second energy term, which expresses the interaction between spatially adjacent pixels in the eight-neighborhood, is computed by using dissimilarity measures between spectral vectors, such as vector norms, spectral angle map, or spectral information divergence. The performance of the proposed method was validated on hyperspectral images captured by the ROSIS and the AVIRIS sensors. Figure 3 compares classification results obtained by applying support vector machines and the proposed approach for the ROSIS hyperspectral image acquired over the University of Pavia. The new method yields higher classification accuracies when compared to the recent state-of-the-art approaches.

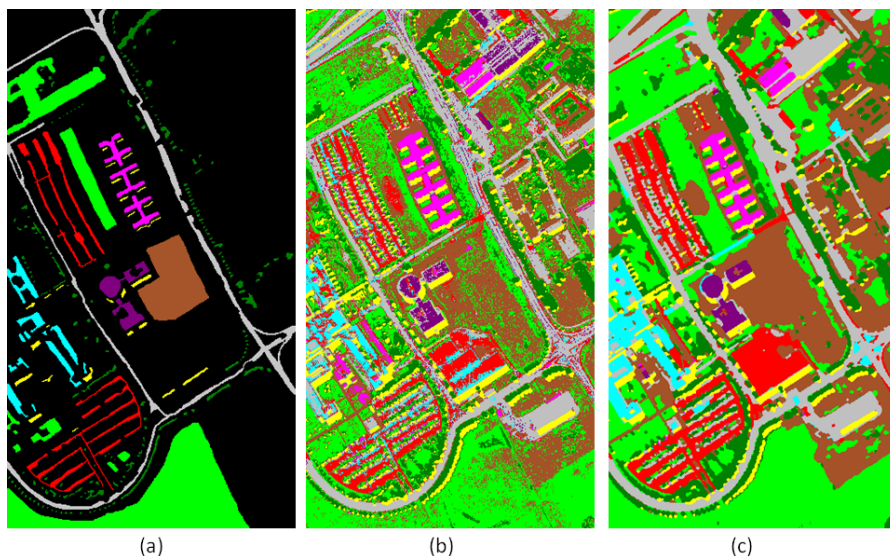


Figure 3. Hyperspectral image of the University of Pavia. (a) Ground-truth (b) Support vector machines classification map. (c) Graph-cut classification map.

## 5.3. Marked point processes

### 5.3.1. Multiple target tracking using spatio-temporal marked point processes

**Participants:** Paula Craciun, Josiane Zerubia [contact].

*This work has been done in collaboration with Mathias Ortner from Airbus D&S (<http://www.space-airbusds.com/fr/>)*

Multiple target tracking, stochastic geometry, point processes, remote sensing

Tracking can be defined as the problem of estimating the trajectories of objects in the image plane, as they move around the scene. Hence, a tracker assigns consistent labels to the objects in different frames of a sequence of images and can additionally provide information about the orientation, shape or size of the objects. Multi-target tracking has been historically achieved using sequential techniques, the major drawback of such methods residing in the impossibility to modify past results in the light of new data. However, applications such as offline video processing or information retrieval are not sequential in nature. Batch processing methods are preferred in this case since they do not suffer from the limitations of sequential methods. Nevertheless, these techniques remain poorly explored and highly underused.

We propose a novel approach based on spatio-temporal marked point processes to detect and track moving objects in a batch of high resolution images [17]. We develop a new, intuitive energy based model consisting of several terms that take into account both the image evidence and physical constraints such as target dynamics, track persistence and mutual exclusion. We construct a suitable optimization scheme that allows us to find strong local minima of the proposed highly non-convex energy [9]. The model has been validated on two types of data: remotely sensed satellite image sequences, characterized by high resolution, high signal to noise ratio and low temporal frequency; and biological image sequences, characterized by high resolution, low signal to noise ratio and high temporal frequency.

Tracking results are shown in Figure 4, which shows the detection (dots) and tracking (lines) results of boats in a sequence of 14 high resolution remotely sensed images. The images are captured with a low temporal frequency at different acquisition angles.

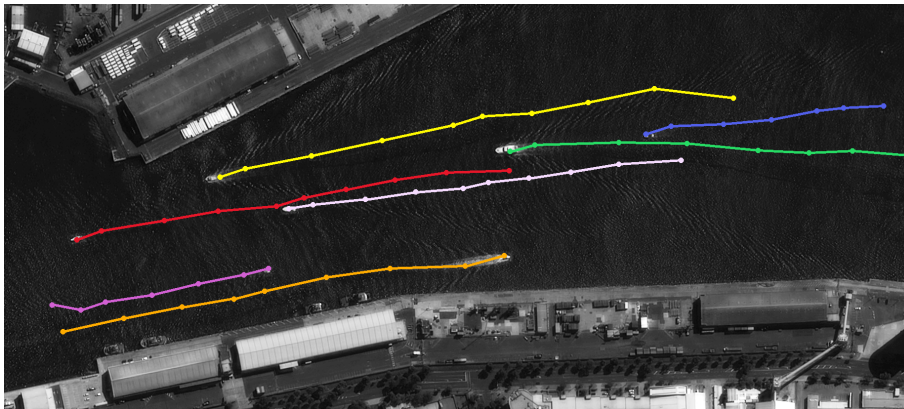


Figure 4. Detection and tracking results on a sequence of satellite images taken at different angles ©Inria/AYIN.

### 5.3.2. 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 Rémi Flamary (Université de Nice Sophia Antipolis), 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 [7], while the second one estimates the parameters by the Stochastic Approximation Expectation-Maximization (SAEM) algorithm [8]. 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. Such a representation is displayed on Figure 5. 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.

Up to recently, the parameters of MPPs were estimated by the Stochastic Expectation-Maximization (SEM) algorithm developed by Celeux & Diebolt (1985). This algorithm consists in alternatively estimating the expected pseudo-likelihood based on a random configuration and updating the parameter value by maximum of pseudo-likelihood. However, since it does not have a pointwise convergence, Ben Hadj et al. (2010) considered running a simulated annealing scheme after few iterations of SEM in order to reach convergence, at the cost of a higher computational time. Instead, we proposed to adapt the Stochastic Approximation Expectation-Maximization (SAEM) algorithm, developed by Delyon et al. (1999), to MPPs. Indeed, it both offers pointwise convergence and a similar computational time as SEM by efficiently taking into account past configurations in the update of the expected pseudo-likelihood.

Using both approaches resulted in the division of the computational time of the estimation of MPPs parameters by 2 and in an increase in performance of detection.

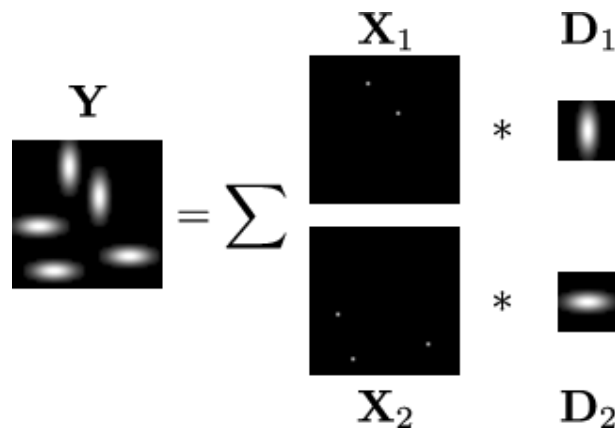


Figure 5. Representation of an image as a sum of convolutions between atoms of a dictionary and matrices of positions.

### 5.3.3. Generic curvilinear structure modeling via marked point process theory

**Participants:** Seong-Gyun Jeong, Yuliya Tarabalka, Josiane Zerubia [contact].

Curvilinear structure extraction, object detection, marked point process, stochastic inference



We proposed a marked point process model to analyze underlying curvilinear structure for wide ranges of input data, for instance, wrinkles, DNA filaments, road cracks, and blood vessels [12], [13]. It is based on sampling technique so that the model represents an arbitrary shape of the line network with a set of small line segments. The line segments should be fit into the given image data, and be harmonic with those of neighborhoods. To take these issues into consideration, we formulate a maximum a posteriori (MAP) estimation as an energy minimization problem. The energy function for given line configuration  $\mathbf{s}$  can be decomposed into data likelihood term  $E_{\text{data}}$  and prior term  $E_{\text{prior}}$ :

$$E(\mathbf{s}) = \sum_i^{\#(\mathbf{s})} E_{\text{data}}(s_i) + \lambda \sum_{i \sim j} E_{\text{prior}}(s_i, s_j),$$

where  $\#(\mathbf{s})$  denotes the total number of line segments in the current configuration,  $i \sim j$  represents symmetric neighborhood system, and  $\lambda$  controls the relative importance of two terms. For the data term, we exploit oriented gradient information and homogeneity of the pixel intensities corresponding to line segment on the image site. The prior energy defines topology of the line configuration in that penalizes overlapping and attracts smooth connections. Another contribution of the work is to reduce parameter dependencies of the marked point process model using aggregation approach. We repeated to perform Markov chain Monte Carlo (MCMC) sampling with different parameter vectors to obtain multiple line hypotheses. Then, we combine line hypotheses to maximize the consensus among detection results.

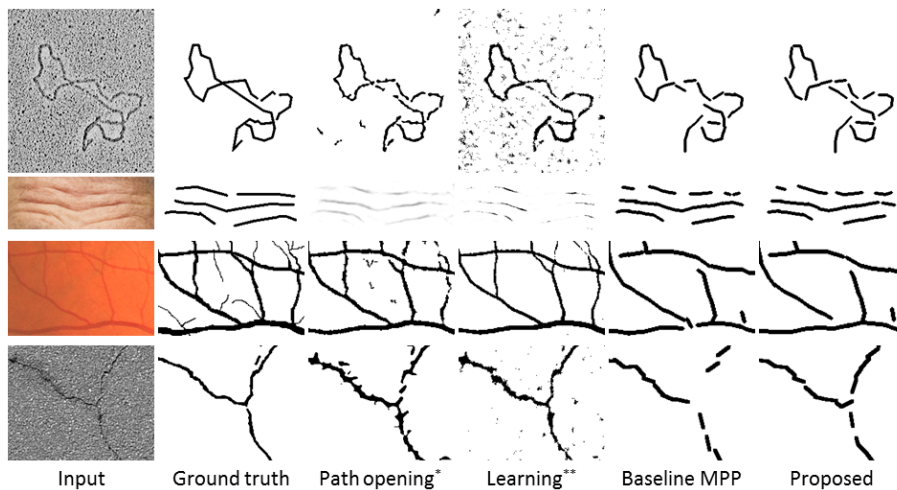


Figure 6. Comparison of the line detection results on DNA filaments, wrinkles, retina, and road cracks (top to bottom).

In figure 6, we have compared line detection results of manually labeled image, morphological filtering (path opening), supervised feature learning, an MPP model using single parameter vector, and the proposed algorithm. The proposed algorithm extracts the most salient line structures for all datasets without any parameter estimation procedure.

## 5.4. Shapes and contours

### 5.4.1. Riemannian metrics on spaces of curves and surfaces

**Participant:** Ian Jermyn [contact].



This work is being done in collaboration with Anuj Srivastava of Florida State University [<https://www.fsu.edu/>].

Shape, Riemannian, metric, elastic, curve, surface, functional data, alignment

Statistical shape modelling has many applications in image processing and beyond. One of the key problems in this area is to develop and understand measures of shape similarity. One approach uses Riemannian metrics induced on ‘shape space’ by Riemannian metrics on the space of embeddings. Current work is focused on generalizing to surfaces the elastic metric used for curve embeddings, and in finding surface representations that simplify computations in the same way that the square root velocity representation simplifies computations in the case of curves. The notion of a ‘square-root normal field’ (SRNF), which leads to a reduced version of the full elastic metric, is a promising possibility in this direction.

The most recent work [16] has focused on estimating the inverse of the SRNF map. If this can be done even approximately, a very efficient framework results: the surfaces, represented by their SRNFs, can be efficiently analyzed using standard Euclidean tools, and only the final results need to be mapped back to the surface space. In this work, we developed a procedure for inverting SRNF maps of star-shaped surfaces, a special case for which analytic results can be obtained. We tested our method via the classification of 34 cases of ADHD (Attention Deficit Hyperactivity Disorder), plus controls, in the Detroit Fetal Alcohol and Drug Exposure Cohort study. We obtained state-of-the-art results.

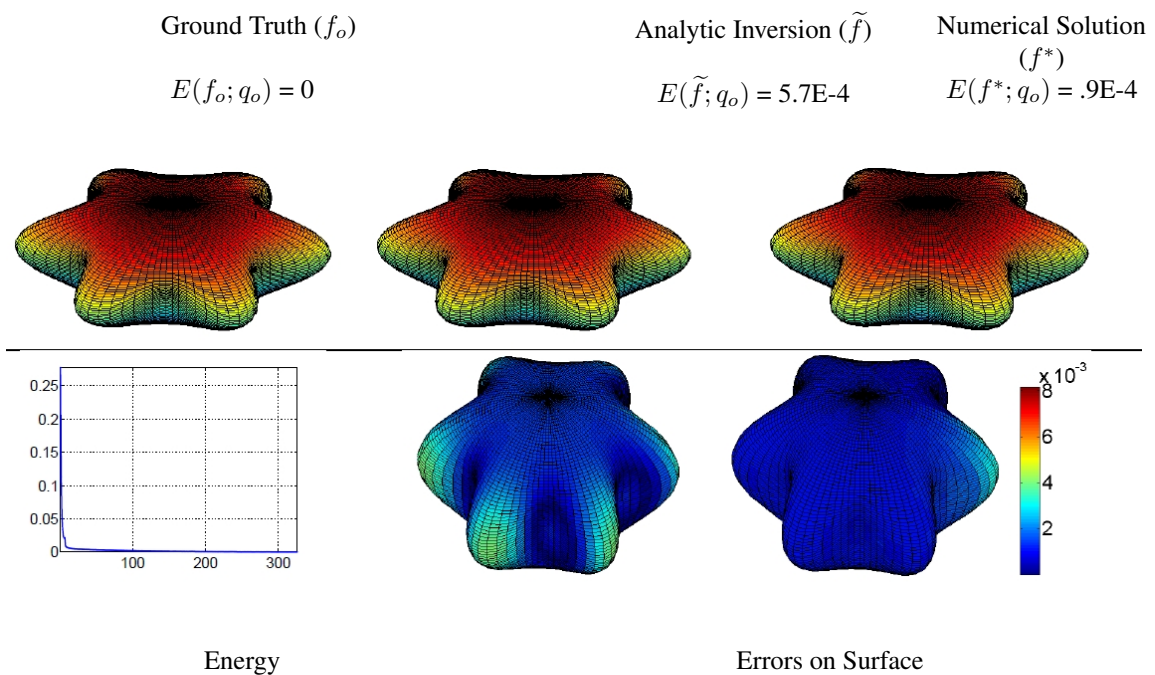


Figure 7. Reconstructing a surface from its SRNF. A target surface ( $f_o$ ) is numerically reconstructed as  $f^*$  with initialization as the unit sphere. The energy plot shows the evolution of energy against iterations with initialization as a unit sphere. The analytically inverted surface  $\tilde{f}$  is shown for comparison. The corresponding energies  $E(\tilde{f}; q_o)$  and  $E(f^*; q_o)$  are also shown. The errors between the reconstructed surfaces and the ground truth are shown on the ground truth surface with colours representing the magnitudes, i.e.  $|f^*(s) - f_o(s)|$  for all  $s \in S^2$ .

#### 5.4.2. Enforcing monotonous shape growth or shrinkage in video segmentation

**Participant:** Yuliya Tarabalka [contact].

This work has been done in collaboration with Dr. Guillaume Charpiat (STARS team, Inria-SAM), Dr. Bjoern Menze (Technische Universität München, Germany and Asclepios team, Inria-SAM), and Dr. Ludovic Brucker (NASA GSFC, USA) [<http://www.nasa.gov/>].

Video segmentation, graph cut, shape analysis, shape growth

Automatic segmentation of objects from video data is a difficult task, especially when image sequences are subject to low signal-to-noise ratio or low contrast between the intensities of neighboring structures. Such challenging data are acquired routinely, for example, in medical imaging or satellite remote sensing. While individual frames can be analyzed independently, temporal coherence in image sequences provides a lot of information not available for a single image. In this work, we focused on segmenting shapes that grow or shrink monotonically in time, from sequences of extremely noisy images.

We proposed a new method for the joint segmentation of monotonically growing or shrinking shapes in a time sequence of images with low signal-to-noise ratio [3]. The task of segmenting the image time series is expressed as an optimization problem using the spatio-temporal graph of pixels, in which we are able to impose the constraint of shape growth or shrinkage by introducing unidirectional infinite-weight links connecting pixels at the same spatial locations in successive image frames. The globally-optimal solution is computed with a graph-cut algorithm. The performance of the proposed method was validated on three applications: segmentation of melting sea ice floes; of growing burned areas from time series of 2D satellite images; and of a growing brain tumor from sequences of 3D medical scans. In the latter application, we imposed an additional inter-sequences inclusion constraint by adding directed infinite-weight links between pixels of dependent image structures. Figure 8 shows a multi-year sea ice floe segmentation result. The proposed method proved to be robust to high noise and low contrast, and to cope well with missing data. Moreover, it showed linear complexity in practice.

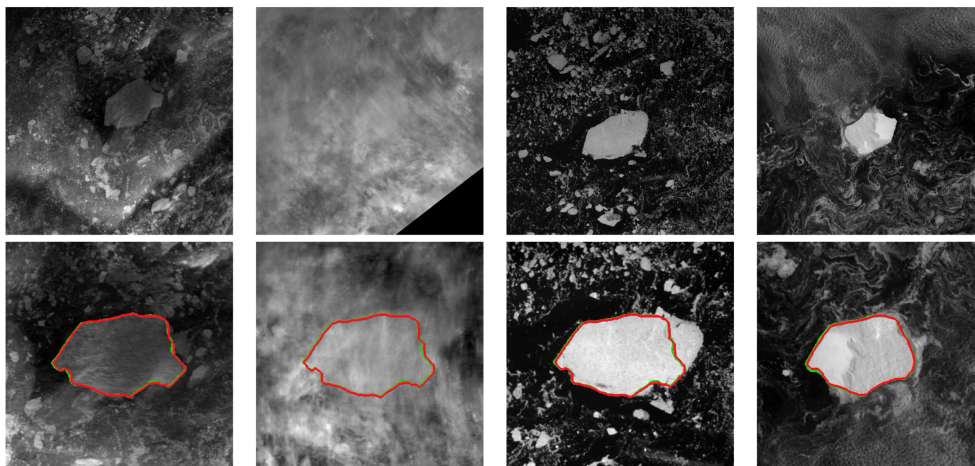


Figure 8. Top: MODIS images for four time moments (days 230, 233, 235 and 267 of 2008, respectively). Bottom: corresponding aligned images with segmentation contours (in red). Manual segmentation is shown in green.

#### 5.4.3. Multi-label image segmentation with partition trees and shape prior

**Participants:** Emmanuel Maggiori, Yuliya Tarabalka [contact].

*This work has been done in collaboration with Dr. Guillaume Charpiat (STARS team, Inria-SAM).*

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

The multi-label segmentation of images is one of the great challenges in computer vision. It consists in the simultaneous partitioning of an image into regions and the assignment of labels to each of the segments. The problem can be posed as the minimization of an energy with respect to a set of variables which can take one of

multiple labels. Throughout the years, several efforts have been done in the design of algorithms that minimize such energies.

We proposed a new framework for multi-label image segmentation with shape priors using a binary partition tree [19]. In the literature, such trees are used to represent hierarchical partitions of images, and are usually computed in a bottom-up manner based on color similarities, then processed to detect objects with a known shape prior. However, not considering shape priors during the construction phase induces mistakes in the later segmentation. This paper proposes a method which uses both color distribution and shape priors to optimize the trees for image segmentation. The method 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 reducing the search space and make the optimization efficient (see Figure 9 (i)) and [19]. Our experiments (see Figure 9) show that the optimization approach succeeds in incorporating shape information into multi-label segmentation, outperforming the state-of-the-art.

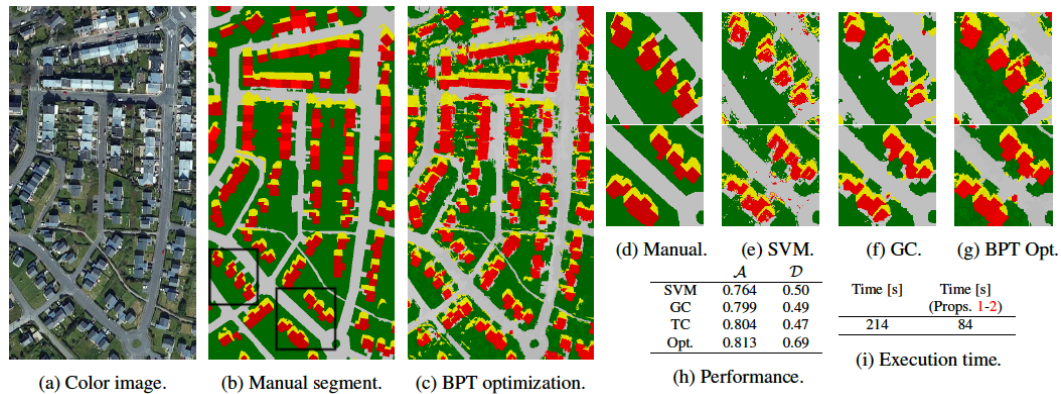


Figure 9. Classification results for the satellite image over Brest.  $\mathcal{A}$  denotes overall classification accuracy, and  $\mathcal{D}$  denotes average building's overlap. The performance of the proposed binary partition tree (BPT) optimization method is compared with the following methods: 1) support vector machines (SVM) classification; 2) graph cut (GC) with  $\alpha$ -expansion; 3) cut on the BPT, regularized by the number of regions without using shape priors (TC).

## 5.5. Other detection approaches

### 5.5.1. Image-based evaluation of treatment responses of facial wrinkles using LDDMM registration and Gabor features

**Participants:** Nazre Batool, Josiane Zerubia [contact].

Face, skin texture, detection of wrinkles, LDDMM registration, response to treatment, Gabor filters, morphological processing

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. The novelty of this application is that a series of images of the same subject over a shorter time period of weeks are analyzed instead of more prevalent inter-person analysis of facial images. To overcome the challenges of detecting and evaluating such subtle changes, we propose a framework to compare image features in key wrinkle sites only while excluding the noise introduced by changes in surrounding skin texture. After initial registration using facial landmarks such as corners of eyes, nose, mouth, we propose a method based on Large Deformation Diffeomorphic Metric Mapping (LDDMM) to achieve finer registration. Fig. 10 (1a-1e) shows an example of registration using LDDMM for a pair of images.

Then we use N. Batool's previously proposed algorithm (Nazre & Chellappa (2015)) to detect key wrinkle sites. The algorithm is based on 'scaled' maximum Gabor filter responses and the incorporation of geometric constraints via morphological image processing. The binary output from the algorithm is used to create a unique wrinkle template for each subject. Fig. 10 (2a-2d) an example of obtaining a unique wrinkle template from an image using Gabor responses and wrinkle detection algorithm in (Nazre & Chellappa (2015)). Gabor responses in this template, in time series images are compared to detect subtle changes for a subject. We do not adopt the direct approach of comparing filter responses in the whole image instead of those in wrinkle template only because such an approach causes intermingling of skin texture variations in non-wrinkle sites with changes in wrinkle sites degrading the overall accuracy.

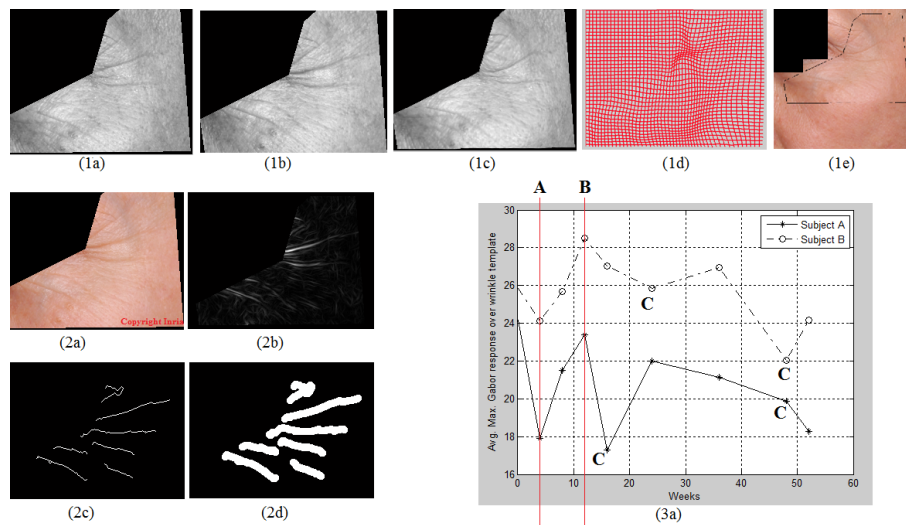


Figure 10. Overview of the evaluation framework. (1a) Week 4 image. (1b) Baseline image. (1c) Week 4 image registered using LDDMM to baseline image. (1d) Deformation of the underlying 2D space. (1e) Deformed week 4 images aligned in the original face image. (2a) Baseline image. (2b) Gabor maximum amplitude response. (2c) Detected wrinkles. (2d) The template for key wrinkle sites. (3) Plot of results for two subjects.

Fig. 10 (3a) shows a plot of results for two subjects where y-axis shows average maximum Gabor amplitude response in key wrinkle sites and x-axis corresponds to the number of weeks after the treatment. For both subjects a significant drop in the average response can be seen 4 weeks after the treatment (event 'A'). An increase in the Gabor response happened at week 12 (event 'B') which coincided with slight darkening/reddening of skin for both subjects. On the other hand, event 'C' represents co-occurrence of skin lightening with a decrease in Gabor response. These preliminary results indicate trends in wrinkle responses to treatment, skin darkening and lightening. In future, these trends will be validated by more rigorous experiments.

### 5.5.2. SAR data classification using generalized Gamma mixture model

**Participant:** Josiane Zerubia [contact].

This work has been performed in collaboration with Dr. Vladimir Krylov (University of Genoa, Italy), Prof. Heng-Chao Li, Prof. Ping-Zhi Fan (Southwest Jiaotong University, Chengdu, China) and Prof. William Emery (University of Colorado, Boulder, USA).

SAR images, statistical modeling, generalized Gamma mixture model



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 (G $\Gamma$ D) in view of its flexibility and compact analytical form. Specifically, we have developed a generalized Gamma mixture model (G $\Gamma$ MM) 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.

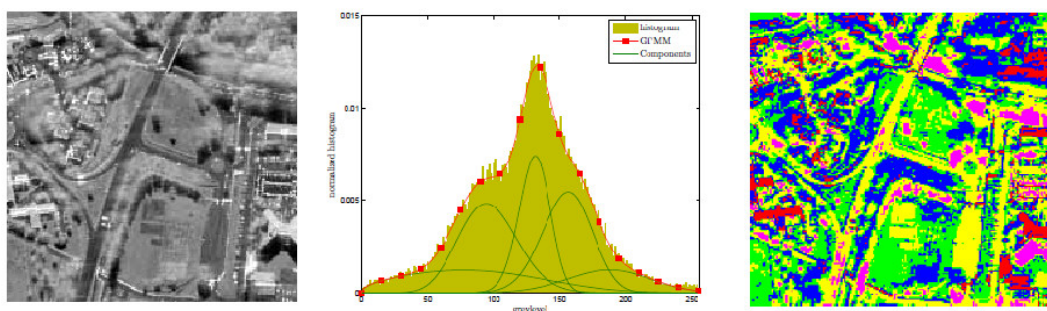


Figure 11. Statistical modeling of a RAMSES (©CNES, ONERA) image (left) by generalized Gamma mixture model (middle) and its visualization by maximum likelihood classification (right).

In Fig. 11 we demonstrate a typical result of the developed statistical modeling technique on a portion of a multilook airborne RAMSES (©CNES, ONERA) sensor acquisition over Toulouse suburbs (single polarization, downsampled to approximately 2m ground resolution). The unsupervised G $\Gamma$ MM 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.

## LEAR Project-Team

# 6. New Results

## 6.1. Highlights of the Year

- Cordelia Schmid received the **Longuet-Higgins prize** for fundamental contributions in computer vision that have withstood the test of time, 2014.
- We participated to the **Trecvid 2014 Multimedia Event Detection** challenge. We ranked first on one of the four tracks (Ad-hoc training videos with 10 examples per class).
- We participated to the **THUMOS 2014 challenge**. We obtained top ranked results in the localization track of the Thumos 2014 Action Recognition Challenge. The goal of the challenge is to evaluate large-scale action recognition in natural settings.

## 6.2. Visual recognition in images

### 6.2.1. Multi-fold MIL Training for Weakly Supervised Object Localization

**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 [13], we 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. This procedure is particularly important when high-dimensional representations, such as the Fisher vectors, are used. We present a detailed experimental evaluation using the PASCAL VOC 2007 and 2010 datasets. 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.

A journal paper is currently in preparation in which extends [13] 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.

### 6.2.2. Transformation Pursuit for Image Classification

**Participants:** Mattis Paulin, Jerome Revaud, Zaid Harchaoui, Florent Perronnin [XRCE], Cordelia Schmid.

In this work [19], [23], we use data augmentation (see Fig 2 for examples) to improve image classification performances in a large-scale context. A simple approach to learning invariances in image classification consists in augmenting the training set with transformed versions of the original images. However, given a large set of possible transformations, selecting a compact subset is challenging. Indeed, all transformations are not equally informative and adding uninformative transformations increases training time with no gain in accuracy. We propose a principled algorithm – Image Transformation Pursuit (ITP) – for the automatic selection of a compact set of transformations. ITP works in a greedy fashion, by selecting at each iteration the one that yields the highest accuracy gain. ITP also allows to efficiently explore complex transformations, that combine basic transformations. We report results on two public benchmarks: the CUB dataset of bird images and the ImageNet 2010 challenge. Using Fisher Vector representations, we achieve an improvement from 28.2% to 45.2% in top-1 accuracy on CUB, and an improvement from 70.1% to 74.9% in top-5 accuracy on ImageNet. We also show significant improvements for deep convnet features: from 47.3% to 55.4% on CUB and from 77.9% to 81.4% on ImageNet.



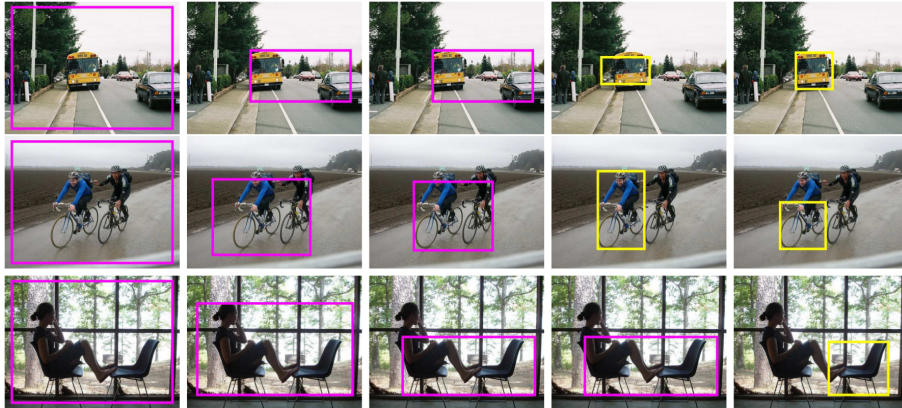


Figure 1. Illustration of our iterative object localization process on several example images, from initialization (left) to final localization (right). Yellow bounding boxes indicate that the object location hypothesis is in agreement with the ground-truth, for pink boxes the hypothesis is incorrect.

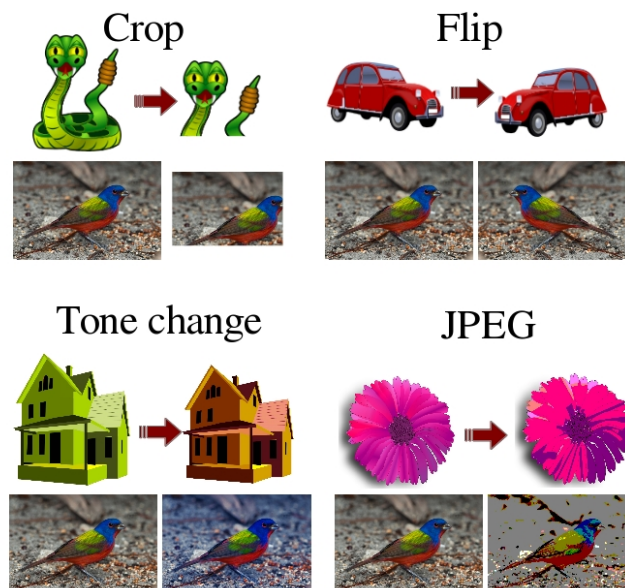


Figure 2. Examples of transformations used in [19], [23].

### 6.2.3. Convolutional Kernel Networks

**Participants:** Julien Mairal, Piotr Koniusz, Zaid Harchaoui, Cordelia Schmid.

An important goal in visual recognition is to devise image representations that are invariant to particular transformations. In this paper [16] we address this goal with a new type of convolutional neural network (CNN) whose invariance is encoded by a reproducing kernel. Unlike traditional approaches where neural networks are learned either to represent data or for solving a classification task, our network learns to approximate the kernel feature map on training data. Such an approach enjoys several benefits over classical ones. First, by teaching CNNs to be invariant, we obtain simple network architectures that achieve a similar accuracy to more complex ones, while being easy to train and robust to overfitting. Second, we bridge a gap between the neural network literature and kernels, which are natural tools to model invariance. We evaluate our methodology on visual recognition tasks where CNNs have proven to perform well, e.g., digit recognition with the MNIST dataset, and the more challenging CIFAR-10 and STL-10 datasets, where our accuracy is competitive with the state of the art. Figure 3 illustrates the architecture of our network.

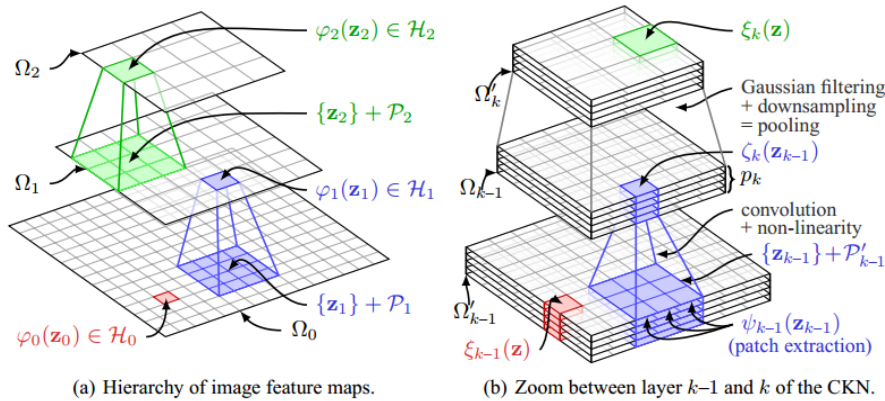


Figure 3. Left: concrete representation of the successive layers for the multilayer convolutional kernel. Right: one layer of the convolutional neural network that approximates the kernel.

### 6.2.4. Scene Text Recognition and Retrieval for Large Lexicons

**Participants:** Udit Roy [IIIT Hyderabad, India], Anand Mishra [IIIT Hyderabad, India], Karteek Alahari, C. v. Jawahar [IIIT Hyderabad, India].

In [21], we propose a framework for recognition and retrieval tasks in the context of scene text images. In contrast to many of the recent works, we focus on the case where an image-specific list of words, known as the small lexicon setting, is unavailable. We present a conditional random field model defined on potential character locations and the interactions between them. Observing that the interaction potentials computed in the large lexicon setting are less effective than in the case of a small lexicon, we propose an iterative method, which alternates between finding the most likely solution and refining the interaction potentials. We evaluate our method on public datasets and show that it improves over baseline and state-of-the-art approaches. For example, we obtain nearly 15% improvement in recognition accuracy and precision for our retrieval task over baseline methods on the IIIT-5K word dataset, with a large lexicon containing 0.5 million words.

### 6.2.5. On Learning to Localize Objects with Minimal Supervision

**Participants:** Hyun On Song [UC Berkeley], Ross Girshick [UC Berkeley], Stefanie Jegelka [UC Berkeley], Julien Mairal, Zaid Harchaoui, Trevor Darrell [UC Berkeley].

Learning to localize objects with minimal supervision is an important problem in computer vision, since large fully annotated datasets are extremely costly to obtain. In this paper [22], we propose a new method that achieves this goal with only image-level labels of whether the objects are present or not. Our approach combines a discriminative submodular cover problem for automatically discovering a set of positive object windows with a smoothed latent SVM formulation. The latter allows us to leverage efficient quasiNewton optimization techniques. Experimental results are presented in Figure 4 .

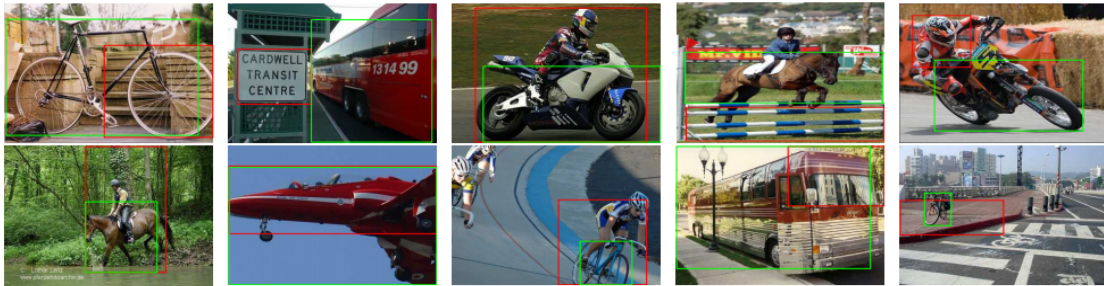


Figure 4. Visualization of some common failure cases of constructed positive windows by (Siva et al., 2012) vs our method. Red bounding boxes are constructed positive windows from (Siva et al., 2012). Green bounding boxes are constructed positive windows from our method.

### 6.2.6. Good Practice in Large-Scale Learning for Image Classification

**Participants:** Zeynep Akata, Florent Perronnin [XRCE], Zaid Harchaoui, Cordelia Schmid.

In this paper [3], we benchmark several SVM objective functions for large-scale image classification. We consider one-vs-rest, multi-class, ranking, and weighted approximate ranking SVMs. A comparison of online and batch methods for optimizing the objectives shows that online methods perform as well as batch methods in terms of classification accuracy, but with a significant gain in training speed. Using stochastic gradient descent, we can scale the training to millions of images and thousands of classes. Our experimental evaluation shows that ranking-based algorithms do not outperform the one-vs-rest strategy when a large number of training examples are used. Furthermore, the gap in accuracy between the different algorithms shrinks as the dimension of the features increases. We also show that learning through cross-validation the optimal rebalancing of positive and negative examples can result in a significant improvement for the one-vs-rest strategy. Finally, early stopping can be used as an effective regularization strategy when training with online algorithms. Following these “good practices”, we were able to improve the state-of-the-art on a large subset of 10K classes and 9M images of ImageNet from 16.7% Top-1 accuracy to 19.1%.

## 6.3. Learning and statistical models

### 6.3.1. Fast and Robust Archetypal Analysis for Representation Learning

**Participants:** Yuansi Chen, Julien Mairal, Zaid Harchaoui.

In [9], we revisit a pioneer unsupervised learning technique called archetypal analysis, which is related to successful data analysis methods such as sparse coding and non-negative matrix factorization. Since it was proposed, archetypal analysis did not gain a lot of popularity even though it produces more interpretable models than other alternatives. Because no efficient implementation has ever been made publicly available, its application to important scientific problems may have been severely limited. Our goal is to bring back into favour archetypal analysis. We propose a fast optimization scheme using an active-set strategy, and provide



an efficient open-source implementation interfaced with Matlab, R, and Python. Then, we demonstrate the usefulness of archetypal analysis for computer vision tasks, such as codebook learning, signal classification, and large image collection visualization.

In Figure 5, we present some archetypes corresponding to the request “Paris” when downloading 36 600 images uploaded in 2012 and 2013, and sorted by relevance on the Flickr website.



Figure 5. Classical landmarks appear on the left, which is not surprising since Flickr contains a large number of vacation pictures. In the middle, we display several archetypes that we did not expect, including ones about soccer, graffiti, food, flowers, and social gatherings. Finally, we display on the right some archetypes that do not seem to have some semantic meaning, but they capture some scene composition or texture that are common in the dataset.

### 6.3.2. Conditional Gradient Algorithms for Norm-Regularized Smooth Convex Optimization

**Participants:** Zaid Harchaoui, Anatoli Juditsky, Arkadii Nemirovski.

In this paper [6], we consider convex optimization problems arising in machine learning in high-dimensional settings. For several important learning problems, such as e.g. noisy matrix completion, state-of-the-art optimization approaches such as composite minimization algorithms are difficult to apply and do not scale up to large datasets. We study three conditional gradient-type algorithms, *i.e.* first-order optimization algorithms that require a linear minimization oracle but do not require a proximal oracle. These new algorithms are suitable for large-scale problems, and enjoy finite-time convergence guarantees. Promising experimental results are presented on two large-scale real-world datasets. The method is illustrated in Figure 6.

### 6.3.3. A Smoothing Approach for Composite Conditional Gradient with Nonsmooth Loss

**Participants:** Federico Pierucci, Zaid Harchaoui, Jérôme Malick [BIPOP Team, Inria].

In [25], we consider learning problems where the nonsmoothness lies both in the convex empirical risk and in the regularization penalty. Examples of such problems include learning with nonsmooth loss functions and atomic decomposition regularization penalty. Such doubly nonsmooth learning problems prevent the use of recently proposed composite conditional gradient algorithms for training, which are particularly attractive for large-scale applications. Indeed, they rely on the assumption that the empirical risk part of the objective is smooth. We propose a composite conditional gradient algorithm with smoothing to tackle such learning

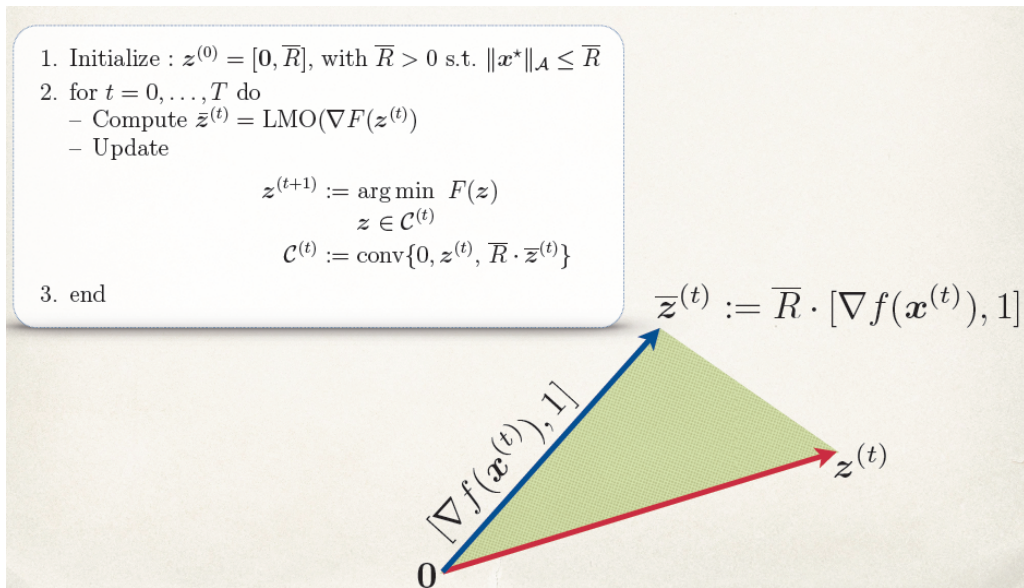


Figure 6. Overview of the composite conditional gradient algorithm which minimizes  $F(x) := f(x) + \lambda \|x\|_{\mathcal{A}}$ , where  $f$  is smooth and  $\|\cdot\|_{\mathcal{A}}$  is an atomic-decomposition norm.

problems. We set up a framework allowing to systematically design parametrized smooth surrogates of nonsmooth loss functions. We then propose a smoothed composite conditional gradient algorithm, for which we prove theoretical guarantees on the accuracy. We present promising experimental results on collaborative filtering tasks (see Figure 7).

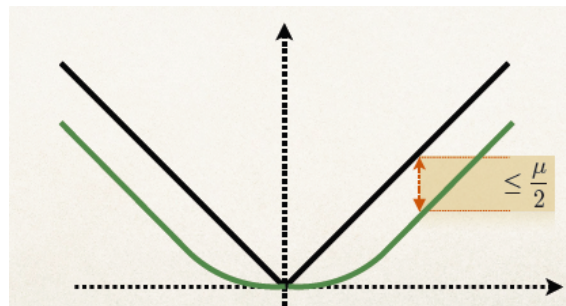


Figure 7. Illustration of the smooth surrogate with parameter  $\mu$  (green) of the absolute value function (black).

#### 6.3.4. Incremental Majorization-Minimization Optimization with Application to Large-Scale Machine Learning

**Participant:** Julien Mairal.

In this paper [27], we study optimization methods consisting of iteratively minimizing surrogates of an objective function, as illustrated in Figure 8. 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.

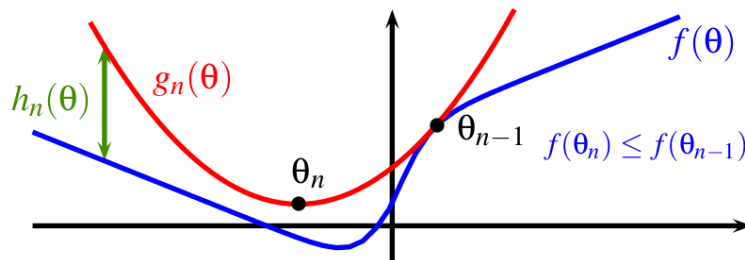


Figure 8. Illustration of the basic majorization-minimization principle. We compute a surrogate  $g_n$  of the objective function  $f$  around a current estimate  $\theta_{n-1}$ . The new estimate  $\theta_n$  is a minimizer of  $g_n$ . The approximation error  $h_n$  is smooth.

### 6.3.5. Efficient RNA Isoform Identification and Quantification from RNA-Seq Data with Network Flows

**Participants:** Elsa Bernard [Institut Curie, Ecoles des Mines-ParisTech], Laurent Jacob [CNRS, LBBE Laboratory], Julien Mairal [correspondant], Jean-Philippe Vert [Institut Curie, Ecoles des Mines-ParisTech].

Several state-of-the-art methods for isoform identification and quantification are based on  $\ell_1$ -regularized regression, such as the Lasso. However, explicitly listing the—possibly exponentially—large set of candidate transcripts is intractable for genes with many exons. For this reason, existing approaches using the  $\ell_1$ -penalty are either restricted to genes with few exons or only run the regression algorithm on a small set of preselected isoforms. In [4], we introduce a new technique called FlipFlop, which can efficiently tackle the sparse estimation problem on the full set of candidate isoforms by using network flow optimization. Our technique removes the need of a preselection step, leading to better isoform identification while keeping a low computational cost. Experiments with synthetic and real RNA-Seq data confirm that our approach is more accurate than alternative methods and one of the fastest available. Figure 9 presents the graph on which the network flow optimization is performed.

### 6.3.6. Riemannian Sparse Coding for Positive Definite Matrices

**Participants:** Anoop Cherian, Suvrit Sra [MPI].

Inspired by the great success of sparse coding for vector valued data, our goal in this work [12] is to represent symmetric positive definite (SPD) data matrices as sparse linear combinations of atoms from a dictionary, where each atom itself is an SPD matrix. Since SPD matrices follow a non-Euclidean (in fact a Riemannian) geometry, existing sparse coding techniques for Euclidean data cannot be directly extended. Prior works have approached this problem by defining a sparse coding loss function using either extrinsic similarity measures (such as the log-Euclidean distance) or kernelized variants of statistical measures (such as the Stein divergence, Jeffrey's divergence, etc.). In contrast, we propose to use the intrinsic Riemannian distance on the manifold of SPD matrices. Our main contribution is a novel mathematical model for sparse coding of SPD matrices; we also present a computationally simple algorithm for optimizing our model. Experiments on several computer vision datasets showcase superior classification and retrieval performance compared against state-of-the-art approaches.



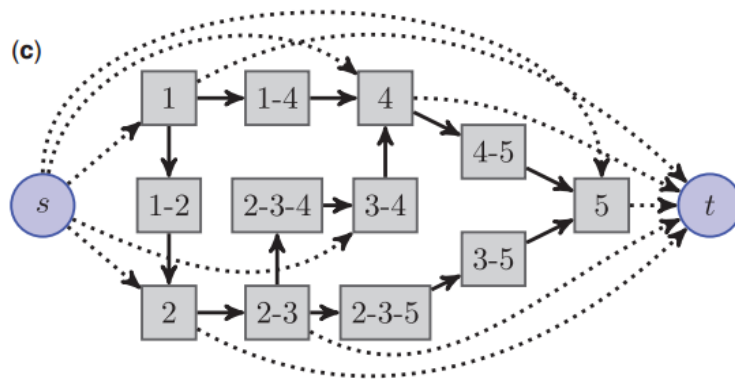


Figure 9. Graph on which we perform network flow optimization. Nodes represent observed reads, and paths on the graph correspond to isoforms.

## 6.4. Recognition in video

### 6.4.1. Occlusion and Motion Reasoning for Long-Term Tracking

**Participants:** Yang Hua, Karteek Alahari, Cordelia Schmid.

Object tracking is a reoccurring problem in computer vision. Tracking-by-detection approaches, in particular Struck, have shown to be competitive in recent evaluations. However, such approaches fail in the presence of long-term occlusions as well as severe viewpoint changes of the object. In this paper we propose a principled way to combine occlusion and motion reasoning with a tracking-by-detection approach. Occlusion and motion reasoning is based on state-of-the-art long-term trajectories which are labeled as object or background tracks with an energy-based formulation. The overlap between labeled tracks and detected regions allows to identify occlusions. The motion changes of the object between consecutive frames can be estimated robustly from the geometric relation between object trajectories. If this geometric change is significant, an additional detector is trained. Experimental results show that our tracker obtains state-of-the-art results and handles occlusion and viewpoints changes better than competing tracking methods. This work corresponds to the publication [15] and is illustrated in Figure 10.

### 6.4.2. Category-Specific Video Summarization

**Participants:** Danila Potapov, Matthijs Douze, Zaid Harchaoui, Cordelia Schmid.

In large video collections with clusters of typical categories, such as “birthday party” or “flash-mob”, category-specific video summarization can produce higher quality video summaries than unsupervised approaches that are blind to the video category. Given a video from a known category, our approach published in [20] first efficiently performs a temporal segmentation into semantically-consistent segments, delimited not only by shot boundaries but also general change points. Then, equipped with an SVM classifier, our approach assigns importance scores to each segment. The resulting video assembles the sequence of segments with the highest scores, as shown in Figure 11. The obtained video summary is therefore both short and highly informative. Experimental results on videos from the multimedia event detection (MED) dataset of TRECVID’11 show that our approach produces video summaries with higher relevance than the state of the art.

### 6.4.3. Efficient Action Localization with Approximately Normalized Fisher Vectors

**Participants:** Dan Oneata, Jakob Verbeek, Cordelia Schmid.

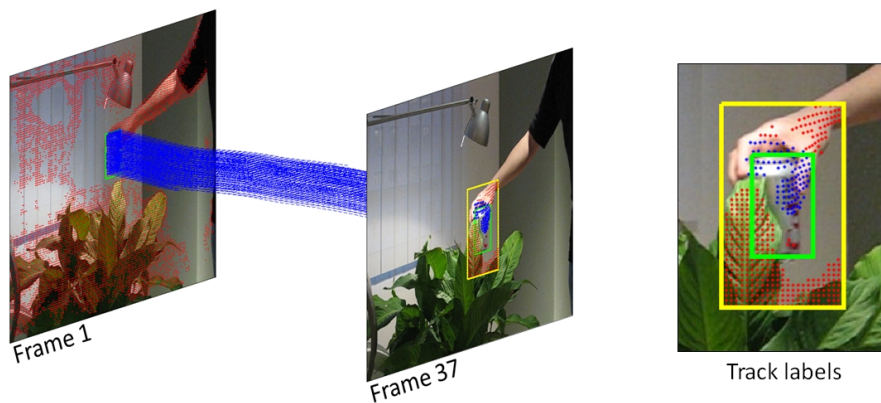


Figure 10. Left: Long-term tracks beginning in frame 1 of the Coke sequence. The yellow box shows the search region used to compute the bounding box most likely to contain the object (green box). We use the tracks to estimate the object state. Right: Close-up of the track labels in frame 37. Here, less than 60% of the tracks within the predicted bounding box are assigned to the object (blue), and the remaining are labelled as background (red). Thus, the object is predicted to be in an occluded state.



Figure 11. Original video, and its video summary for the category “birthday party”.

The Fisher vector (FV) representation is a high-dimensional extension of the popular bag-of-words representation. Transformation of the FV by power and  $\ell_2$  normalizations has shown to significantly improve its performance, and led to state-of-the-art results for a range of image and video classification and retrieval tasks. These normalizations, however, render the representation non-additive over local descriptors. Combined with its high dimensionality, this makes the FV computationally expensive for the purpose of localization tasks. In [18] we present approximations to both these normalizations (see Figure 12), which yield significant improvements in the memory and computational costs of the FV when used for localization. Second, we show how these approximations can be used to define upper-bounds on the score function that can be efficiently evaluated, which enables the use of branch-and-bound search as an alternative to exhaustive sliding window search. We present experimental evaluation results on classification and temporal localization of actions in videos. These show that our approximations lead to a speedup of at least one order of magnitude, while maintaining state-of-the-art action recognition and localization performance.

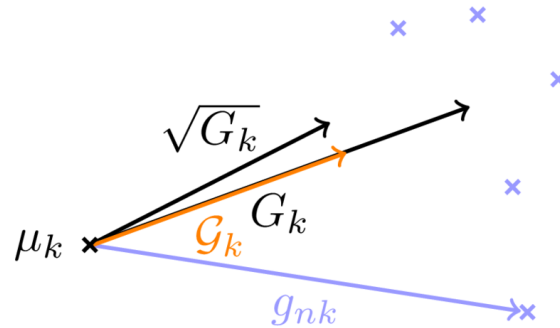


Figure 12. Schematic illustration of the proposed approximation for the square-root normalization. We depict a Fisher vector  $G_k$  as an aggregation of individual gradients  $g_{nk}$ . Both the exact ( $\sqrt{G_k}$ ) and the approximated ( $\mathcal{G}_k$ ) square-root normalizations scale similarly the Fisher vector  $G_k$ ; the approximated variant has the property of preserving the orientation of the Fisher vector  $G_k$ .

#### 6.4.4. Spatio-Temporal Object Detection Proposals

**Participants:** Dan Oneata, Jakob Verbeek, Cordelia Schmid, Jerome Revaud.

Spatio-temporal detection of actions and events in video is a challenging problem. Besides the difficulties related to recognition, a major challenge for detection in video is the size of the search space defined by spatio-temporal tubes formed by sequences of bounding boxes along the frames. Recently methods that generate unsupervised detection proposals have proven to be very effective for object detection in still images. These methods open the possibility to use strong but computationally expensive features since only a relatively small number of detection hypotheses need to be assessed. In [17] we make two contributions towards exploiting detection proposals for spatio-temporal detection problems. First, we extend a recent 2D object proposal method, to produce spatio-temporal proposals by a randomized supervoxel merging process (see Figure 13). We introduce spatial, temporal, and spatio-temporal pairwise supervoxel features that are used to guide the merging process. Second, we propose a new efficient supervoxel method. We experimentally evaluate our detection proposals, in combination with our new supervoxel method as well as existing ones. This evaluation shows that our supervoxels lead to more accurate proposals when compared to using existing state-of-the-art supervoxel methods.

#### 6.4.5. EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow

**Participants:** Revaud Jerome, Weinzaepfel Philippe, Harchaoui Zaid, Cordelia Schmid.

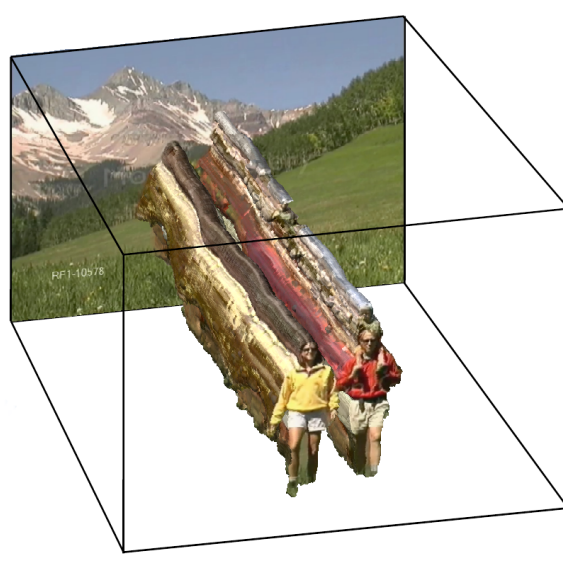


Figure 13. An example of a spatio-temporal proposal generated by our randomized supervoxel merging process. The video sample belongs to the UCF Sports dataset.

We propose a novel approach [29] 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 (see Figure 14), 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. It significantly outperforms the state of the art on MPI-Sintel and performs on par on KITTI and Middlebury.

#### 6.4.6. Weakly Supervised Action Labeling in Videos Under Ordering Constraints.

**Participants:** Piotr Bojanowski [Willow team, Inria], Rémi Lajugie [Willow team, Inria], Francis Bach [Sierra team, Inria], Ivan Laptev [Willow team, Inria], Jean Ponce [Willow team, Inria], Cordelia Schmid, Josef Sivic [Willow team, Inria].

Suppose we are given a set of video clips, each one annotated with an ordered list of actions, such as “walk” then “sit” then “answer phone” extracted from, for example, the associated text script. See Fig. 15 for an illustration. In this work [8], we seek to temporally localize the individual actions in each clip as well as to learn a discriminative classifier for each action. We formulate the problem as a weakly supervised temporal assignment with ordering constraints. Each video clip is divided into small time intervals and each time interval of each video clip is assigned one action label, while respecting the order in which the action labels appear in the given annotations. We show that the action label assignment can be determined together with learning a classifier for each action in a discriminative manner. We evaluate the proposed model on a new and challenging dataset of 937 video clips with a total of 787720 frames containing sequences of 16 different actions from 69 Hollywood movies.

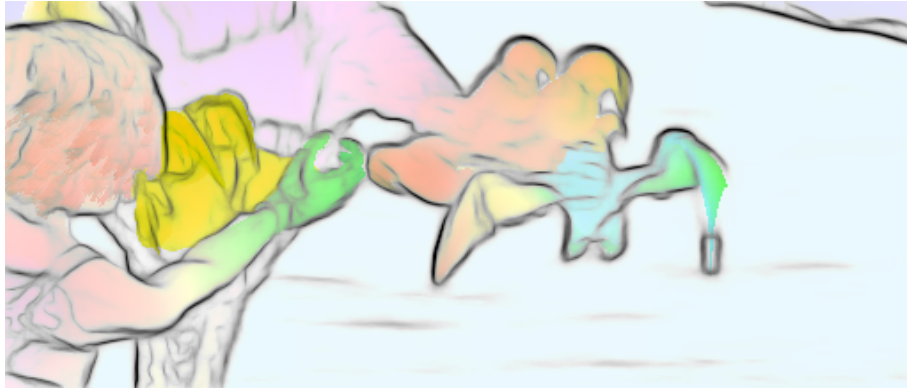


Figure 14. Image edges detected with SED and ground-truth optical flow. Motion discontinuities appear most of the time at image edges

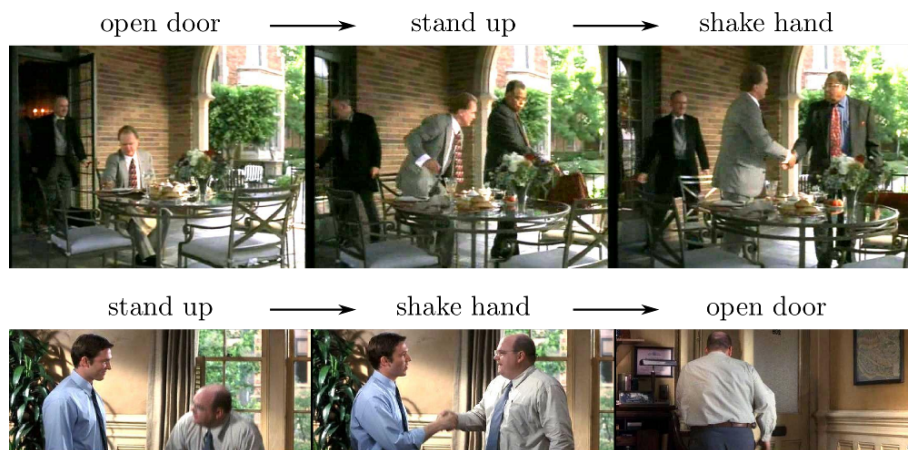


Figure 15. Sample data used as input to our method. Every video clip comes with an ordered list of actions that appears in it. These actions are not temporally localized, only the order is known. The goal of our paper is to correctly localize these actions according to a discriminative criterion.



### 6.4.7. Mixing Body-Part Sequences for Human Pose Estimation

**Participants:** Cherian Anoop, Mairal Julien, Alahari Karteek, Schmid Cordelia.

This work [11] presents a method for estimating articulated human poses in videos. We cast this as an optimization problem defined on body parts with spatio-temporal links between them. The resulting formulation is unfortunately intractable and previous approaches only provide approximate solutions. Although such methods perform well on certain body parts, e.g., head, their performance on lower arms, i.e., elbows and wrists, remains poor. We present a new approximate scheme with two steps dedicated to pose estimation. First, our approach takes into account temporal links with subsequent frames for the less-certain parts, namely elbows and wrists. Second, our method decomposes poses into limbs, generates limb sequences across time, and re-composes poses by mixing these body part sequences (See Figure 16 for an illustration). We introduce a new dataset "Poses in the Wild", which is more challenging than the existing ones, with sequences containing background clutter, occlusions, and severe camera motion. We experimentally compare our method with recent approaches on this new dataset as well as on two other benchmark datasets, and show significant improvement.

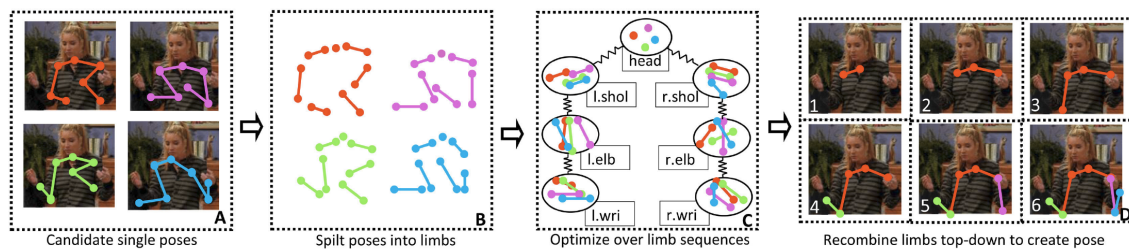


Figure 16. Illustration of our limb recombination scheme. From left to right: Block-A: An image and four candidate poses, where only a part of each pose is well-aligned with the person. Block-B: We divide each candidate pose into limb parts. Block-C: We allow the recombination of limbs from different pose candidates with constraints between two limbs that have a joint in common. Block-D: An example where recombination builds an accurate pose, which is not in the original candidate set.

### 6.4.8. The LEAR Submission at Thumos 2014

**Participants:** Dan Oneata, Jakob Verbeek, Cordelia Schmid.

In [28] we describe the submission of our team to the THUMOS workshop in conjunction with ECCV 2014. Our system is based on Fisher vector (FV) encoding of dense trajectory features (DTF), which we also used in our 2013 submission. The dataset is based on the UCF101 dataset, which is currently the largest action dataset both in terms of number of categories and clips, with more than 13000 clips drawn from 101 action classes. This year special attention was paid to classification of uncropped videos, where the action of interest appears in videos that contain also non-relevant sections. This year's submission additionally incorporated static-image features (SIFT, Color, and CNN) and audio features (ASR and MFCC) for the classification task. For the detection task, we combined scores from the classification task with FV-DTF features extracted from video slices. We found that these additional visual and audio feature significantly improve the classification results. For localization we found that using the classification scores as a contextual feature besides local motion features leads to significant improvements. In Figure 17 we show the middle frame from the top four ranked videos corresponding to the three hardest classes (as evaluated on the validation data). Our team has ranked second on the classification challenge (out of eleven teams) and first on the detection challenge (out of three teams).



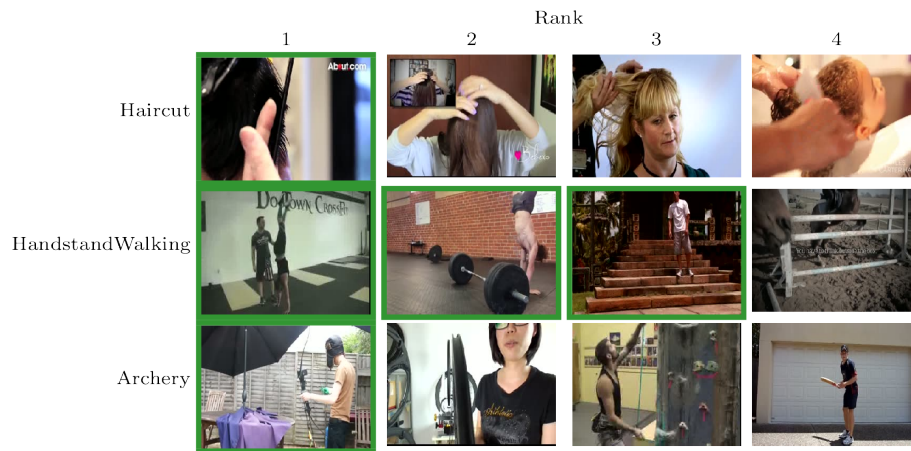


Figure 17. Snapshots from the top four ranked test videos for the three hardest classes; green borders indicate true positives.

#### 6.4.9. The LEAR Submission at TrecVid MED 2014

**Participants:** Matthijs Douze, Dan Oneata, Mattis Paulin, Clément Leray, Nicolas Chesneau, Danila Potapov, Jakob Verbeek, Karteek Alahari, Zaid Harchaoui, Lori Lamel [Spoken Language Processing group, LIMSI, CNRS], Jean-Luc Gauvain [Spoken Language Processing group, LIMSI, CNRS], Christoph Schmidt [Fraunhofer IAIS, Sankt Augustin], Cordelia Schmid.

In [26] we describe our participation to the 2014 edition of the TrecVid Multimedia Event Detection task. Our system is based on a collection of local visual and audio descriptors, which are aggregated to global descriptors, one for each type of low-level descriptor, using Fisher vectors. Besides these features, we use two features based on convolutional networks: one for the visual channel, and one for the audio channel. Additional high-level features are extracted using ASR and OCR features. Finally, we used mid-level attribute features based on object and action detectors trained on external datasets. In the notebook paper we present an overview of the features and the classification techniques, and experimentally evaluate our system on TrecVid MED 2011 data.

We participated in four tasks, which differ in the amount of training videos for each event (either 10 or 100), and the time that is allowed for the processing. For the 20 pre-specified events several weeks are allowed to extract features, train models, and to score the test videos (which consisted of 8,000 hours of video this year). For the 10 ad-hoc events, we only have five days to do all processing. Across the 11 participating teams, our results ranked first for the 10-example ad-hoc task, and fourth and fifth place for the other tasks.

## LINKMEDIA Project-Team

# 6. New Results

## 6.1. Highlights of the Year

BEST PAPER AWARD :

[21] **European Symposium on Research in Computer Security**. R. BALU, T. FURON, S. GAMBS.

## 6.2. Unsupervised motif discovery

### 6.2.1. *Clustering by diverting supervised machine learning*

**Participants:** Vincent Claveau, Patrick Gros, Abir Ncibi.

Knowledge discovery aims at bringing out coherent groups of objects and is usually based on clustering which necessitates defining a notion of similarity between objects. In practice, this strong prior is often neither possible nor welcome. We proposed to divert supervised machine learning (ML) techniques in order to calculate, indirectly and without supervision, similarities among objects. Our approach consists in generating artificial labeling problems on the data to reveal regularities between objects through their labeling. In [28], we show how this framework can be implemented and experimented on two information extraction/discovery tasks concerned with named entities. The ML technique diverted to exhibit similarities between with the named entities are the Conditional Random Fields. This same method can also be applied with less common ML techniques: In [59], we show that Inductive Logic Programming can also be used to cluster complex data. Thanks to the ability of ILP to handle data that cannot be expressed under the usual attribute-value representation, we use it to make emerge clusters of TV broadcasts based only on their broadcasting information (date, time, length, etc.).

### 6.2.2. *Spoken term discovery applied to audio thumbnailing*

**Participants:** Sébastien Campion, Guillaume Gravier.

We evaluated a system to create audio thumbnails of spoken content, i.e., short audio summaries representative of the entire content, without resorting to a lexical representation. As an alternative to searching for relevant words and phrases in a transcript, unsupervised motif discovery is used to find short, word-like, repeating fragments at the signal level without acoustic models. The output of the word discovery algorithm is exploited via a maximum motif coverage criterion to generate a thumbnail in an extractive manner. A limited number of relevant segments are chosen within the data so as to include the maximum number of motifs while remaining short enough and intelligible. Evaluation is performed on broadcast news reports with a panel of human listeners judging the quality of the thumbnails. Results indicate that motif-based thumbnails stand between random thumbnails and ASR-based keywords, however still far behind thumbnails and keywords humanly authored [34].

### 6.2.3. *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.*

Unsupervised approaches were introduced a few years ago to analyze the structure of TV programs, relying on the discovery of repeated elements within a program or across multiple episodes of the same program. These methods can discover key repeating elements, such as jingles and separators, however they cannot infer the entire structure of a program. In [48], we studied a hierarchical use of grammatical inference to yield a temporal grammar of a program from a collection of episodes, discovering both the vocabulary of the grammar and the temporal organization of the words from the vocabulary. Using a set of basic event detectors and simple filtering techniques to detect repeating elements of interest, a symbolic representation of each episode is derived based on minimal domain knowledge. Grammatical inference based on multiple sequence alignment is then used in a hierarchical manner to provide a temporal grammar of the program at various levels of details.

### 6.2.4. Efficient indexing for content retrieval

**Participants:** Raghavendran Balu, Teddy Furon, Hervé Jégou.

*In collaboration with Miajing Shi during, visiting Ph. D. student from Pekin University.*

Many nearest neighbor search algorithms rely on encoding real vectors into binary vectors. The most common strategy projects the vectors onto random directions and takes the sign to produce so-called sketches. In [22], we discuss the sub-optimality of this choice, and propose a better encoding strategy based on the quantization and reconstruction points of view. Our second contribution is a novel asymmetric estimator for the cosine similarity. Similar to previous asymmetric schemes, the query is not quantized and the similarity is computed in the compressed domain. We tackled the same similarity estimation problem with a rather different approach in [52], where we assume that only a few vectors of the database, so-called heavy hitters, have a similarity to the query that significantly deviates from 0. For this purpose, we have introduced a group testing framework for detecting large similarities between high-dimensional vectors, such as descriptors used in state-of-the-art description of multimedia documents. We produce a set of group representations that jointly encode several vectors into a single one, in the spirit of group testing approaches. By comparing a query vector to several of these intermediate representations, we screen the large values taken by the similarities between the query and all the vectors, at a fraction of the cost of exhaustive similarity calculation. Unlike concurrent indexing methods that suffer from the curse of dimensionality, our method exploits the properties of high-dimensional spaces.

## 6.3. Multimedia content description and structuring

### 6.3.1. Linguistic knowledge extraction

#### 6.3.1.1. Identifying events in texts

**Participant:** Vincent Claveau.

*In collaboration with Béatrice Arnulphy, former team member now with ANR, Xavier Tannier and Anne Vilnat, LIMSI.*

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 was available for French. In [20], we tried 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 and set new standard for French.

#### 6.3.1.2. Morpho-semantic analysis of terms

**Participants:** Vincent Claveau, Ewa Kijak.

In most Indo-European languages, many biomedical terms are rich morphological structures composed of several constituents mainly originating from Greek or Latin. The interpretation of these compounds are keystones to access information. Following our work on morphology in the biomedical domain, we proposed different techniques to generate probabilistic morph-semantic resources and we show how these alignment information can be used for segmenting compounds, attaching semantic interpretation to each part, proposing definitions (gloses) of the compounds [26]. When possible, these tasks are compared with state-of-the-art tools, and the results show the interest of our automatically built probabilistic resources.

#### 6.3.1.3. Distributional semantics

**Participants:** Vincent Claveau, Ewa Kijak.

*In collaboration with Olivier Ferret, CEA-LIST.*

We addressed the issue of building and improving a distributional thesaurus. We first show that existing tools from the information retrieval domain can be directly used in order to build a thesaurus with state-of-the-art performance. Secondly, we focus more specifically on improving the obtained thesaurus, seen as a graph of  $k$ -nearest neighbors. By exploiting information about the neighborhood contained in this graph, we propose several contributions. 1) We show how the lists of neighbors can be globally improved by examining the reciprocity of the neighboring relation, that is, the fact that a word can be close to another and vice-versa. 2) We also propose a method to associate a confidence score to any lists of nearest neighbors (i.e., any entry of the thesaurus). 3) Last, we demonstrate how these confidence scores can be used to reorder the closest neighbors of a word. These different contributions are validated through experiments and offer significant improvement over the state-of-the-art [27], [60].

### 6.3.2. OCR and speech content-based description

#### 6.3.2.1. Use of stress information for robust speech recognition

**Participant:** Guillaume Gravier.

*In collaboration with S. Ziegler, PANAMA Inria team and Laboratoire de Sciences Cognitives et Psycholinguistique.*

[44] presents a study on the robustness of stress information for automatic speech recognition in the presence of noise. The syllable stress, extracted from the speech signal, was integrated in the recognition process by means of a previously proposed decoding method. Experiments were conducted for several signal-to-noise ratio conditions and the results show that stress information is robust in the presence of medium to low noise. This was found to be true both when syllable boundary information was used for stress detection and when this information was not available. Furthermore, the obtained relative improvement increased with a decrease in signal quality, indicating that the stressed parts of the signal can be considered islands of reliability.

#### 6.3.2.2. Boosting bonsai trees for handwritten/printed text discrimination

**Participant:** Christian Raymond.

*In collaboration with Yann Ricquebourg, Baptiste Poirriez, Aurélie Lemaitre and Bertrand Coüasnon, IRISA.*

Boosting over decision-stumps proved its efficiency in natural language processing, essentially with symbolic features, and its good properties (fast, few and not critical parameters, not sensitive to overfitting) could be of great interest in the numeric world of pixel images. In [51], we investigated the use of boosting over small decision trees in image classification processing for the discrimination of handwritten/printed text. We conducted experiments to compare with usual SVM-based classification revealing convincing results with very close performance, but with faster predictions and behaving far less as a black-box. Those promising results tend to make use of this classifier in more complex recognition tasks like multiclass problems.

#### 6.3.2.3. Speaker role detection from spoken document

**Participant:** Christian Raymond.

*In collaboration with LIMSI and LIUM.*

In [40] and [41], we tackle the problem of speaker role detection in broadcast news shows. In the literature, many proposed solutions are based on the combination of various features coming from acoustic, lexical and semantic information with a machine learning algorithm. Many previous studies mention the use of boosting over decision stumps to combine efficiently these features. We proposed a modification of this state-of-the-art machine learning algorithm changing the weak learner (decision stumps) by small decision trees, denoted bonsai trees. Experiments show that using bonsai trees as weak learners for the boosting algorithm largely reduces both system error rate and learning time.

### 6.3.3. Image and video description and classification

#### 6.3.3.1. Fine-grain image classification

**Participants:** Teddy Furon, Philippe-Henri Gosselin, Hervé Jégou.

*In collaboration with Xerox Research Center Europe.*

We have addressed the problem of instance classification: our goal is to annotate images with tags corresponding to objects classes which exhibit small intra-class variations such as logos, products or landmarks. Our first contribution on image classification [13] describes the processing pipeline, which has won FGCOMP challenge associated with Imagenet. It improves a standard method based on Fisher vectors to adapt it to the context of fine-grained classes, where the difference between classes rely on few but typical visual differences. On the same task, we have proposed a novel algorithm [39] for the selection of class-specific *prototypes* which are used in a voting-based classification scheme.

#### 6.3.3.2. Aggregation of local descriptors

**Participants:** Teddy Furon, Hervé Jégou, Giorgos Toliás.

*In collaboration with the University of Oxford.*

For unsupervised particular object and image recognition, we have considered the design of a single vector representation for an image that embeds and aggregates a set of local patch descriptors such as SIFT. In [36], we make two contributions, both aimed at regularizing the individual contributions of the local descriptors in the final representation. The first is a novel embedding method that avoids the dependency on absolute distances by encoding directions. The second contribution is a “democratization” strategy that further limits the interaction of unrelated descriptors in the aggregation stage. In [36], we addressed another issue inherent to existing encoding algorithms: Image search systems based on local descriptors typically achieve orientation invariance by aligning the patches on their dominant orientations. This choice introduces too much invariance because it does not guarantee that the patches are rotated consistently. To address this problem, we have introduced another aggregation strategy of local descriptors that achieves this covariance property by jointly encoding the angle in the aggregation stage in a continuous manner. It is combined with an efficient monomial embedding to provide a codebook-free method to aggregate local descriptors into a single vector representation.

#### 6.3.3.3. Action localization in videos

**Participants:** Mihir Jain, Hervé Jégou.

*In collaboration with the University of Amsterdam and the project-team SERPICO.*

We have tackled the problem of action localization in videos [35], where the objective is to determine when and where certain actions appear. We introduce a sampling strategy, called tubelets and inspired a method recently introduced for image detection. It drastically reduces the number of hypotheses that are likely to include the action of interest. By using super-voxels and employing a criterion that reflects how action related motion deviates from background motion, the method is specifically adapted to 2D+t sequences and establishes the new state-of-the-art for action localization on the public datasets UCF Sports and MSR-II.

#### 6.3.4. Text description for information retrieval

**Participants:** Vincent Claveau, Sébastien Le Maguer.

*In collaboration with Natalia Grabar, STL UMR8163, and Thierry Hamon, LIMSI*

Following previous work, we investigated the interest of “bag of bags of features” representation for texts in an vector-space information retrieval setting. Each text is thus represented as a bag of vector. With this representation, computing the similarity between two texts necessitates to aggregate every vector to vector similarity for the two bags. In [58], we examine the expected properties of such an aggregation function and show their influence through different experiments. When some specific conditions are met, we show that the gains over standard representation can be very important.

With a team composed with members of TEXMEX/LINKMEDIA, LIMSI and STL, we have participated to the biomedical information retrieval challenge proposed in the framework of CLEF eHealth [25]. For this first participation, our approach relies on a state-of-the-art IR system called Indri, based on statistical language modeling, and on semantic resources. The purpose of semantic resources and methods is to manage the term variation such as synonyms, morpho-syntactic variants, abbreviation or nested terms. Different combinations of resources and Indri settings are explored, mostly based on query expansion. We obtained good overall results (3rd in terms of MAP) and confirmed the interest of query expansion to retrieve a maximum of relevant documents.

## 6.4. Linking, navigation and analytics

### 6.4.1. NLP-driven hyperlink construction in broadcast videos

**Participants:** Rémi Bois, Vincent Claveau, Guillaume Gravier, Pascale Sébillot, Anca-Roxana Şimon.

*In collaboration with Sien Moens, Katholieke Universiteit Leuven, Éric Jamet and Martin Ragot, Univ. Rennes 2.*

The hyperlinking sub-task of the MediaEval Search and Hyperlinking task aims at creating hyperlinks between predefined anchor segments, i.e., fragments of videos, and short related video segments, called targets, that have to be automatically extracted from videos of a large collection. Capitalizing on the experience acquired in previous participations [54], we proposed a two step approach exploiting speech material: Potential target segments are first generated relying on a topic segmentation technique; For each anchor, the best 20 target segments are then selected according to two distinct strategies. The first strategy focuses on the identification of very similar targets using n-grams and named entities, while the second one makes use of an intermediate structure built from topic models, which offers the possibility to control serendipity and to explain the links created [53].

In 2014, we also initiated the CominLabs project “Linking media in acceptable hypergraphs” dedicated to the creation of explicit and meaningful links between multimedia documents or fragments of documents. Two main issues were addressed: The construction of a corpus, composed of audio and video news, reports and debates, newspapers and blog websites, as well as social networks; A preliminary study of the perceived usefulness of various types of links by end-users.

### 6.4.2. Analytics in collections of art critics

**Participant:** Vincent Claveau.

*In collaboration with Fabienne Moreau and Nicolas Thély, Univ. Rennes 2.*

We aim at exploiting text mining techniques in the service of digital humanities, and more precisely in the field of art criticism. It relies on a collaboration between our team, linguists and art and aesthetics specialists. In preliminary work [56], we adapted term extraction, named entity recognition and information retrieval techniques to this field to extract multiple linguistic clues from art review articles. Future work will make the most of these clues and clustering approaches to build a navigable and structured collection of the articles.

### 6.4.3. Data models for navigation

**Participant:** Laurent Amsaleg.

*In collaboration with Björn Þór Jónsson, Grímur Tómasson, Hlynur Sigurþórsson, Áslaug Eriksdóttir and Marta Kristin Larusdóttir, School of Computer Science, Reykjavík University.*

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 propose a multi-dimensional model for media browsing, called  $M^3$ , that combines MDA concepts with concepts from faceted browsing. 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 [38]. A photo navigation prototype was demonstrated at the Intl. Conf. on Multimedia Modeling [37].

### 6.4.4. Exploiting k-nn graphs for image retrieval

**Participants:** Laurent Amsaleg, Hervé Jégou, Giorgos Tolias.



We have proposed two techniques exploiting the relationship between the images with a collection. In [29], we revisit how to exploit the  $k$ -reciprocal nearest neighbors to produce, for a given query, a neighborhood that improves over the one obtained with the original metric. This strategy is simpler than concurrent prior work, yet it is both effective and less sensitive to parameters. Second, we propose to employ measures defined on sets of shared nearest neighbors in order to re-rank the shortlist. Both methods are simple, yet they significantly improve the accuracy of image search engines on standard benchmarks. We also introduced a query expansion technique [18] for image search that is faster and more precise than the existing ones. The expansion generates an enriched representation which refines the initial local descriptors *individually* by aggregating those of the database, while new descriptors are produced from the images that are deemed relevant. The technique has two computational advantages over other query expansion techniques. First, the size of the enriched representation is comparable to that of the initial query. Second, the technique is effective even without using any geometry, in which case searching a database comprising 105k images typically takes 80 ms on a desktop machine. Overall, our technique significantly outperforms the visual query expansion state of the art on popular benchmarks.

## 6.5. Content-based information retrieval and security

### 6.5.1. Analysis of privacy preserving data aggregation for recommendation systems

**Participants:** Raghavendran Balu, Teddy Furon.

*Work under the Alcatel-Lucent / Inria common Lab, in collaboration with Armen Aghasaryan, Dimitre Davidov, Makram Bouzid (ALU) and Sébastien Gambis (Inria Rennes Cidre team project).*

We consider personalized recommendation systems in which before publication, the profile of a user is sanitized by a non-interactive mechanism compliant with the concept of differential privacy. We analyze two existing schemes offering a differentially private representation of profiles: BLIP (BLoom-and-FLIP) and JLT (Johnson-Lindenstrauss Transform). For assessing their security levels, we play the role of an adversary aiming at reconstructing a user profile [21]. We compare two inference attacks, namely single and joint decoding. The first one decides of the presence of a single item in the profile, and sequentially explores all the item set. The second one decides whether a subset of items is likely to be the user profile, and considers all the possible subsets. Our contributions are a theoretical analysis as well as a practical implementation of both attacks, which were evaluated on datasets of real user profiles. The results obtained clearly demonstrates that joint decoding is the most powerful attack, while also giving useful insights on how to set the differential privacy parameter  $\epsilon$ .

### 6.5.2. Content based image retrieval with privacy

**Participants:** Laurent Amsaleg, Teddy Furon, Li Weng.

*Work initiated during a collaboration with A. Morton, L. Weng (with LINKMEDIA since May 2014) and S. Marchand-Maillet, Université de Genève.*

We propose a privacy protection framework for large-scale content-based information retrieval. It offers two layers of protection. First, robust hash values are used as queries to prevent revealing original content or features. Second, the client can choose to omit certain bits in a hash value to further increase the ambiguity for the server. Due to the reduced information, it is computationally difficult for the server to know the client's interest. The server has to return the hash values of all possible candidates to the client. The client performs a search within the candidate list to find the best match. Since only hash values are exchanged between the client and the server, the privacy of both parties is protected. We introduce the concept of tunable privacy, where the privacy protection level can be adjusted according to a policy. It is realized through hash-based piece-wise inverted indexing. The idea is to divide a feature vector into pieces and index each piece with a sub-hash value. Each sub-hash value is associated with an inverted index list. The framework has been extensively tested using a large image database. We have evaluated both retrieval performance and privacy-preserving performance for a particular content identification application.

### 6.5.3. Privacy protection in face recognition

**Participant:** Teddy Furon.

*In collaboration with B. Bhattarai, A. Mignon, F. Jurie, GREYC, Université of Caen.*

We investigated a new approach for de-identifying face images, *i.e.* for preventing automatic matching with public face collections. The overall motivation is to offer tools for privacy protection on social networks. We address this question by drawing a parallel between face de-identification and oracle attacks in digital watermarking. In our case, the identity of the face is seen as the watermark to be removed. Inspired by oracle attacks, we forge de-identified faces by superimposing a collection of carefully designed noise patterns onto the original face. The modification of the image is controlled to minimize the probability of good recognition while minimizing the distortion. In addition, these de-identified images are by construction made robust to counter attacks such as blurring. We present an experimental validation in which we de-identify LFW faces and show that resulting images are still recognized by human beings while deceiving a state-of-the-art face recognition algorithm [23].

#### **6.5.4. Tardos code in practice**

**Participant:** Teddy Furon.

*Joint work with the Inria supported start-up LAMARK.*

We deal with active fingerprinting a.k.a. traitor tracing where a collusion of dishonest users merges their individual versions of a content to yield a pirated copy. The Tardos codes are one of the most powerful tools to fight against such collusion process by identifying the colluders. Instead of studying as usual the necessary and sufficient code length in a theoretical setup, we adopt the point of view of the practitioner. We call this the *operational mode*, *i.e.* a practical setup where a Tardos code has already been deployed and a pirated copy has been found. This new paradigm shows that the known bounds on the probability of accusing an innocent in the theoretical setup are way too pessimistic. Indeed the practitioner can resort to much tighter bounds because the problem is fundamentally much simpler under the operational mode. In the end, we benchmark under the operational mode several single decoders recently proposed in the literature [32].

## MAGRIT Project-Team

# 6. New Results

## 6.1. Highlights of the Year

We were invited to present our work on *Impact of Soft Tissue Heterogeneity on Augmented Reality for Liver Surgery* in the TVCG Special Session at SIGGRAPH Asia 2014.

## 6.2. Matching and 3D tracking

### Pose initialization

Automating the camera pose initialization is still a problem in non instrumented environments. Difficulties originate in the possibly large viewpoint changes and lighting variations between the data stored in the model and the current view. One year ago, we began to investigate the use of viewpoint simulation techniques for re-localization within P. Rolin's PhD thesis. We especially consider challenging situations where the current view is distant from the image sequence used for model construction. We here consider scene models built from image sequence using Structure from Motion techniques. A point is then represented by its 3D coordinates and small image patches arising from the images where the point is detected. The underlying idea is to enrich 3D points by descriptors generated from virtual viewpoints chosen away from the learning sequence. For each 3D point of the model, a local image patch is generated from a set of virtual viewpoints, taking into account the local 3D normal and the images of the learning sequence. View synthesis is performed with an affine or an homography model. Though one possible shortcoming of simulation is to generate too many incorrect patches at discontinuities in the scene and thus to degrade the matching step, our preliminary results are very promising [25] and show a noticeable increase of the inlier ratio in the matching stage and an improved stability of the computed pose, especially when homography models are considered. We exhibit many examples where our method successfully computes the camera pose whereas the traditional methods fail.

Current investigations are about the development of scalable solutions for pose computation in large environments with several leverage actions in view. Designing efficient probabilistic techniques for matching and defining strategies based on the geometry of the scene for choosing a reduced set of virtual views are lines of research under investigation for jointly limiting the redundancy and improving the performance of the matching.

### Tracking 3D deformable objects

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 SHACRA 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 [2]. Specific models which best suit the considered organs, such as a vascularized model of the liver, have been introduced in this framework. During this year, we have first performed quantitative evaluation of the method [17]. Promising results were obtained through in-vivo experimentation on a human liver and ex-vivo validation on a porcine liver. In this latter case, artificial tumors were introduced in the liver, thus allowing a quantitative evaluation of the error between the predicted and the actual tumor. These experiments show that localization errors 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 [16], whereas previous works were guided by 3D features obtained with a stereo-endoscope. The only

parameter involved in the method is the Young's modulus but we show in experiments that a rough estimate of the Young's modulus is sufficient to obtain a good reconstruction. Experiments on computer-generated and real data have shown the effectiveness of the approach. The method is currently restricted to the orthographic projection and its extension to full projective geometry is under investigation.

A bio-mechanical model-based approach has also been considered in the context of tongue tracking in ultrasound images with a view to produce an augmented head for language learning. A crucial issue is the robustness of the tracking due to the strong speckle noise in ultrasound (US) data. Here, a small number of points are used to guide the model. Selection of feature points is based on the uncertainty associated to the tracked points and on spatial constraints. This model has proven to be especially efficient in the case of non uniform and fast movements [19].

#### **Use of AR in educational sciences**

In collaboration with the Ecole supérieure du professorat et de l'éducation and the PErSEUs laboratory at Université de Lorraine, we designed an inquiry-based AR learning environment (AIBLE) for teaching and learning astronomy in primary school (children of 8-11 years old). The novelty of this environment is the combination of Inquiry Based Sciences Education principles and didactics principles (here of astronomy) with AR capabilities. In this context, a GPL-licensed software called AIBLE-AstroAR has been developed based on the ARToolkit library. This software basically consists of a tangible user interface, which allows the children to move virtual celestial objects "as for real" and investigate in order to find origins of Moon phases evolution, alternation of day and night, seasons and Moon/Sun eclipses.

Last year, a study has been carried out to compare AIBLE with a physical model traditionally used in primary school. This study indicated that AIBLE significantly enhances learning compared to classical support. During this year, we performed further investigation with a larger panel of children to assess which characteristics of the environment facilitate learning [14]. Analyses of the marker positions as moved by the children indicated that AIBLE really facilitates heuristic investigation, which fosters consciousness of the origin of astronomical phenomena. This work provides new opportunities for teachers to identify solving problem strategies initiated by learners. These results also contribute to the understanding of the ways through which AR can be used in formal teaching curricula in K-12 schools.

### **6.3. Image-based modeling**

#### **Modeling vasculature for real time simulation**

One of our objectives to benefit interventional neuroradiology is to offer a patient-based interactive simulator to the interventional radiologists. Our contributions address vasculature modeling from patient data, namely 3D rotational angiography (3DRA) volumes. During Ahmed YUREIDINI's PhD thesis (2010-2014), a new model was developed consisting of a tree of local implicit blobby models.

We've been collaborating with SHACRA Inria project-team (Lille-Nord Europe) and the Department of Interventional Neuroradiology from Nancy University Hospital, in the context of the SOFA-InterMedS Inria Large-Scale Initiative. Ahmed YUREIDINI defended his PhD thesis in May this year with highest honors [9]. In particular, a detailed study was made to compare our tree of local implicits with triangular meshes in a view to model synthetic shapes as well as vasculatures from patient data. Increased performances with regard to processing speed, numerical stability and realism of the behavior were demonstrated.

#### **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. A solution in a two-view reconstruction context was proposed this year based on the extraction of the guide-wire as a skeleton in the images. The large stereo basis (views are almost orthogonal) and the segmentation errors (such as both missing parts and spurious segments in the skeleton) make the reconstruction especially difficult. The skeletons are subdivided in simple curves that are matched to build all corresponding potential 3D curves. These curves are nodes in a graph whose edge weights express a connection cost that takes into account both distance and orientation at the curves extremities. The solution 3D curve is provided by following the path of minimal cost in the graph. This algorithm demonstrated very good reconstruction results on synthetic and phantom data. A paper on this subject has been accepted for publication at SPIE Medical Imaging 2015.

#### **Patient-specific heart valve modeling**

Many pathologies damage heart valve anatomy producing undesired backflow, or regurgitation, decreasing cardiac efficiency and potentially leading to heart failure if left untreated. Such cases could be treated by surgical repair for the valve. However it is technically difficult and outcomes are highly dependent upon the experience of the surgeon: he must essentially predict the displacement and deformation of complex valve leaflets and supporting structures. One way to facilitate the repair is to simulate the mechanical behavior of the pathological valve with patient-specific data. This is the objective of Pierre-Frédéric Villard's one-year CNRS delegation in the Harvard Bio-robotics Laboratory (HBL). During the initial three first months of the sabbatical leave, various tasks have been performed: i) Study of the physiology of pathological valve behavior with medical experts. Following anatomical book reading and medical expert interviews the anatomy and the physiology are now understood. ii) Evaluation of HBL material for 4D ultrasound segmentation. HBL has previously developed a method to extract mitral valve geometry from a home-made high temporal resolution 3D ultrasound and iii) Automatic segmentation of a Mitral Valve microCT to feed a biomechanical model. A method to semi-automatically segment the leaflet-chordae set has been developed.

## **6.4. Parameter estimation**

#### **Metrologic performance assessment in experimental mechanics**

A problem of interest in experimental solid mechanics is strain map estimation on the surface of a specimen subjected to a load or a tensile test. One of the available approaches is based on images of a pseudo-periodic grid transferred on the surface of the specimen. Sensor noise is a major source of uncertainty in the strain map, and quantifying the propagation of the sensor noise to the measured strain components is a major problem when metrological performances are in view. We have proposed in [12] a study of the mathematical properties of the popular method based on windowed Fourier analysis, under a Gaussian white noise assumption. In the case of a more realistic signal-dependent, heteroscedastic noise, we have quantified in [10] (see also [15], [26]) the trade-off between the noise amplitude, the measurement resolution and the spatial resolution of the method. We have also investigated image stacking for noise reduction. While averaging a serie of images is certainly the most basic option to reduce the noise, it is not effective for studying grid images under a high magnification factor, because of unavoidable residual vibrations carried for instance by concrete floor slabs. We have shown in [13] that, while these vibrations indeed blur grid images, they still permit to reduce the noise amplitude in the displacement and strain maps.

#### **Sensor noise measurement.**

While searching for a low-cost on-the-fly estimation of the sensor parameters based on a serie of grid images (thus with no need of changing the experimental setting), we have proposed in [11] an algorithm which is able to deal with the vibrations biasing the estimations. More generally, we have investigated in [21] the problem of sensor parameter estimation from a series of images, under light flickering and vibrations. Light 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**

In the IDEaS ANR project we propose to target 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. This approach first calls for an improved knowledge of the statistical behavior of the simulation, which we addressed in the past year through experimental studies. We described our experimental setup in [20], which, in particular uses high speed stereo reconstruction to be able to study non quasi-static effects. Preliminary measures of the catheter speed during stick and slip transitions back up our conviction that quasi-static mechanical models fail to simulate such rapid motions of the tool. Our on-going analysis of the simulation sensitivity to mechanical parameters also sets forward friction as critical for high-fidelity simulation.



## MORPHEO Project-Team

# 6. New Results

## 6.1. Human Shape and Pose Tracking Using Keyframes

In this work we consider human tracking in multi-view set-ups and investigate a robust strategy that learns online key poses to drive a shape tracking method. The interest arises in realistic dynamic scenes where occlusions or segmentation errors occur. The corrupted observations present missing data and outliers that deteriorate tracking results. We propose to use key poses of the tracked person as multiple reference models. In contrast to many existing approaches that rely on a single reference model, multiple templates represent a larger variability of human poses. They provide therefore better initial hypotheses when tracking with noisy data. Our approach identifies these reference models online as distinctive keyframes during tracking. The most suitable one is then chosen as the reference at each frame. In addition, taking advantage of the proximity between successive frames, an efficient outlier handling technique is proposed to prevent from associating the model to irrelevant outliers. The two strategies are successfully experimented with a surface deformation framework that recovers both the pose and the shape. Evaluations on existing datasets also demonstrate their benefits with respect to the state of the art. This work was presented at CVPR'14 [5].

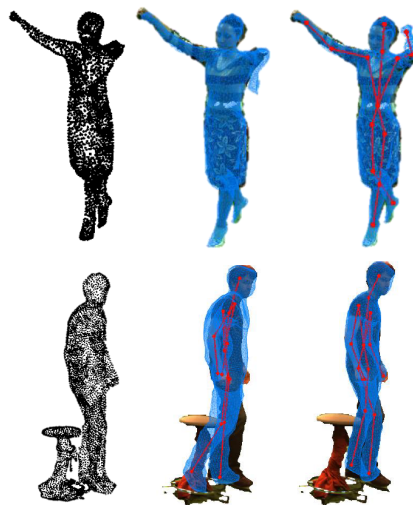


Figure 4. Shape tracking with keyframes[5]

## 6.2. On Mean Pose and Variability of 3D Deformable Models

We present a novel methodology for the analysis of complex object shapes in motion observed by multiple video cameras. In particular, we propose to learn local surface rigidity probabilities (i.e., deformations), and to estimate a mean pose over a temporal sequence. Local deformations can be used for rigidity-based dynamic surface segmentation, while a mean pose can be used as a sequence keyframe or a cluster prototype and has therefore numerous applications, such as motion synthesis or sequential alignment for compression or morphing. We take advantage of recent advances in surface tracking techniques to formulate a generative model of 3D temporal sequences using a probabilistic framework, which conditions shape fitting over all

frames to a simple set of intrinsic surface rigidity properties. Surface tracking and rigidity variable estimation can then be formulated as an Expectation-Maximization inference problem and solved by alternatively minimizing two nested fixed point iterations. We show that this framework provides a new fundamental building block for various applications of shape analysis, and achieves comparable tracking performance to state of the art surface tracking techniques on real datasets, even compared to approaches using strong kinematic priors such as rigid skeletons.



Figure 5. Rigidity probability of the shape tracked with mean pose [4]

### 6.3. Segmentation multi-vues par coupure de graphes

In this paper, we address the problem of object segmentation in multiple views when two or more viewpoints of the same scene are available. We propose a new approach that propagates segmentation coherence information in space, hence allowing evidence in one image to be shared over the complete set. To this aim the segmentation is cast as a single efficient labeling problem over space and time with graph cuts. In contrast to most existing multi-view segmentation methods that rely on some form of dense reconstruction, ours only requires a sparse 3D sampling to propagate information between viewpoints. The approach is thoroughly evaluated on standard multi-view datasets. The obtained results compete with state of the art methods but they are achieved with significantly fewer viewpoints.

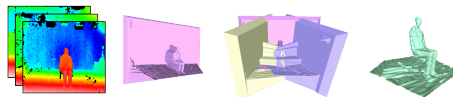


Figure 6. Results of our multi-view segmentation approach over 3 input views, with no user interaction (completely automated). [9]

### 6.4. Combined Visible and X-Ray 3D Imaging

This work considers 3D imaging of moving objects and introduces a technique that exploits visible and x-ray images to recover dense 3D models. While recent methods such as tomography from cone-beam x-ray can advantageously replace more expensive and higher-dose CT scanners, they still require specific equipment and immobilised patients. We investigate an alternative strategy that combines a single x-ray source and a set of colour cameras to capture rigidly moving samples. The colour cameras allow for coarse markerless motion tracking, which is further refined with the x-ray information. Once the sample poses are correctly estimated, a dense 3D attenuation model is reconstructed from the set of x-ray frames. Preliminary results on simulated data compared to ground-truth as well as actual in-vivo experiments were presented at the conference MIUA'14 [6].

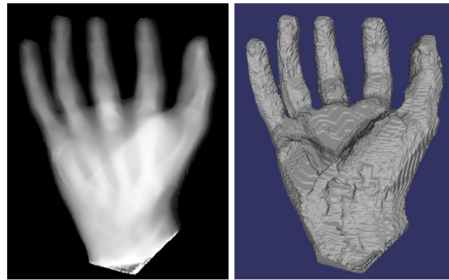


Figure 7. Human hand reconstruction with a combined visible and x-ray system [6]

## 6.5. Non-Rigid Registration meets Surface Reconstruction

Non rigid registration is an important task in computer vision with many applications in shape and motion modeling. A fundamental step of the registration is the data association between the source and the target sets. Such association proves difficult in practice, due to the discrete nature of the information and its corruption by various types of noise, e.g. outliers and missing data. In this work we investigate the benefit of the implicit representations for the non-rigid registration of 3D point clouds. First, the target points are described with small quadratic patches that are blended through partition of unity weighting. Then, the discrete association between the source and the target can be replaced by a continuous distance field induced by the interface. By combining this distance field with a proper deformation term, the registration energy can be expressed in a linear least square form that is easy and fast to solve. This significantly eases the registration by avoiding direct association between points. Moreover, a hierarchical approach can be easily implemented by employing coarse-to-fine representations. Experimental results were conducted with point clouds from multi-view data sets. The qualitative and quantitative comparisons show the outperformance and robustness of our framework. This work was presented at 3DV'14[7].

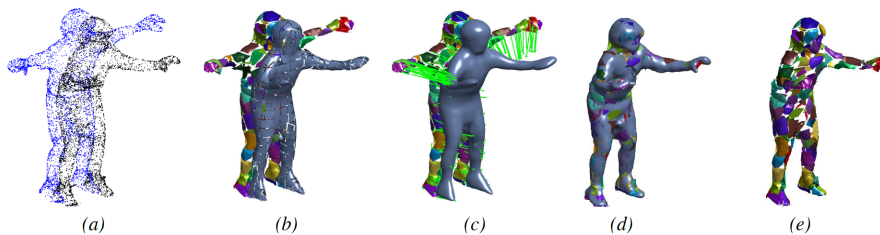


Figure 8. Using implicit interface for registration: (a) initial pose of the source and target sets; (b) source patches and the local quadrics representing the target; (c) the implicit interface induces a gradient field; (d) deformed source patches fitting the interface; a coarse-to-fine interface has been used in (c) - (d); (e) the final deformation of the template [7].

## 6.6. High Resolution 3D Shape Texture from Multiple Videos

We examine the problem of retrieving high resolution textures of objects observed in multiple videos under small object deformations. In the monocular case, the data redundancy necessary to reconstruct a high-

resolution image stems from temporal accumulation. This has been vastly explored and is known as super-resolution. On the other hand, a handful of methods have considered the texture of a static 3D object observed from several cameras, where the data redundancy is obtained through the different viewpoints. We introduce a unified framework to leverage both possibilities for the estimation of a high resolution texture of an object. This framework uniformly deals with any related geometric variability introduced by the acquisition chain or by the evolution over time. To this goal we use 2D warps for all viewpoints and all temporal frames and a linear projection model from texture to image space. Despite its simplicity, the method is able to successfully handle different views over space and time. As shown experimentally, it demonstrates the interest of temporal information that improves the texture quality. Additionally, we also show that our method outperforms state of the art multi-view super-resolution methods that exist for the static case. This work was presented at CPVR'14 [8].

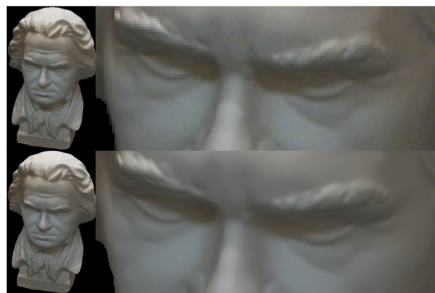


Figure 9. Input view  $768 \times 576$  resolution with up-sampling by factor of three, BEETHOVEN dataset. Super-resolved  $2304 \times 1728$  output of our algorithm rendered from identical viewpoint [8].

## PERCEPTION Project-Team

### 5. New Results

#### 5.1. Highlights of the Year

- In 2014 Antoine Deleforge (team member 2009-2013) received the **Signal, Image and Vision best PhD prize** for his thesis “*Acoustic Space Mapping: A Machine Learning Approach to Sound Source Separation and Localization*”, defended in December 2013 and advised by Radu Horaud. The prize is jointly awarded by GDR ISIS, Club EEA, and GRETSI.

Website: <http://www.inria.fr/centre/grenoble/actualites/apprendre-a-rester-attentif-a-ses-locuteurs>

- Radu Horaud was awarded an **ERC Advanced Grant** for his five year project VHIA “*Vision and Hearing in Action*”, grant number 340113, 2014-2019.

Website: <https://team.inria.fr/perception/vhia/>.

- The PERCEPTION team was awarded an **ANR BLANC** two year project MIXCAM “*Real-Time Visual Reconstruction by Mixing Multiple Depth and Color Cameras*”, in collaboration with 4D View Solutions, 2014-2016.

Website: <https://team.inria.fr/perception/mixcam-project/>

- The PERCEPTION team was awarded an **FP7 STREP** three year project EARS “*Embodied Audition for Robots*”, in collaboration with Friedrich Alexander Universiteit, coordinator (Germany), Ben Gurion University (Israel), Imperial College (UK), Humboldt University Berlin (Germany) and Aldebaran Robotics (France), 2014-2017.

Website: <https://team.inria.fr/perception/ears/>

#### 5.2. Acoustic Space Learning on Binaural Manifolds

We addressed the problems of modeling the acoustic space generated by a full-spectrum sound source and of using the learned model for the localization and separation of multiple sources that simultaneously emit sparse-spectrum sounds. We lay theoretical and methodological grounds in order to introduce the binaural manifold paradigm. We perform an in-depth study of the latent low-dimensional structure of the high-dimensional interaural spectral data, based on a corpus recorded with a human-like audiomotor robot head, namely the POPEYE robot shown on Fig 3 (right). A non-linear dimensionality reduction technique is used to show that these data lie on a two-dimensional (2D) smooth manifold parameterized by the motor states of the listener, or equivalently, the sound source directions, e.g., Fig. 4. We propose a probabilistic piecewise affine mapping model (PPAM) specifically designed to deal with high-dimensional data exhibiting an intrinsic piecewise linear structure. We derive a closed-form expectation-maximization (EM) procedure for estimating the model parameters, followed by Bayes inversion for obtaining the full posterior density function of a sound source direction. We extend this solution to deal with missing data and redundancy in real world spectrograms, and hence for 2D localization of natural sound sources such as speech. We further generalize the model to the challenging case of multiple sound sources and we propose a variational EM framework. The associated algorithm, referred to as variational EM for source separation and localization (VESSL) yields a Bayesian estimation of the 2D locations and time-frequency masks of all the sources. Comparisons of the proposed approach with several existing methods reveal that the combination of acoustic-space learning with Bayesian inference enables our method to outperform state-of-the-art methods [18], [24].

Website: <https://team.inria.fr/perception/research/acoustic-learning/>

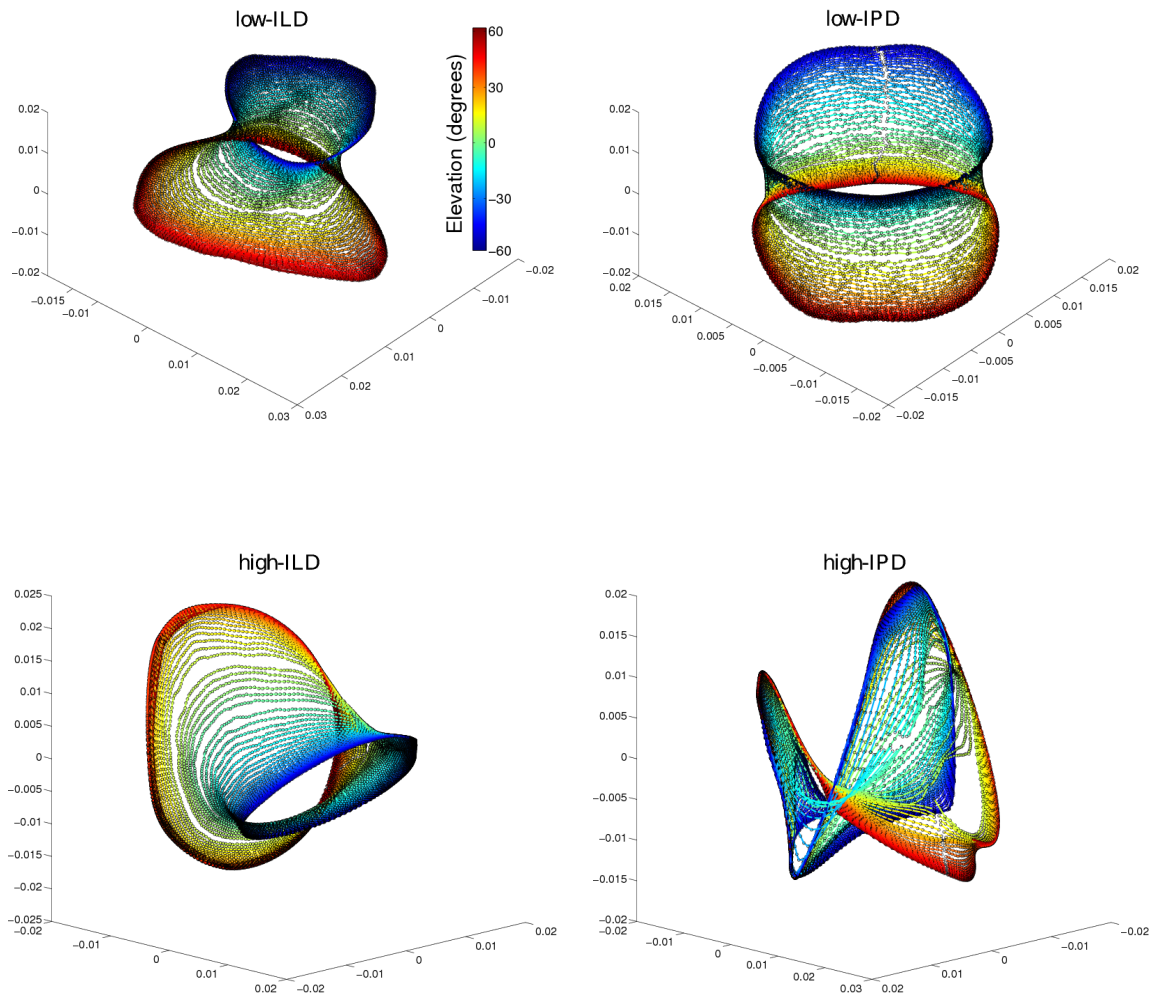


Figure 4. This figure illustrates the concept of binaural manifold. A wide-spectrum sound is recorded with a binaural acoustic dummy head and an interaural high-dimensional spectral representation of this sound is mapped onto a low-dimensional (2) space. This reveals the two-dimensional manifold of possible sound-source directions that is embedded in the interaural spectral features. Please consult [18] for more details.



### 5.3. Geometric Sound Source Localization

We address the problem of sound-source localization from time-delay estimates using arbitrarily-shaped non-coplanar microphone arrays. A novel geometric formulation is proposed, together with a thorough algebraic analysis and a global optimization solver [15]. The proposed model is thoroughly described and evaluated. The geometric analysis, stemming from the direct acoustic propagation model, leads to necessary and sufficient conditions for a set of time delays to correspond to a unique position in the source space. Such sets of time delays are referred to as *feasible sets*. We formally prove that every feasible set corresponds to exactly one position in the source space, whose value can be recovered using a closed-form localization mapping. Therefore we seek for the optimal feasible set of time delays given, as input, the received microphone signals. This time delay estimation problem is naturally cast into a programming task, constrained by the feasibility conditions derived from the geometric analysis. A global branch-and-bound optimization technique is proposed to solve the problem at hand, hence estimating the best set of feasible time delays and, subsequently, localizing the sound source. Extensive experiments with both simulated and real data are reported; we compare our methodology to four state-of-the-art techniques. This comparison shows that the proposed method combined with the branch-and-bound algorithm outperforms existing methods. These in-depth geometric understanding, practical algorithms, and encouraging results, open several opportunities for future work.

Website: <https://team.inria.fr/perception/research/geometric-sound-source-localization/>

### 5.4. Joint Registration of Multiple Point Sets

We developed a probabilistic generative model and its associated algorithm to jointly register multiple point sets. The vast majority of state-of-the-art registration techniques select one of the sets as the *model* and perform pairwise alignments between the other sets and this set. The main drawback of this mode of operation is that there is no guarantee that the model-set is free of noise and outliers, which contaminates the estimation of the registration parameters. Unlike previous work, the proposed method treats all the point sets on an equal footing: they are realizations of a Gaussian mixture (GMM) and the registration is cast into a clustering problem [26]. We formally derive an EM algorithm that estimates both the GMM parameters and the rotations and translations that map each individual set onto the *central* model. The mixture means play the role of the registered set of points while the variances provide rich information about the quality of the registration. We thoroughly validate the proposed method with challenging datasets, we compare it with several state-of-the-art methods, and we show its potential for fusing real depth data.

Website: <https://team.inria.fr/perception/research/jrmpc/>

### 5.5. High-Dimensional Regression

The problem of approximating high-dimensional data with a low-dimensional representation is addressed. The article makes the following contributions. An inverse regression framework is proposed, which exchanges the roles of input and response, such that the low-dimensional variable becomes the regressor, and which is tractable. A mixture of locally-linear probabilistic mapping model is introduced, that starts with estimating the parameters of the inverse regression, and follows with inferring closed-form solutions for the forward parameters of the high-dimensional regression problem of interest. Moreover, a partially-latent paradigm is introduced, such that the vector-valued response variable is composed of both observed and latent entries, thus being able to deal with data contaminated by experimental artifacts that cannot be explained with noise models. The proposed probabilistic formulation could be viewed as a latent-variable augmentation of regression. Expectation-maximization (EM) procedures are introduced, based on a data augmentation strategy which facilitates the maximum-likelihood search over the model parameters. Two augmentation schemes are proposed and the associated EM inference procedures are described in detail; they may well be viewed as generalizations of a number of EM regression, dimension reduction, and factor analysis algorithms. The proposed framework is validated with both synthetic and real data. Experimental evidence is provided that the method outperforms several existing regression techniques [19], [25].

Website: <https://team.inria.fr/perception/research/high-dim-regression/>

## 5.6. Audiovisual Speaker Detection, Localization and Interaction with NAO

In this research we address the problem of audio-visual speaker detection. We introduce an online system working on the humanoid robot NAO. The scene is perceived with two cameras and two microphones. A *multimodal* Gaussian mixture model (GMM) fuses the information extracted from the auditory and visual sensors. The system is implemented based on a platform-independent middleware library and it is able to process the information online (17 visual frames per second). A detailed method description and the system implementation are provided, with special emphasis on the online processing issues that must be addressed, and the proposed solutions. Experimental validation is done over five different scenarios, with no special lighting, nor special acoustic conditions, leading to good results [16].

Website: <https://team.inria.fr/perception/research/audiovisual-nao/>

## 5.7. EM for Weighted-Data Clustering



Figure 5. We developed a novel multimodal clustering method that is based on expectation-maximization (EM) with weighted data. The left image shows auditory features (green), namely sound source positions mapped onto the image plane using [24] and visual features (blue, lip landmarks), as well as the active speaker (yellow square). The right image shows the results of our weighted-data EM algorithm that finds three clusters. Among these clusters, the active audio-visual cluster is marked with a transparent blue circle.

Data clustering has received a lot of attention and many methods, algorithms and software packages are currently available. Among these techniques, parametric finite-mixture models play a central role due to their interesting mathematical properties and to the existence of maximum-likelihood estimators based on expectation-maximization (EM). In this paper we propose a new mixture model that associates a weight with each observed data point. We introduce a Gaussian mixture with weighted data and we derive two EM algorithms [29]: the first one considers the weight of each observed datum to be fixed, while the second one treats each weight as a hidden variable drawn from a gamma distribution. We provide a general-purpose scheme for weight initialization and we thoroughly validate the proposed algorithms by comparing them with several parametric and non-parametric clustering techniques. We demonstrate the utility of our method for clustering heterogeneous data, namely data gathered with different sensorial modalities, e.g., audio and vision.

Website: <https://team.inria.fr/perception/research/wdgmm/>

## 5.8. Continuous Action Recognition

Continuous action recognition is more challenging than isolated recognition because classification and segmentation must be simultaneously carried out. We build on the well known dynamic time warping (DTW) framework and devise a novel visual alignment technique, namely dynamic frame warping (DFW), which performs isolated recognition based on per-frame representation of videos, and on aligning a test sequence with a model sequence. Moreover, we propose two extensions which enable to perform recognition concomitant with segmentation, namely one-pass DFW and two-pass DFW. These two methods have their roots in the domain of continuous recognition of speech and, to the best of our knowledge, their extension to continuous visual action recognition has been overlooked. We test and illustrate the proposed techniques with a recently released dataset (RAVEL) [32] and with two public-domain datasets widely used in action recognition (Hollywood-1 and Hollywood-2). We also compare the performances of the proposed isolated and continuous recognition algorithms with several recently published methods [22].

Website: <https://team.inria.fr/perception/research/car/>

## 5.9. Skeletal Quads

Recent advances on human motion analysis have made the extraction of human skeleton structure feasible, even from single depth images. This structure has been proven quite informative for discriminating actions in a recognition scenario. In this context, we propose a local skeleton descriptor that encodes the relative position of joint quadruples. Such a coding implies a similarity normalization transform that leads to a compact (6D or 5D) view-invariant skeletal feature, referred to as skeletal quad. In the references below, we use this descriptor in conjunction with Fisher kernel in order to encode gesture or action (sub)sequences. The short length of the descriptor compensates for the large inherent dimensionality associated to Fisher vectors. We investigate the performance in both isolated [28] and continuous [27] recognition scenarios.

Website: <https://team.inria.fr/perception/research/skeletalquads/>

## PRIMA Project-Team

### 5. New Results

#### 5.1. Highlights of the Year

On March 14, 2014, James Crowley was named Chevalier de l'Ordre national du Mérite.

On August 2014, the paper "Human-Robot Motion: an Attention-Based Navigation Approach" [14] by Thierry Fraichard, Remi Paulin & Patrick Reignier has been nominated for the best paper award at the IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN 2014), Edinburgh (UK).

On December 2014, Patrick Reignier was a member of the EDF grand jury for smart energy

BEST PAPER AWARD :

[14] **IEEE Int. Symp. on Robot and Human Interactive Communication (ROMAN)**. T. FRAICHARD, R. PAULIN, P. REIGNIER.

#### 5.2. Attention-Based Navigation

**Participants:** Thierry Fraichard, Remi Paulin, Patrick Reignier..

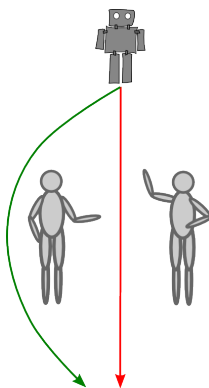


Figure 6. People are not pieces of furniture! Motion in red is definitely shorter but it is not appropriate.

The domain of service-robots is growing fast and has become the focus of many researchers and industrials alike. Their application areas been 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 minimising resources such as time or distance travelled anymore, the robot motion must abide by social/cultural rules and be **appropriate**, e.g. Fig. 6 .

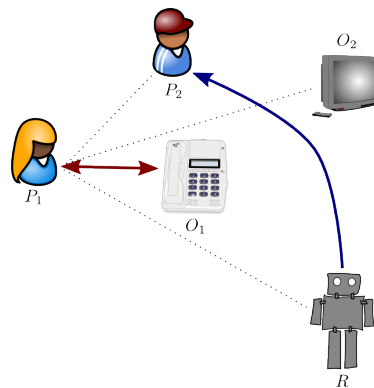


Figure 7. Attention vs activity: although  $P_1$ 's current activity is being on the phone, part of her attention may be directed towards  $P_2$ , the TV set or the robot  $R$ . Suppose now that  $R$  moves towards  $P_2$  in a way that hides the TV from  $P_1$ . Such a behavior would not be appropriate should  $P_1$  be actually paying attention to the TV.

Most of the approaches proposed so far relies 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” [42], or the activity it is currently engaged in, *e.g.* “Interaction Space” [47] and “Activity Space” [51]. 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. Such approaches are interesting however their spatial nature (being inside or outside the space) make them less suitable when facing more complicated situations, *e.g.* Fig. 7. To overcome those limits, we suggest using the psychological concept of **attention**, which plays a central role when humans navigate around each other. Besides lifting the limits of social spaces, 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 enable 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 [53], 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 mapping from the state space of the robot to  $\mathbb{R}$ . The attention field can serve as an attention predictor that can be used to predict potential attentional situations. This knowledge can in turn be used to decide what the robot should do in the future depending on its current task.

Let us illustrate this on a simple scenario featuring a person, a TV and a robot (denoted by  $P_1$ ,  $O_1$  and  $R$  in Fig. 8 -left). The person is currently watching the TV: this is his current *activity*. This activity relates to his *intention* and is modeled by the yellow vector  $\vec{I}$  in Fig. 8 -left that is directed from the person to the TV. Fig. 8 -right depicts the attention field for the person; it is a mapping from  $\mathbb{R}^2$  to  $\mathbb{R}$  that gives the amount of attention that the person is paying to the robot when it is at a given position  $(x, y)$ . Fig. 8 -right should be interpreted as follows: the warmer the color, the higher the amount of attention given by the person to the robot. It integrates both the visual and auditory perception capabilities of  $P_1$ .

In 2014, we have furthered the development of the concept of attention field and demonstrated different ways to use its attention prediction capability on various scenarios. The main results obtained have been reported in a conference article that has been nominated for the Best Paper Award [14]. Work is ongoing to quantify

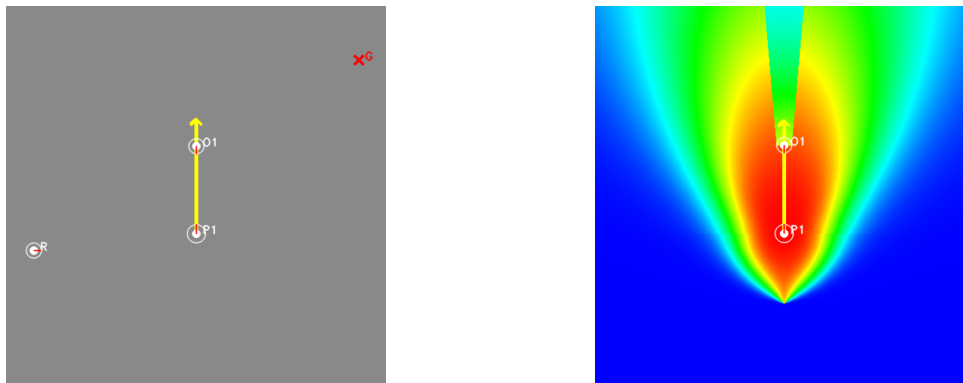


Figure 8. Person-TV-Robot scenario (left), Attention field for the person  $P_1$  (right).

the social “goodness” of the paths provided by our approach, to further the use of the concept of attention on more challenging and dynamic scenarios and to offer an approach to fill the gap between appropriate motion and interaction through motion.

### 5.3. 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.

## **5.4. 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.

## SIROCCO Project-Team

# 6. New Results

## 6.1. 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

### 6.1.1. *Salient object detection*

**Participants:** Olivier Le Meur, Zhi Liu.

Salient object detection consists in extracting in an automatic manner the most interesting object in an image or video sequence. From an input image, an object, with well-defined boundaries, is detected based on its saliency. This subject knows an renewed interest these last years. A number of datasets serving as ground truth has been released and can be used to benchmark methods.

In 2013, we proposed a new method for detecting salient objects in still color images. In 2014, this method has been extended to video sequences [21]. Based on the superpixel representation of video frames, motion histograms and color histograms are computed at local and global levels. From these histograms, a superpixel-level temporal saliency measure as well as a spatial saliency measure are obtained. Finally, a pixel-level saliency derivation method is proposed to generate pixel-level temporal saliency map and spatial saliency map. An adaptive fusion method allows to integrate them into an unique spatiotemporal saliency map. Experimental results on two public datasets demonstrate that the proposed model outperforms state-of-the-art spatiotemporal saliency model in terms of both saliency detection and human fixation prediction.

### 6.1.2. *Saliency aggregation*

**Participants:** Olivier Le Meur, Zhi Liu.

In this study [32], we investigate whether the aggregation of saliency maps allows to outperform the best saliency models. Today there exist a number of saliency models for predicting the most visually salient locations within a scene. Although all existing models follow the same objective, they provide results which could be, to some extent, different. The discrepancies are related to the quality of the prediction but also to the saliency map representation. Indeed some models output very focused saliency maps whereas the distribution of saliency values is much more uniform in other models. Others tend to emphasize more on the image edges, the color or luminance contrast. This saliency map manifold contains a rich resource that should be used and from which new saliency maps could be inferred. Combining saliency maps generated using different models might enhance the prediction quality and the robustness of the prediction. Our goal is then to take saliency maps from this manifold and to produce the final saliency map.

This study discussed various aggregation methods; six unsupervised and four supervised learning methods are tested on two existing eye fixation datasets. Results show that a simple average of the TOP 2 saliency maps significantly outperforms the best saliency models. Considering more saliency models tends to decrease the performance, even when robust aggregation methods are used. Concerning the supervised learning methods, we provide evidence that it is possible to further increase the performance, under the condition that an image similar to the input image can be found in the training dataset. Our results might have an impact for critical applications which require robust and relevant saliency maps.

### 6.1.3. *Models for 3D video quality assessment*

**Participants:** Darya Khaustova, Olivier Le Meur.

This work is carried out in collaboration with Orange labs. The goal is to design objective metrics for quality assessment of 3D video content, by establishing links between human visual perception (visual comfort) and video parameters such as quality and depth quantity, and between visual comfort and visual attention. In 2013 we investigated the differences in 2D visual attention in comparison with 3D visual attention [31]. In 2014, we have focused on the design of an objective stereoscopic quality metric. In stereoscopic video quality, the assessment of spatial and temporal distortions by conventional quality metrics became incomplete because of the added depth dimension. Improperly captured or rendered, depth information can induce visual discomfort, impacting the overall video 3D QoE quality independently of image quality. The model is based on perceptual thresholds, namely visual annoyance, and acceptability. The visual annoyance threshold defines the boundary between annoying and not annoying sensation: 50% of subjects consider a stimulus annoying and 50% as not annoying. Acceptability determines the viewer's expectation level for the perceived video quality in a certain context and situation (inspired by the acceptability for the customer defined as an adequate service).

In order to compute the quality score, the proposed metric requires in input the distortion level of a technical and particular parameter, annoyance threshold and acceptability threshold of the targeted parameter. The performance of proposed objective mode is evaluated by considering five view asymmetries with five degradation levels. Generated contents were assessed by 30 subjects for each asymmetry (focal length mismatch, vertical shift, and rotation, green and white level reduction). The results of the subjective test have demonstrated that it is possible to classify detected problem to one of the objective categories using corresponding acceptability and visual annoyance thresholds.

#### 6.1.4. Epitome-based video representation

**Participants:** Martin Alain, Christine Guillemot.

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. Experiments show that interesting complexity results can be obtained without degrading the epitome quality for both proposed methods. By limiting the number of exhaustive searches we limit the memory occupation and the processing time, while keeping a good epitome quality (down to 18.08 % of the original memory occupation and 41.39 % of the original processing time) [25]. As an example, images reconstructed using the different techniques are visible in Fig. 1. The epitome construction method is currently being extended from still images to groups of images in video sequences. Denoising and super-resolution algorithms based on the constructed epitomes are also under study.

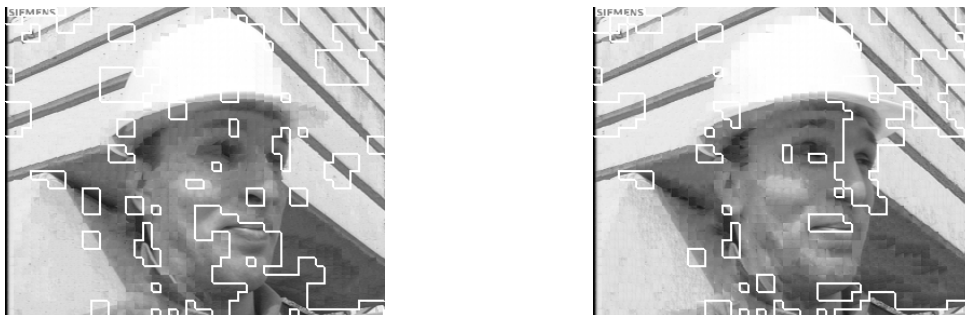


Figure 1. Reconstructed images using the list-based (left) and clustering-based methods. Epitome patches are highlighted in white.

### 6.1.5. Light field tomographic reconstruction from a fixed camera focal stack

**Participants:** Christine Guillemot, Elif Vural.

Thanks to the internship of Antoine Mousnier (student at Ecole Centrale Lyon), we have developed a novel approach to partially reconstruct high-resolution 4D light fields from a stack of differently focused photographs taken with a fixed camera. First, a focus map is calculated from this stack using a simple approach combining gradient detection and region expansion with graph cut. Then, this focus map is converted into a depth map thanks to the calibration of the camera. We proceed after this with the tomographic reconstruction of the epipolar images by back-projecting the focused regions of the scene only. We call it masked back-projection. The angles of back-projection are calculated from the depth map. Thanks to the high angular resolution we achieve, we are able to render puzzling perspective shifts although the original photographs were taken from a single fixed camera at a fixed position and render images with extended focus (see Fig. 2). To the best of our knowledge, our method is the first one to reconstruct a light field by using a focal stack captured with an ordinary camera at a fixed viewpoint.

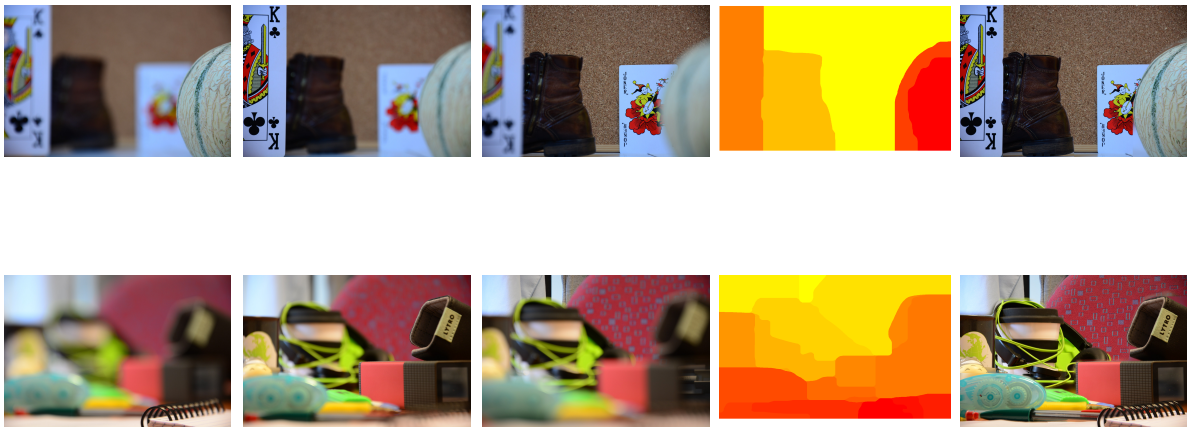


Figure 2. Three images of the focal stack (left); estimated depth map and image with extended focus (right). The focal stack images of the first and second rows have been captured with a Nikon 5200 camera.

## 6.2. Rendering, inpainting and super-resolution

image-based rendering, inpainting, view synthesis, super-resolution

### 6.2.1. Video inpainting

**Participants:** Mounira Ebdelli, Christine Guillemot, Olivier Le Meur.

Image (and video) inpainting refers to the process of restoring missing or damaged areas in an image (or a video). This field of research has been very active over the past years, boosted by numerous applications: restoring images from scratches or text overlays, loss concealment in a context of impaired image transmission, object removal in a context of editing, disocclusion in image-based rendering of viewpoints different from those captured by the cameras. Inpainting is an ill-posed inverse problem: given observations, or known samples in a spatial (or spatio-temporal) neighborhood, the goal is to estimate unknown samples of the region to be filled in. Many methods already exist for image inpainting, either based on PDE (Partial Derivative Equation)-based diffusion schemes, either using sparse or low rank priors or following texture synthesis principles exploiting statistical or self-similarity priors.

In 2014, the problem of video inpainting has been further addressed with free-moving cameras. The algorithm developed first compensates the camera motion between the current frame and its neighboring frames in a sliding window, using a new region-based homography computation which better respects the geometry of the scene compared to state-of-the-art methods. The source frame is first segmented into regions in order to find homogeneous regions. Then, the homography for mapping each region into the target frame is estimated. The overlapping of all aligned regions forms the registration of the source frame into the target one. Once the neighboring frames have been aligned, they form a stack of images from which the best candidate pixels are searched in order to replace the missing ones. The best candidate pixel is found by minimizing a cost function which combines two energy terms. One energy term, called the data term, captures how stationary is the background information after registration, hence enforcing temporal coherency. The second term aims at favoring spatial consistency and preventing incoherent seams, by computing the energy of the difference between each candidate pixel and its 4-neighboring pixels in the missing region. The minimization of the energy term is performed globally using Markov Random Fields and graph cuts. A method of Poisson blending has been implemented in order to further enhance the visual quality of the inpainted videos. The proposed approach, although less complex than state-of-the-art methods, provides more natural results.

### 6.2.2. Image and video super-resolution in the example-based framework

**Participants:** Marco Bevilacqua, Christine Guillemot, Aline Roumy.

Super-resolution (SR) refers to the problem of creating a high-resolution (HR) image, given one or multiple low-resolution (LR) images as input. The SR process aims at adding to the LR input(s) new plausible high frequency details, to a greater extent than traditional interpolation methods. We mostly focused on the single-image problem, where only a single LR image is available. We have adopted the example-based framework on one hand and the sparse approximation framework on the other hand.

In the example-based framework, the relation between the LR and HR image spaces is modeled with the help of pairs of small “examples”, i.e. texture patches. Each example pair consists of a LR patch and its HR version that also includes high-frequency details; the pairs of patches form a dictionary of patches. For each patch of the LR input image, one or several similar patches are found in the dictionary, by performing a nearest neighbor search. The corresponding HR patches in the dictionary are then combined to form a HR output patch; and finally all the reconstructed HR patches are re-assembled to build the superresolved image. In this procedure, one important aspect is how the dictionary of patches is built. At this regard, two choices are possible: an external dictionary, formed by sampling HR and LR patches from external training images; and an internal dictionary, where the LR/HR patch correspondences are learned by putting in relation directly the input image and scaled versions of it. The advantage of having an external dictionary is that it is built in advance: this leads to a reduction of the computational time, whereas in the internal case the dictionary is generated online at each run of the algorithm. However, external dictionaries have a considerable drawback: they are fixed and so non-adapted to the input image. To be able to satisfactorily process any input image, we need then to include in the dictionary a large variety of patch correspondences, leading to a high computational time. In 2013, external dictionaries have been designed to bridge the gap between external and internal dictionary based methods.

In 2014 instead, we proposed a novel SR method for internal dictionaries [16]. The internal dictionary contains pair of LR/HR patches taken from the image to be processed and is by construction well adapted to the data. However, its size is limited since it results from the sampling of a single image. This leads to an undersampling of the LR space and even more of the HR space. To overcome this problem, state of the art methods select, for each input LR patch, a local neighborhood, learn the local geometry of this neighborhood, and apply it in the HR domain. Therefore, an underlying hypothesis is that the local neighborhoods in the LR and HR domain are similar. To avoid this hypothesis, we employ a regression-based method to directly map LR input patches into their related HR output patches. To make this regression more robust, first the LR patches have been first oversampled (by a bicubic interpolation) such that LR and HR spaces have the same dimension, and second a Tikhonov regularization has been added. When compared to other state-of-the-art algorithms, our proposed algorithm shows the best performance, both in terms of objective metrics and subjective visual results. As for the former, it presents considerable gains in PSNR and SSIM values. When observing the super-resolved images, also, it turns out to be the most capable in producing fine artifact-free HR details.



### 6.2.3. Image super-resolution in a sparse and manifold learning framework

**Participants:** Julio Cesar Ferreira, Christine Guillemot, Olivier Le Meur, Elif Vural.

The problem of image super-resolution has also been addressed in a sparse approximation framework. This led to a novel algorithm based on sparse representations in which a structure tensor-based regularization has been introduced [29]. The relative discrepancy between the two eigenvalues of the structure tensor is an indicator of the degree of anisotropy of the gradient in a region of the image. The eigenvalues and eigenvectors of the structure tensor are used to compute, for each pixel belonging to a salient edge, a stream line in the direction perpendicular to the edge (given by the eigenvector corresponding to the highest eigenvalue of the structure tensor). The saliency of an edge is given by the S-norm of the highest eigenvalue. An energy term dealing with the sharpness of edges is then computed and used as a regularization constraint to modify the current estimated high resolution image inside the Iterative Shrinkage Thresholding algorithm. This extra constraint forces the value of the current pixel along the stream line to be as close as possible to pixel values having lowest saliency. The resulting single-image algorithm, called Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS) allows sharpening edges and reducing the ringing artefacts compared to existing methods. This is illustrated in Fig.3



Figure 3. Comparison of SR results ( $\times 3$ ). (a) LR image; (b) Nearest-neighbor; (c) Sparse method without structure-based regularization; (d) SE-ASDS results. (e) Comparison between (c) and (d) on patches: edges of (d) are more contrasted than (c).

In the previous method, the dictionaries used for the sparse approximation method are defined as a union of PCA basis learned on clusters of patches of the input image. The clusters are constructed using the classical k-means algorithm with patch distances computed with the Euclidean distance. This study is being pursued by assuming manifold models for the patches of the input images. A method using graph-based clustering has then been used for clustering patches on the manifold, and this method has been extended to cope with the out-of-sample problem. Dedicated dictionary learning methods are currently under development to have dictionaries best adapted to the manifold structure.

## 6.3. Representation and compression of large volumes of visual data



Sparse representations, data dimensionality reduction, compression, scalability, perceptual coding, rate-distortion theory

### 6.3.1. *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 thus do a theoretical analysis of separability of data representations given by supervised manifold learning. In particular, we focus on the nonlinear supervised extensions of the Laplacian eigenmaps algorithm and examine the linear separation between different classes in the learned embedding. We first consider a setting with two classes and show that the two classes become linearly separable even with a one-dimensional embedding. We characterize the linear separation in terms of the data graph properties such as edge weights, diameter, and volume and some algorithm parameters. We then extend these results to a setting with multiple classes, where the classes are assumed to be categorizable into a few groups with high intra-group affinities. We show 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 are also confirmed by experimentation on synthetic data sets and image data.

Next, we consider 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 are designed for unsupervised methods. We study the problem for the particular application of data classification and propose an algorithm to compute a continuous function from the original data space to the low-dimensional space of embedding. In particular, we construct 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 show that the proposed method gives promising results in the classification of low-dimensional image data such as face images.

### 6.3.2. *Dictionary learning for sparse coding and classification of satellite images*

**Participants:** Jeremy Aghaei Mazaheri, Christine Guillemot, Claude Labit.

In the context of the national partnership Inria-Astrium, we explore novel methods to encode images captured by a geostationary satellite. These pictures have to be compressed on-board before being sent to earth. Each picture has a high resolution and so the rate without compression is very high (about 70 Gbits/sec) and the goal is to achieve a rate after compression of 600 Mbits/sec, that is a compression ratio higher than 100. On earth, the pictures are decompressed with a high reconstruction quality and visualized by photo-interpreters. The goal of the study is to design novel transforms based on sparse representations and learned dictionaries for satellite images.

We have developed methods for learning adaptive tree-structured dictionaries. Each dictionary in the structure is learned on a subset of residuals from the previous level, with the K-SVD algorithm. The tree structure offers better rate-distortion performance than a "flat" dictionary learned with K-SVD, especially when only a few atoms are selected among the first levels of the tree. The tree-structured dictionary allows efficient coding of the indices of the selected atoms. Besides coding, these structured dictionaries turn out to be useful tools for

MTF (Modulation Transfer Function) estimation and supervised classification. The MTF estimation consists in estimating the MTF of the instrument used to take this picture. The learned structured dictionaries are currently studied to perform supervised classification in a context of scene recognition in satellite images. In that case, dictionaries should be learned for specific scenes. Then, patches (around each pixel) of a test picture to classify are decomposed over the different dictionaries to determine for each pixel the dictionary giving the best approximation and thus the corresponding class. A graph-cut algorithm can be applied to smooth the classification results. We are currently trying to learn more discriminant dictionaries for this specific application. For that purpose, the objective function to minimize to learn the dictionaries should not only be reconstructive, but also discriminative.

### 6.3.3. 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. The method uses 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, the codebooks are learned using the k-means algorithm. The learning algorithm proceeds in two passes so that the training set of residual vectors corresponds to the case where the vector quantization is the best mode in rate-distortion sense for the second-order prediction. It has been observed that the codebooks learned for different Quantization Parameters (QP) are very similar, leading eventually to QP-independent codebooks. A second method is being developed using clustering with Kohonen self-organizing maps in the codebook learning stage.

### 6.3.4. HDR video compression

**Participants:** Christine Guillemot, Mikael Le Pendu.

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, leading to new compression challenges. In collaboration with Technicolor, we have developed a method for converting the floating point RGB values to high bit depth integers with an approximate logarithmic encoding that is reversible without loss. This bit depth reduction is performed adaptively depending on the minimum and maximum values which characterize the dynamic of the data. A 50% rate saving has been obtained at high bitrates compared to the well-known adaptive LogLuv transform [33]. A reversible tone mapping-operator (TMO) has also been designed for efficient compression of High Dynamic Range (HDR) images using a Low Dynamic Range (LDR) encoder. Based on a statistical model of the HDR compression scheme and assumptions on the rate of the encoded LDR image, a closed form solution has been derived for the optimal tone curve in a rate-distortion sense [34].

### 6.3.5. 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 problem of HEVC encoding complexity reduction by proposing a strategy to infer UHD coding modes and quadtree from those optimized on the lower (HD) resolution version of the input video. A speed-up by a factor of 3 is achieved compared to directly encoding the UHD format at the expense of a limited PSNR-rate loss [28]. Another method which is still under investigation is to extract from the input video sequence a number of low-level features for

adapting the coding decision such as the decomposition of the quadtree. The low-level features are related to gradient-based statistics, structure tensors statistics or entropy etc.

## 6.4. 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

### 6.4.1. Universal distributed source coding

**Participant:** Aline Roumy.

In 2012, we started a new collaboration with Michel Kieffer and Elsa Dupraz (Supelec, L2S) on universal distributed source coding. Distributed source coding (DSC) refers to the problem where several correlated sources need to be compressed without any cooperation at the encoders. Decoding is however performed jointly. This problem arises in sensor networks but also in video compression techniques, where the successive frames are seen as distributed such that the correlation between the frames is not directly used at the encoder. Traditional approaches for DSC (from an information theoretical but also practical point of view) assume that the joint distribution of the sources is perfectly known. Since this assumption is not satisfied in practice, a way to get around this is to use a feedback channel (from the decoder to the encoder), that can trigger the encoder. Instead, we consider universal distributed source coding, where the joint source distribution is unknown.

More precisely, we considered the problem of compressing one source, while a second source, called side information, is available at the decoder. Further, we assumed that the conditional distribution of the side information given the source is unknown at both encoder and decoder. First, we proposed in [18] four uncertainty models for this conditional distribution, and derived the information theoretical bounds. These models differ through the (partial) knowledge on the distribution the user has. This partial knowledge includes the variation speed (slow/fast), the set of possible distributions, and eventually, some a priori distribution on the class of distributions. A complete coding scheme has also been proposed that works well for any distribution in the class. At the encoder, the proposed scheme encompasses the determination of the coding rate and the design of the encoding process. These determinations directly result from the information-theoretical compression bounds. Then a novel decoder is proposed that jointly estimate the source symbols and the conditional distribution. As the proposed decoder is based on the Expectation-Maximization algorithm, which is very sensitive to initialization, we also propose a method to produce first a coarse estimate of the distribution. The proposed scheme avoids the use of a feedback channel or the transmission of a learning sequence, which both result in a rate increase at finite length. Moreover, the proposed algorithm use non-binary LDPC codes, such that the usual binarization of the source, which induce compression inefficiency, can be avoided.

### 6.4.2. Rate Distortion analysis of Compressed sensing and distributed Compressed sensing

**Participant:** Aline Roumy.

In collaboration with Enrico Magli and Giulio Coluccia (Polito, Torino, Italy), we studied Compressed sensing as a communication tool. 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. However, the obtained measurements are reals, and therefore require an infinite precision representation. Therefore, using CS as a compression tool (in the information theoretical sense), requires to determine the trade-off between the rate necessary to encode the measurements and the distortion obtained on the data. In [17], we derive the rate-distortion (RD) function of CS and distributed CS, under the assumption that the sparsity support is perfectly known at the decoder. This provides a lower bound for any practical reconstruction algorithm.

The proof technique developed in [17] has application beyond information theory. It also provides novel analyses of CS reconstruction algorithms [27]. Classical performance analysis of reconstruction algorithms, rely on parameters that are difficult to compute (RIP, coherence of the measurement matrix), for which bounds are used. Instead, we derive exact characterization, by performing either averaged (over the measurement matrix) or asymptotic (in the size of the data) analysis.

## STARS Project-Team

## 6. New Results

### 6.1. Highlights of the Year

NeoSensys, a spin off of the Stars team which aims at commercializing video surveillance solutions for the retail domain, has been created in September 2014.

### 6.2. Introduction

This year Stars has proposed new algorithms related to its three main research axes : perception for activity recognition, semantic activity recognition and software engineering for activity recognition.

#### 6.2.1. Perception for Activity Recognition

**Participants:** Julien Badie, Slawomir Bak, Piotr Bilinski, François Brémond, Bernard Boulay, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Carolina Garate, Michal Koperski, Ratnesh Kumar, Filipe Martins, Malik Souded, Anh Tuan Nghiem, Sofia Zaidenberg, Monique Thonnat.

For perception, the main achievements are:

- Our new covariance descriptor has led to many publications and applications already. The work on this topic is now more about the precise use of the descriptor in varied applications than the design of new descriptors.
- The new action descriptors have led to finer gesture classification. As our target application is the detection of the Alzheimer syndrome from gesture analysis, which requires still finer descriptors, we will continue the work on this topic.
- The different shape priors developed (for shape growth enforcement, shape matching, articulated motion) have been formulated and designed so that efficient optimization tools could be used, leading to global optimality guarantees. These particular problems can thus be considered as solved, but there is still much work to be done on shape and related optimization, in particular to obtain shape statistics for human action recognition.
- The success obtained in the control of trackers is a proof of concept, but this work still needs to be pursued to get more practical and to be applied on more real world videos.

More precisely, the new results for perception for activity recognition are:

- People Detection for Crowded Scenes (6.3 ),
- Walking Speed Detection on a Treadmill using an RGB-D camera : experimentations and results (6.4 ),
- Head detection using RGB-D camera (6.5 ),
- Video Segmentation and Multiple Object Tracking (6.6 ),
- Enforcing Monotonous Shape Growth or Shrinkage in Video Segmentation (6.7 ),
- Multi-label Image Segmentation with Partition Trees and Shape Prior (6.8 ),
- Automatic Tracker Selection and Parameter Tuning for Multi-object Tracking (6.9 ),
- An Approach to Improve Multi-object Tracker Quality using Discriminative Appearances and Motion Model Descriptor (6.10 ),
- Person re-identification by pose priors(6.11 ),
- Global tracker : an online evaluation framework to improve tracking quality (6.12 ),
- Human action recognition in videos (6.13 ),
- Action Recognition using 3D Trajectories with Hierarchical Classifier (6.14 ),
- Action Recognition using Video Brownian Covariance Descriptor for Human (6.15 ),
- Towards Unsupervised Sudden Group Movement Discovery for Video Surveillance (6.16 ).

### 6.2.2. Semantic Activity Recognition

**Participants:** Vania Bogorny, Luis Campos Alvares, Vasanth Bathrinathan, Guillaume Charpiat, Duc Phu Chau, Serhan Cosar, Carlos F. Crispim Junior, Giuseppe Donatiello, Baptiste Fosty, Carolina Garate, Alvaro Gomez Uria Covella, Alexandra Konig, Farhood Negin, Anh-Tuan Nghiem, Philippe Robert, Carola Strumia.

For activity recognition, the main advances on challenging topics are:

- The utilization by clinicians for their everyday work of a first monitoring system able to recognize complex activities, to evaluate in real-time older people performance in an ecological room at Nice Hospital.
- The successful processing of over 80 older people videos and matching their performance for autonomy at home (e.g. walking efficiency) and cognitive disorders (e.g. realisations of executive tasks) with gold standard scales (e.g. NPI, MMSE). This research work contributes to the early detection of deteriorated health status and the early diagnosis of illness.
- The fusion of events coming from camera networks and heterogeneous sensors (e.g. RGB videos, Depth maps, audio, accelerometers).
- The management of the uncertainty of primitive events.
- The generation of event models in an unsupervised manner.

For this research axis, the contributions are :

- Autonomous Monitoring for Securing European Ports (6.17 ),
- Video Understanding for Group Behavior Analysis (6.18 ),
- Evaluation of an event detection framework for older people monitoring: from minute to hour-scale monitoring and Patients autonomy and dementia assessment (6.19 ),
- Uncertainty Modeling Framework for Constraint-based Event Detection in Vision Systems (6.20 ),
- Assisted Serious Game for older people (6.21 ),
- Enhancing Pre-defined Event Models using Unsupervised Learning (6.22 ),
- Using Dense Trajectories to Enhance Unsupervised Action Discovery (6.23 ),
- Abnormal Event Detection in Videos and Group Behavior Analysis (6.24 ).

### 6.2.3. Software Engineering for Activity Recognition

**Participants:** François Brémond, Daniel Gaffé, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Omar Abdalla, Mohamed Bouatira, Ines Sarray, Luis-Emiliano Sanchez.

For the software engineering part, the main achievements are the Software Engineering methods and tools applied to video analysis. We have demonstrated that these approaches are appropriate and useful for video analysis systems:

- Run time adaptation using MDE is a promising approach. Our current prototype resorts to tools and technologies which were readily available. This made possible a proof of concepts.
- Introducing metrics in feature models was valuable to reduce the huge set of valid configurations after a dynamic context change and to provide a real time selection of an appropriate running configuration.
- The synchronous approach is well suited to describe reactive systems in a generic way, it has a well-established formal foundation allowing for automatic proofs, and it interfaces nicely with most model-checkers.

The contributions for this research axis are:

- Model-Driven Engineering for Activity Recognition Systems(6.25 ),
- Scenario Analysis Module (6.26 ),
- The Clem Workflow (6.27 ),
- Multiple Services for Device Adaptive Platform for Scenario Recognition (6.28 ).



### 6.3. People Detection for Crowded Scenes

**Participants:** Malik Souded, François Brémond.

**keywords:** people detection, crowded scenes, features, boosting.

This work aims at proposing an efficient people detection algorithm which can deal with crowded scenes.

#### 6.3.1. Early Work

We have previously proposed an approach which optimizes state-of-the-art methods [Tuzel 2007, Yao 2008], based on training cascade of classifiers using LogitBoost algorithm on region covariance descriptors. This approach performs in real time and provides good detection performances in low to medium density scenes (see some examples in figure 10). However, this approach shows its limits on crowded scenes. Both detection accuracy and detection time are highly impacted in this case. The detection time increases dramatically due to the number of people in images, which forces the evaluation of many cascade levels, while the numerous partial occlusions highly decrease the detection rate (the considered detector is a full-body detector). To deal with these issues, we are working on a new approach.

#### 6.3.2. Current Work

Our approach is based on training a cascade of classifiers using Boosting algorithms too, but on large sets of various features with several parameters for each of them (LBP, Haar-Like, HOG, Region Covariance Descriptor, etc.). The variety of features is motivated by three main reasons:

- Using fast features like LBP and Haar-like in the first levels of the cascade allows a fast rejection of a high part of negatives. The remaining ones will be rejected by a more sophisticated feature like Covariance Descriptor. This will highly decrease the detection time.
- Covariance Descriptor are not discriminative enough for very small regions. Our aim is to train the new detector on specific body parts, especially the upper one (shoulders and heads) to increase detection rate in highly crowded scenes (with a high rate of partial occlusions). Using a large set of various features allows the training system to select the ones which provide the best discriminative power for these regions.
- The possibility to combine several features to describe the same region, even by a simple concatenation, providing more discriminative power than using single features.

Another part of this approach consists in the optimization of the detector at two levels:

- Optimizing the training process by first clustering both positive and negative training samples. This clustering allows to focus on the hard samples which are too close to the other class from a classification point of view, providing more accurate detectors.
- Iterative training of several detectors on randomly selected samples, and weighting of the training samples according to their classification confidence, which allows to improve the clustering process.

The evaluation of this approach is still in progress.

### 6.4. Walking Speed Detection on a Treadmill using an RGB-D camera : experimentations and results

**Participants:** Baptiste Fosty, François Brémond.

**keywords:** RGB-D camera analysis, walking speed, serious games

Within the context of the development of serious games for people suffering from Alzheimer disease (Az@Game project), we have developed an algorithm to compute the walking speed of a person on a treadmill. The goal is to use this speed inside the game to control the displacement of the avatar, and then for the patient to perform some physical as well as cognitive tasks. For the evaluation of the accuracy of the algorithm, we collected a video data set of healthy people walking on a motorized treadmill.



Figure 10. Some examples of detection using the previously proposed approach (see section Early Work).

**Protocol.** With the help of a specialist in the domain of physical activities, a protocol has been set up to cover the spectrum of the possible walking speeds and to prove the reproducibility of the results. This protocol consists in performing three times ten minutes of walking on the motorized treadmill, each attempt being itself divided in five times two minutes at the following speeds : 1.5 km/h, 2.5 km/h, 3.5 km/h, 4.5 km/h and 5.5 km/h. Participants, mostly people from the age of 18 to 60 without any physical disorder that could influence the gait, were asked to keep a natural gait and to follow the rotation of the treadmill.

**Ground truth.** The quantitative performances of the walking speed computation are evaluated by comparison with the speed of the walking person. The speed references are twofold :

- a theoretical value : the speed displayed by the treadmill, set up by the participant but imposed by the protocol (see Figure 12 , red graph),
- a practical value : white marks have been painted on the treadmill to recompute the real speed of the rotation and so the walking speed (see Figure 12 , green and blue graphs).

**Results.** The results presented herein are based on the videos of 36 participants who performed the protocol described above, with 17 males and 19 females, with an average age of  $32.1 \pm 7.7$  years, an average height of  $171.1 \pm 9.1$  cm and an average weight of  $67.4 \pm 13.6$  kg.

Speed (km/h)	Number of observations	Average (km/h)	Standard deviation (km/h)	Mean absolute error (km/h)
1.5 (WM)	11823 (3170)	1.59 (1.508)	0.20 (0.014)	0.15 (0.011)
2.5 (WM)	14549 (5113)	2.56 (2.478)	0.20 (0.026)	0.16 (0.027)
3.5 (WM)	17399 (7133)	3.53 (3.453)	0.20 (0.040)	0.16 (0.051)
4.5 (WM)	19800 (9129)	4.47 (4.427)	0.21 (0.059)	0.17 (0.079)
5.5 (WM)	22163 (11180)	5.36 (5.397)	0.26 (0.116)	0.21 (0.116)
<b>Total (WM)</b>	<b>85734 (35725)</b>	<b>3.5 (3.453)</b>	<b>0.22 (0.074)</b>	<b>0.17 (0.071)</b>

Figure 11. Walking speed results. The number of observations corresponds to the number of time the speed has been computed. The WM lines refer to the results of the detection of the white mark.

The table in figure 11 shows the statistical evaluation of the performances of the system. The average column shows that the accuracy of the system is better for the median speeds (around 4.5km/h). When the person is walking slower, the system overestimate the speed due to the wrongly detected steps whereas when faster, there is an underestimation because of missing the exact time when the distance between feet is maximum (framerate too low).

A paper reporting this work is actually under writing process.

## 6.5. Head Detection Using RGB-D Camera

**Participants:** Marine Chabran, François Brémond.

**keywords:** RGB-D camera analysis, head detection, serious games

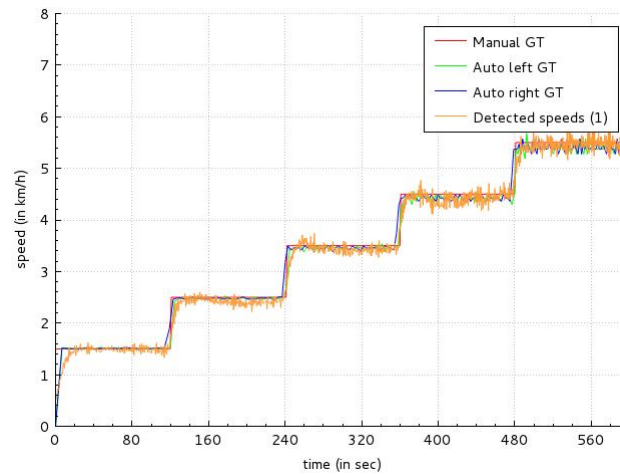


Figure 12. Graphical results of one attempt of protocol issue. The red graph corresponds to the theoretical speed, green and blue graphs correspond respectively to the detected speed of the left and right white marks on the treadmill, and the orange graph is the detected walking speed of the person.

The goal of this work is to improve a head detection algorithm using RGB-D sensor (like a Kinect camera) for action recognition as part of a study of autism. The psychologists want to compare the learning process of children with autism syndrome depending on games (digital or physical toys).

The algorithm described in [79] represents a head by its center position. It takes three steps to determine this point :

- Determine possible head center positions using a head model : inner circle radius=6 cm, outer circle radius=20 cm (Figure 13 ).

A good inner point is a point on the inner circle verifying :

$$depthHeadCenter + 30cm > depthInnerPoint > depthHeadCenter - 30cm.$$

A good outer point is a point on the outer circle verifying :

$$depthHeadCenter < depthOuterPoint + 15cm.$$

- Merge close head centers separated by less than 4 pixels.
- Select final head center according to its score (calculated according to the number of good inner and outer points).

For now, it works well within video where people are close to the camera (about 1 meter) and without any background just behind them (Figure 14 ).

The problem is when the person is sitting and the head is ahead of the body (Figure 15 ) or close to a wall, the difference between head depth and outer circle depth becomes not sufficient (about 10 cm).

We have evaluated the performance of this algorithm with two data sets (Table 1 ). For Lenval Hospital data set, we have evaluated 2 series of 200 frames, for the Smart Home data set, we have evaluated 3 series of 300 frames (a total of 1300 heads).

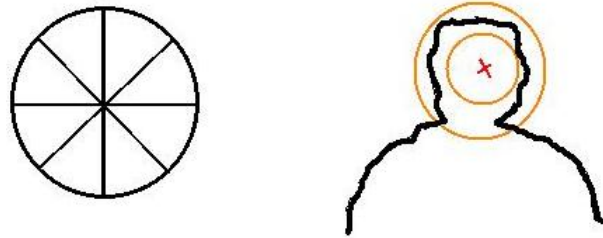


Figure 13. Each circle is divided in  $n$  parts ( $n=8$ ). The points on the inner circle must have a similar depth with the center point, the points on the outer circle must be further than the center point compared to the camera

Table 1. Performance of head detection and people detection on two different data sets.

Videos	Head Detection (%)	People detection (%)
Lenval Hospital dataset (Figure 14 )	89.7	96.9
Rest home dataset (Figure 15 )	62.8	85.3



Figure 14. Result of head detection - good detection. The bounding box represents the person, the small blue circle represents the head center.



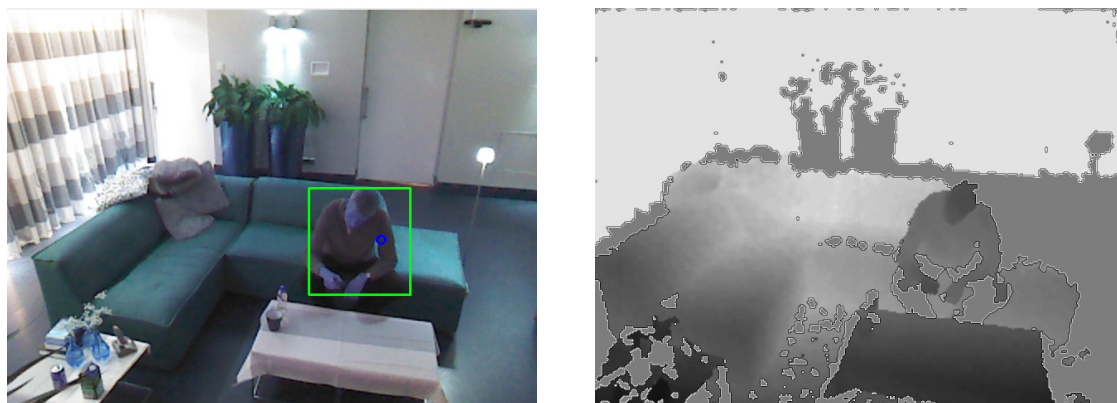


Figure 15. Result of head detection - wrong detection. The green bounding box represents the person, the small blue circle represents the head center.

## 6.6. Video Segmentation and Multiple Object Tracking

**Participants:** Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

**keywords:** Fibers, Graph Partitioning, Message Passing, Iterative Conditional Modes, Video Segmentation, Video Inpainting

This year we focussed on multiple object tracking, and writing of the thesis manuscript of Ratnesh (defense on December 2014).

The first contribution of this thesis is in the domain of video segmentation wherein the objective is to obtain a dense and coherent spatio-temporal segmentation. We propose joining both spatial and temporal aspects of a video into a single notion *Fiber*. A Fiber is a set of trajectories which are spatially connected by a mesh. Fibers are built by jointly assessing spatial and temporal aspects of the video. Compared to the state-of-the-art, a fiber based video segmentation presents advantages such as a natural spatio-temporal neighborhood accessor by a mesh, and temporal correspondences for most pixels in the video. Furthermore, this fiber-based segmentation is of quasi-linear complexity w.r.t. the number of pixels. The second contribution is in the realm of multiple object tracking. We proposed a tracking approach which utilizes cues from point tracks, kinematics of moving objects and global appearance of detections. Unification of all these cues is performed on a Conditional Random Field. Subsequently this model is optimized by a combination of message passing and an Iterated Conditional Modes (ICM) variant to infer object-trajectories. A third, minor, contribution relates to the development of suitable feature descriptor for appearance matching of persons. All of our proposed approaches achieve competitive and better results (both qualitatively and quantitatively) than state-of-the-art open source datasets.

This first part of the thesis was published at IEEE WACV at the beginning of this year [43], and the work on *multiple object tracking* was recently presented at Asian Conference on Computer Vision [44]

Sample visual results from our recent publication [44] can be seen in Figure 16 .

## 6.7. Enforcing Monotonous Shape Growth or Shrinkage in Video Segmentation

**Participant:** Guillaume Charpiat [contact].

*This work has been done in collaboration with Yuliya Tarabalka (Ayin team, Inria-SAM), Bjoern Menze (Technische Universität München, Germany), and Ludovic Brucker (NASA GSFC, USA) [<http://www.nasa.gov>].*



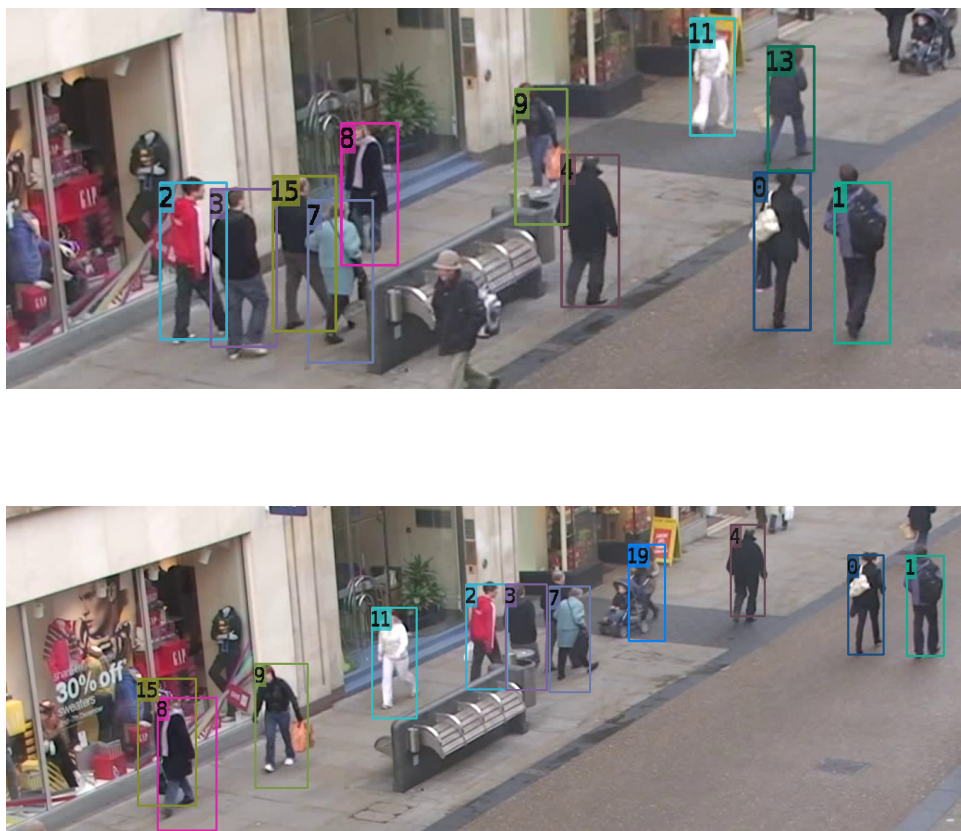


Figure 16. Consistent people crossing in dense scenarios. The two images are 121 frames apart.

**keywords:** Video segmentation, graph cut, shape analysis, shape growth

The automatic segmentation of objects from video data is a difficult task, especially when image sequences are subject to low signal-to-noise ratio or low contrast between the intensities of neighboring structures. Such challenging data are acquired routinely, for example, in medical imaging or satellite remote sensing. While individual frames can be analyzed independently, temporal coherence in image sequences provides a lot of information not available for a single image. In this work, we focused on segmenting shapes that grow or shrink monotonically in time, from sequences of extremely noisy images.

We proposed a new method for the joint segmentation of monotonically growing or shrinking shapes in a time sequence of images with low signal-to-noise ratio [32]. The task of segmenting the image time series is expressed as an optimization problem using the spatio-temporal graph of pixels, in which we are able to impose the constraint of shape growth or shrinkage by introducing unidirectional infinite-weight links connecting pixels at the same spatial locations in successive image frames. The globally-optimal solution is computed with graph-cuts. The performance of the proposed method was validated on three applications: segmentation of melting sea ice floes; of growing burned areas from time series of 2D satellite images; and of a growing brain tumor from sequences of 3D medical scans. In the latter application, we imposed an additional inter-sequences inclusion constraint by adding directed infinite-weight links between pixels of dependent image structures. Figure 17 shows a multi-year sea ice floe segmentation result. The proposed method proved to be robust to high noise and low contrast, and to cope well with missing data. Moreover, in practice, its complexity was linear in the number of images.

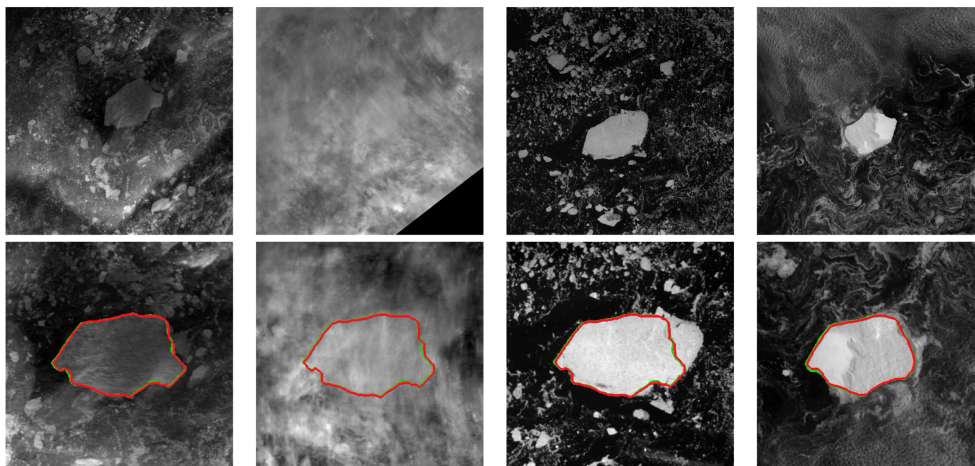


Figure 17. Top: MODIS images at four time moments (days 230, 233, 235 and 267 of 2008, respectively). Bottom: corresponding aligned images with segmentation contours (in red). Manual segmentation (ground truth) is shown in green.

## 6.8. Multi-label Image Segmentation with Partition Trees and Shape Prior

**Participant:** Guillaume Charpiat [contact].

*This work has been done in collaboration with Emmanuel Maggiori and Yuliya Tarabalka (Ayin team, Inria-SAM).*

**keywords:** partition trees, multi-class segmentation, shape priors, graph cut

The multi-label segmentation of images is one of the great challenges in computer vision. It consists in the simultaneous partitioning of an image into regions and the assignment of labels to each of the segments. The problem can be posed as the minimization of an energy with respect to a set of variables which can take one of multiple labels. Throughout the years, several efforts have been done in the design of algorithms that minimize such energies.

We propose a new framework for multi-label image segmentation with shape priors using a binary partition tree [50]. In the literature, such trees are used to represent hierarchical partitions of images, and are usually computed in a bottom-up manner based on color similarities, then processed to detect objects with a known shape prior. However, not considering shape priors during the construction phase induces mistakes in the later segmentation. This study proposes a method which uses both color distribution and shape priors to optimize the trees for image segmentation. The method 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 reducing the search space and make the optimization efficient. Our experiments (see Figure 18) show that the optimization approach succeeds in incorporating shape information into multi-label segmentation, outperforming the state-of-the-art.

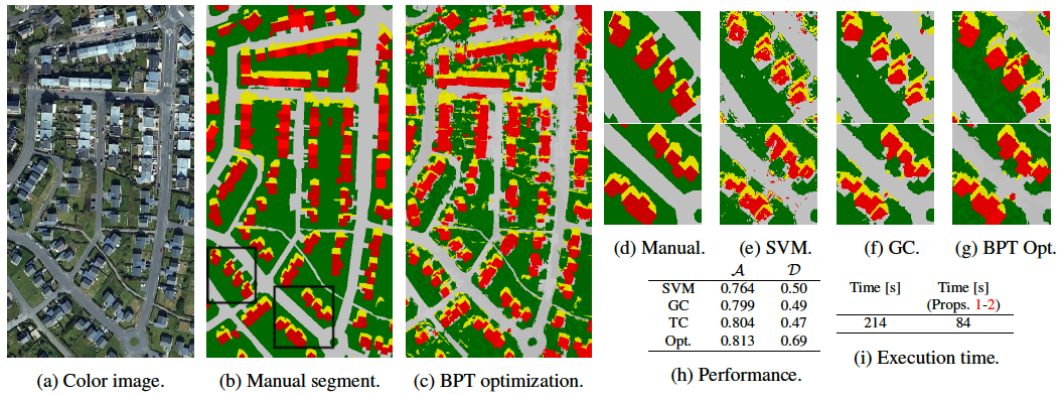


Figure 18. Classification results for the satellite image over Brest.  $\mathcal{A}$  denotes overall classification accuracy, and  $\mathcal{D}$  denotes average building's overlap. The performance of the proposed binary partition tree (BPT) optimization method is compared with the following methods: 1) support vector machines (SVM) classification; 2) graph cut (GC) with  $\alpha$ -expansion; 3) cut on the BPT, regularized by the number of regions without using shape priors (TC).

## 6.9. Automatic Tracker Selection and Parameter Tuning for Multi-object Tracking

**Participants:** Duc Phu Chau, Slawomir Bak, François Brémond, Monique Thonnat.

**Keywords:** object tracking, machine learning, tracker selection, parameter tuning

Many approaches have been proposed to track mobile objects in a scene [87], [45]. However the quality of tracking algorithms always depends on video content such as the crowded level or lighting condition. The selection of a tracking algorithm for an unknown scene becomes a hard task. Even when the tracker has already been determined, there are still some issues (e.g. the determination of the best parameter values or the online estimation of the tracking reliability) for adapting online this tracker to the video content variation. In order to overcome these limitations, we propose the two following approaches.

The main idea of the first approach is to learn offline how to tune the tracker parameters to cope with the tracking context variations. The tracking context of a video sequence is defined as a set of six features: density of mobile objects, their occlusion level, their contrast with regard to the surrounding background, their contrast variance, their 2D area and their 2D area variance. In an offline phase, training video sequences are classified by clustering their contextual features. Each context cluster is then associated to satisfactory tracking parameters using tracking annotation associated to training videos. In the online control phase, once a context change is detected, the tracking parameters are tuned using the learned parameter values. This work has been published in [30].

A limitation of the first approach is the need of annotated data for training. Therefore we have proposed a second approach without training data. In this approach, the proposed strategy combines an appearance tracker and a KLT tracker for each mobile object to obtain the best tracking performance (see figure 19). This helps to better adapt the tracking process to the spatial distribution of objects. Also, while the appearance-based tracker considers the object appearance, the KLT tracker takes into account the optical flow of pixels and their spatial neighbours. Therefore these two trackers can improve alternately the tracking performance.

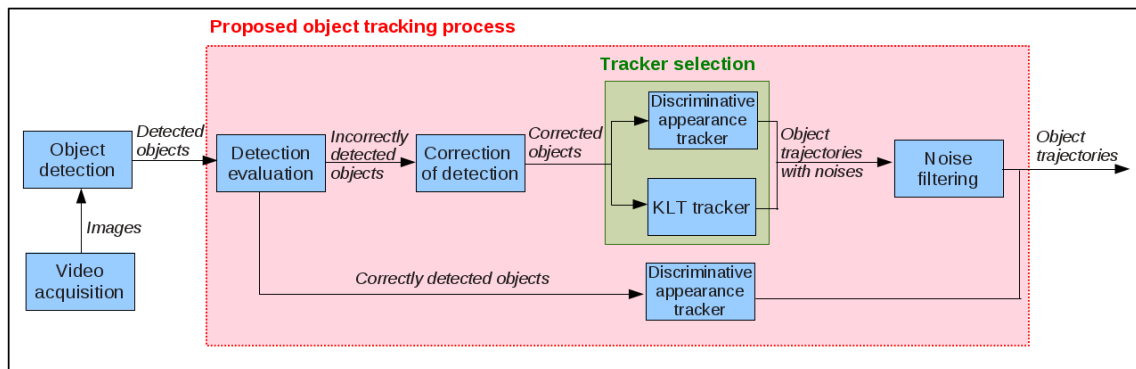


Figure 19. The scheme of the second approach

The second approach has been experimented on three public video datasets. Figure 20 presents correct tracking results of this approach even with strong object occlusion in PETS 2009 dataset. Table 2 presents the evaluation results of the proposed approach, the KLT tracker, the appearance tracker and different trackers from the state of the art. While using separately the KLT tracker or the appearance tracker, the performance is lower than other approaches from the state of the art. The proposed approach by combining these two trackers improves significantly the tracking performance and obtains the best values for both metrics. This work has been published in [39].

Table 2. Tracking results on the PETS sequence S2.L1, camera view 1, sequence time 12.34. The best values are printed in bold.

Method	MOTA	MOTP
Berclaz et al. [60]	0.80	0.58
Shitrit et al. [86]	0.81	0.58
KLT tracker	0.41	<b>0.76</b>
Appearance tracker	0.62	0.63
<b>Proposed approach</b>	<b>0.86</b>	0.72



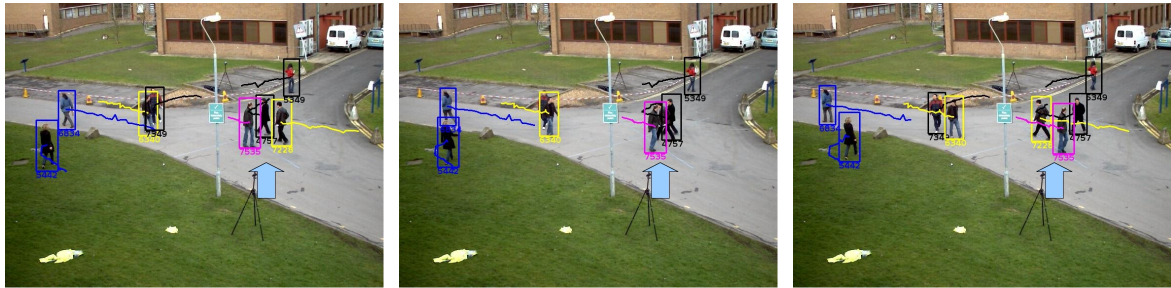


Figure 20. Tracker result: Three persons of Ids 7535, 7228 and 4757 (marked by the cyan arrow) are occluded each other but their identities are kept correctly after occlusion.

## 6.10. An Approach to Improve Multi-object Tracker Quality Using Discriminative Appearances and Motion Model Descriptor

**Participants:** Thi Lan Anh Nguyen, Duc Phu Chau, François Brémond.

**Keywords:** Tracklet fusion, Multi-object tracking

Many recent approaches have been proposed to track multi-objects in a video. However, the quality of trackers is remarkably affected by video content. In the state of the art, several algorithms are proposed to handle this issue. The approaches in [39] and [64] propose methods which compute online or learn descriptor weights during tracking process. These algorithms adapt the tracking to the scene variations but are less effective when mis-detection occurs in a long period of time. Inversely, the algorithms in [59] and [58] can recover a long-term mis-detection by fusing tracklets. However, the descriptor weights in these tracklet fusion algorithms are fixed in the whole video. Furthermore, above algorithms track objects based on object appearance which is not reliable enough when objects look similar to each other.

In order to overcome mentioned issues, the proposed approach brings three contributions: (1) appearance descriptors and motion model combination, (2) online discriminative descriptor weight computation and (3) discriminative descriptors based tracklet fusion. In particular, the appearance of one object can be discriminative with other objects in this scene but can be similar with other objects in another scene. Therefore, tracking objects based on only object appearance is less effective. In order to improve tracker quality, assuming that objects move with constant velocity, this approach firstly combines a constant velocity model from [70] and other appearance descriptors. Continuously, discriminative descriptor weights are computed online to adapt the tracking to each video scene. The more a descriptor discriminates one tracklet over other tracklets, the higher its weight value is. Next, based on these descriptor weights, the similarity score between the target tracklet with its candidate is computed. In the last step, tracklets are fused to a long trajectory by Hungarian algorithm with the optimization of global similarity scores.

The proposed approach gets results of tracker in [63] as input and is tested on challenge datasets. This approach achieves comparable results with other trackers from the state of the art. Figure 1 shows that the tracklet keeps its ID even when occlusion occurs. Table 1 shows the better performance of this approach compared to other trackers from the state of the art.

## 6.11. Person Re-identification by Pose Priors

**Participants:** Slawomir Bak, Sofia Zaidenberg, Bernard Boulay, Filipe Martins, Francois Brémond.

**keywords:** re-identification, pose estimation, metric learning

**Human appearance registration, alignment and pose estimation**

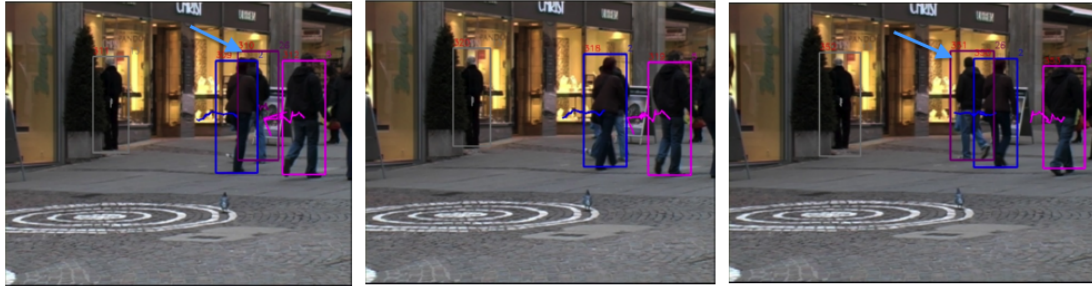
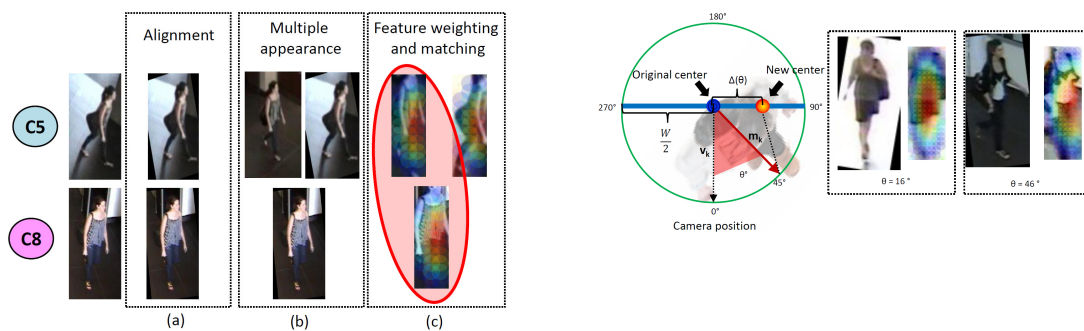


Figure 21. The proposed approach with TUD dataset: Object ID<sub>26</sub> (presented by pink bounding box) keeps its ID correctly after 11 frames of mis-detection.

Table 3. Tracking results on datasets: TUD-Stadtmitte and TUD-crossing. The best values are printed in bold

Dataset	Method	MT(%)	PT(%)	ML(%)
TUD-Stadtmitte	[57]	60.0	30.0	10.0
TUD-Stadtmitte	[30]	<b>70.0</b>	10.0	20.0
TUD-Stadtmitte	[71]	<b>70.0</b>	30.0	<b>0.0</b>
TUD-Stadtmitte	[95]	<b>70.0</b>	30.0	<b>0.0</b>
TUD-Stadtmitte	Ours	<b>70.0</b>	30.0	<b>0.0</b>
TUD-Crossing	[89]	<b>53.8</b>	38.4	7.8
TUD-Crossing	Ours	<b>53.8</b>	46.2	<b>0.0</b>



(a) left

(b) right

Figure 22. Improvements on re-identification using viewpoint cues: (a) target alignment; (b) multiple target appearance based on clustering; (c) pose orientation-driven weighting. The left illustration shows an example of the same person viewed from two different cameras. The right image presents pose estimation algorithm.



Re-identifying people in a network of cameras requires an invariant human representation. State of the art algorithms are likely to fail in real-world scenarios due to serious perspective changes. Most of existing approaches focus on invariant and discriminative features, while ignoring the body alignment issue. In this work we proposed 3 methods for improving the performance of person re-identification. We focus on eliminating perspective distortions by using 3D scene information. Perspective changes are minimized by affine transformations of cropped images containing the target (1). Further we estimate the human pose for (2) clustering data from a video stream and (3) weighting image features. The pose is estimated using 3D scene information and motion of the target. Pose orientation is computed by dot product between viewpoint vector and motion of the target (see figure 22 ). We validated our approach on a publicly available dataset with a network of 8 cameras. The results demonstrated significant increase in the re-identification performance over the state of the art [36].

#### Matching employing pose priors

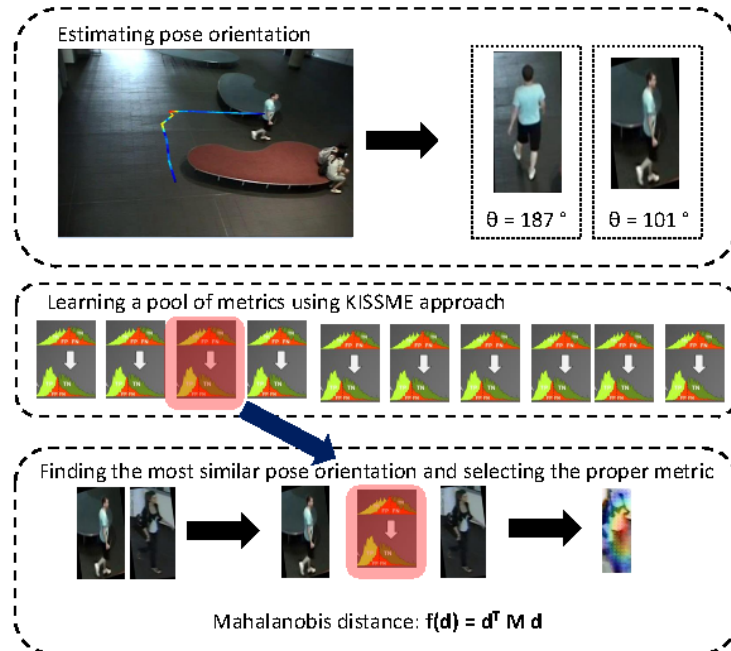


Figure 23. The proper metric is selected from the pool of previously learned metric for matching different poses.

Currently we are working on learning the matching strategy of appearance extracted from different poses. We employ well known metric learning tools for matching given poses. Let us assume that pose can be described by the angle between the motion vector of the target and the viewpoint vector of the camera (see figure 22 ). Thus for each target appearance we can express the pose as the angle in the range of  $[0,360)$ . We decide to divide this range into  $n$  bins. Given  $n$  bins of estimated poses, we learn how to match different poses corresponding to different bins. In the result, we learn  $n * (n + 1)/2$  metrics. While learning metrics, we follow a well known scheme based on image pairs, containing two different poses of the same target as positives and pairs of different poses containing different targets as negatives. The learned metrics stand for the metric pool. This metric pool is learned offline and does not depend on camera pair. In the result, once metric pool is learned, it can be used for any camera pair.

Given two images from different (or the same) camera, we first estimate the poses for each image. Having two poses, we select a corresponding metric from the metric pool. The selected metric provides the strategy to compute similarity between two images (see figure 23 ).

## 6.12. Global Tracker : An Online Evaluation Framework to Improve Tracking Quality

**Participants:** Julien Badie, Slawomir Bak, Duc Phu Chau, François Brémond, Monique Thonnat.

**keywords:** online quality estimation, re-identification, tracking results improvements

This work addresses the problem of estimating the reliability of a tracking algorithm during runtime and correcting the anomalies found. Evaluating and tuning a tracking algorithm generally requires multiple runs and ground truth. The proposed framework called global tracker overcomes these limitations by combining an online evaluation algorithm and a recovering post-process.

Designing an evaluation framework that does not require ground truth has many different applications. One of them is to provide feedback to the tracking algorithm that can tune its own parameters to improve the results on the next frame. Another convenient application is to filter the reliable information from the tracking algorithm that can be used by the next processing step such as event recognition or re-identification.

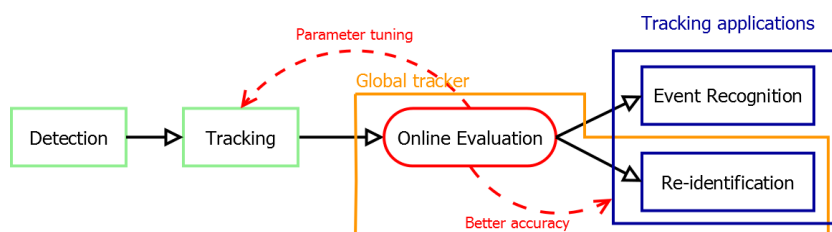


Figure 24. Integration of the global tracker, combining online evaluation and re-identification

The proposed online evaluation framework is based on control features. It means that several representative values or characteristics (the features) are chosen to monitor what is happening. Control features are the features that the online evaluation framework uses to monitor the status of the tracked objects. The framework is divided into two steps :

- computing the control features related to each tracked object of the scene on the current frame
- detecting the possible anomalies and then classifying them into two categories : anomalies due to tracking errors or benign anomalies (when a person leaves the scene or cross an obstacle for example).

This approach has been tested on two datasets (PETS 2009 and Caviar) with two different tracking algorithms (a multi-feature tracker and a tracker based on graph partitioning). The results show that the global tracker, even associated with a tracking algorithm that does not have good results, can perform as well as the state of the art.

This approach has been published in AVSS 2014 [33] which details the differences between real errors and benign anomalies.

## 6.13. Human Action Recognition in Videos

**Participants:** Piotr Bilinski, François Brémond.

**keywords:** Action Recognition; Human Action Recognition

Table 4. Tracking results on sequence S2.L1.View1 of the PETS2009 dataset using CLEAR metrics

Methods	MOTA	MOTP
Heili et al. [72]	0.89	0.66
Zamir et al. [96]	<b>0.90</b>	0.69
Milan et al. [77]	<b>0.90</b>	<b>0.74</b>
<b>Tracker 1</b>	0.62	0.63
<b>Tracker 1 + global tracker</b>	0.85	0.71
<b>Tracker 2</b>	0.85	<b>0.74</b>
<b>Tracker 2 + global tracker</b>	<b>0.90</b>	<b>0.74</b>

Table 5. Tracking results on the Caviar dataset using Mostly Tracked (MT), Partially Tracked (PT) and Mostly Lost (ML) metrics

Method	MT (%)	PT (%)	ML (%)
Li et al. [76]	84.6	14.0	1.4
Kuo et al. [74]	84.6	14.7	<b>0.7</b>
<b>Tracker 1</b>	78.3	16.0	5.7
<b>Tracker 1 + global tracker</b>	<b>86.4</b>	<b>8.3</b>	5.3

This Ph.D. thesis targets the automatic recognition of human actions in videos. Human action recognition is defined as a requirement to determine what human actions occur in videos. This problem is particularly hard due to enormous variations in visual and motion appearance of people and actions, camera viewpoint changes, moving background, occlusions, noise, and enormous amount of video data.

Firstly, we review, evaluate, and compare the most popular and the most prominent state-of-the-art techniques, and we propose our action recognition framework based on local features, which we use throughout this thesis work embedding the novel algorithms. Moreover, we introduce a new dataset (CHU Nice Hospital) with daily self care actions of elder patients in a hospital.

Then, we propose two local spatio-temporal descriptors for action recognition in videos. The first descriptor is based on a covariance matrix representation, and it models linear relations between low-level features. The second descriptor is based on a Brownian covariance, and it models all kinds of possible relations between low-level features.

Then, we propose three higher-level feature representations to go beyond the limitations of the local feature encoding techniques.

The first representation is based on the idea of relative dense trajectories. We propose an object-centric local feature representation of motion trajectories, which allows to use the spatial information by a local feature encoding technique.

The second representation encodes relations among local features as pairwise features. The main idea is to capture the appearance relations among features (both visual and motion), and use geometric information to describe how these appearance relations are mutually arranged in the spatio-temporal space.

The third representation captures statistics of pairwise co-occurring visual words within multi-scale feature-centric neighbourhoods. The proposed contextual features based representation encodes information about local density of features, local pairwise relations among the features, and spatio-temporal order among features.

Finally, we show that the proposed techniques obtain better or similar performance in comparison to the state-of-the-art on various, real, and challenging human action recognition datasets (Weizmann, KTH, URADL, MSR Daily Activity 3D, HMDB51, and CHU Nice Hospital).

The Ph.D. thesis was defended on December 5, 2014.

## 6.14. Action Recognition Using 3D Trajectories with Hierarchical Classifier

**Participants:** Michal Koperski, Piotr Bilinski, François Brémond.

**keywords:** action recognition, computer vision, machine learning, 3D sensors

The goal of our work is to extend recently published approaches ([61], [93]) for Human Action Recognition to take advantage of the depth information from 3D sensors.

We propose to add depth information to trajectory based algorithms ([61], [93]). Currently mentioned algorithms compute trajectories by sampling video frames and then tracking points of interest - creating the trajectory. Our contribution is to create even more discriminative features by adding depth information to previously detected trajectories. In our work we propose methods to deal with noise and missing measurements in depth map.

The second contribution is a technique to deal with actions which do not contain enough motion to compute discriminative trajectory descriptors. Actions like sitting, standing, laptop use do not contain large amount of motion, or motion is occluded by the object. For such cases we proposed LDP (Local Depth Pattern) descriptor which does not require motion to be computed.

Proposed descriptors are further processed using a Bag of Words method and SVM classifier. We use hierarchical approach where at first level we train classifier to recognize if given example contains high or low amount of motion. Then at second layer we train SVM classifier to recognize action labels.



Figure 25. Visualization of MSR Daily Activity 3D data set (left) - video input frame, (center) - frame with detected trajectories (red - static points, green detected trajectories, (right) - corresponding depth map

The evaluation of our method was conducted on "Microsoft Daily Activity3D" data set [94] which consists of 16 actions (drink, eat, read book, call cellphone, write on a paper, use laptop etc.) performed by 10 subjects. We achieve superior performance among techniques which do not require skeleton detection. This work was published in proceedings of the 21st IEEE International Conference on Image Processing, ICIP 2014 [42]

## 6.15. Action Recognition using Video Brownian Covariance Descriptor for Human

**Participants:** Piotr Bilinski, Michal Koperski, Slawomir Bak, François Brémond.

**keywords:** action recognition, computer vision, machine learning

This work addresses a problem of recognizing human actions in video sequences. Recent studies have shown that methods which use bag-of-features and space-time features achieve high recognition accuracy [61], [93], [42]. Such methods extract both appearance-based and motion-based features. In image processing, a novel trend has emerged that ignores explicit values of given features, focusing instead on their pairwise relations. The most known example of such an approach is covariance descriptor [92]. Inspired by Brownian motion statistics [88] and application in people Re-identification [35]; we propose to model relationships between different pixel-level appearance features such as intensity and gradient using Brownian covariance, which is a natural extension of classical covariance measure. While classical covariance can model only linear relationships, Brownian covariance models all kinds of possible relationships. We propose a method to compute Brownian covariance on space-time volume of a video sequence. We show that proposed Video Brownian Covariance (VBC) descriptor carries complementary information to the Histogram of Oriented Gradients (HOG) descriptor. The fusion of these two descriptors gives a significant improvement in performance on three challenging action recognition datasets. The result of this work was published in proceedings of the 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2014 [38].

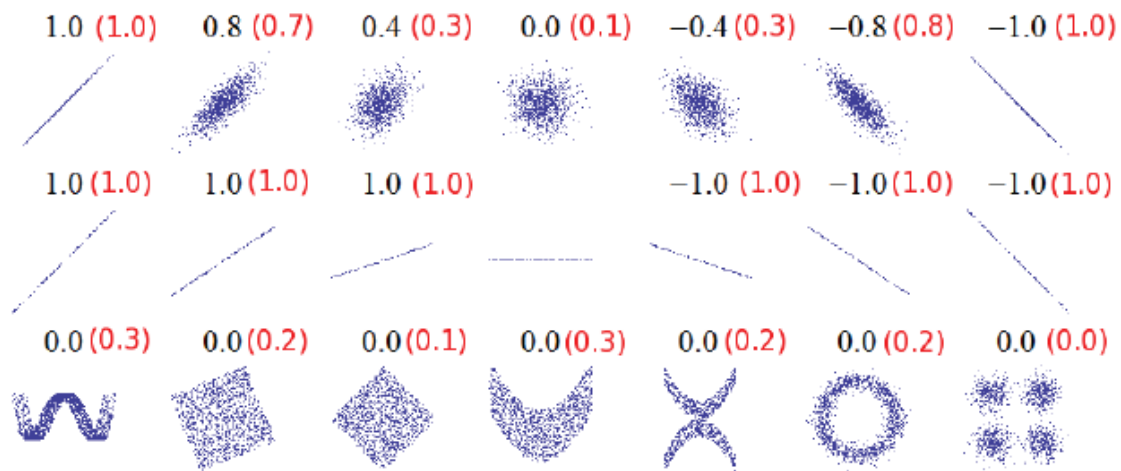


Figure 26. Comparison between Covariance and Brownian distance correlation. Covariance values in black, Brownian values in red.

## 6.16. Towards Unsupervised Sudden Group Movement Discovery for Video Surveillance

**Participants:** Sofia Zaidenberg, Piotr Bilinski, François Brémond.

**keywords:** Event detection; Motion estimation; Anomaly estimation; Situation awareness; Scene Understanding; Group Activity Recognition; Stream Selection

We present a novel and unsupervised approach for discovering “sudden” movements in surveillance videos. The proposed approach automatically detects quick motions in a video, corresponding to any action. A set of possible actions is not required and the proposed method successfully detects potentially alarm-raising actions without training or camera calibration. Moreover, the system uses a group detection and event recognition framework to relate detected sudden movements and groups of people, and to provide a semantical interpretation of the scene. We have tested our approach on a dataset of nearly 8 hours of videos recorded from

two cameras in the Parisian subway for a European Project. For evaluation, we annotated 1 hour of sequences containing 50 sudden movements.

This work has been published in [47].

## 6.17. Autonomous Monitoring for Securing European Ports

**Participants:** Vasanth Bathrinathan, François Brémont.

**Keywords:** Event Recognition, Port Surveillance

This work is done for the European research project SUPPORT (Security UPgrade for PORTs). This project addresses potential threats on passenger life and the potential for crippling economic damage arising from intentional unlawful attacks on port facilities, by engaging representative stakeholders to guide the development of next generation solutions for upgraded preventive and remedial security capabilities in European ports. The overall benefit is securing and efficient operation of European ports enabling uninterrupted flows of cargo and passengers while suppressing attacks on high value port facilities, illegal immigration and trafficking of drugs, weapons and illicit substances.

Scene understanding platform was tested on this new dataset, which has archived footage from past incidents and some acted scenarios. The processing pipeline of algorithms contains camera calibration, background subtraction using GMM (Gaussian Mixture Model), people detection using DPM (Deformable Parts Model), Tracking (Frame to Frame), Event recognition.

We collected several hours of videos which contained security related events like Intrusion to port by different methods (sea, gates, fences), Spying activities from outside the port, robbery or theft, ticketless travelling, restricted zone access, abandon luggage and some abnormal behaviors. The system was modelled and validated for all the above events to be detected and also a live real time demo was done for the completion of the project. All the events from our systems are later sent to project partners for fusion of the data with other sensors data and information from police, internal and external reports, etc., to detect complex security threats (see figure 27).

## 6.18. Video Understanding for Group Behavior Analysis

**Participants:** Carolina Garate, François Brémont.

**keywords:** Computer vision, group tracking, scene understanding, group behavior recognition, video surveillance, event detection.

The main work in this PhD thesis concerns the recognition of the behaviors of a group of people (2-5 persons) involved in a scene depicted by a video sequence.

Our goal focuses on the automatic recognition of behavior patterns in video sequence for groups of people (2-5 persons). We want to build a real time system able to recognize various group scenarios.

The approach includes different tasks to achieve the final recognition. The first one consists in tracking groups of moving regions detected in the video sequence acquired by the cameras. The second task attempts to classify these moving regions into people classes. Finally, the last task recognizes group scenarios using a priori knowledge containing scenario models predefined by experts and also 3D geometric and semantic information of the observed environment.

Our approach considers a chain process consisting of 5 consecutive steps for video processing. The steps are : 1) segmentation, 2) physical object detection, 3) physical objects tracking, 4) group tracking and 5) group behavior recognition. Our research focuses on the last two phases.

First, group scenarios have been defined (and then recognized) using the general scenario description language. Second, the likelihood of the group scenario recognition has been quantified. Third, machine learning techniques have been investigated to learn and recognize these scenarios.



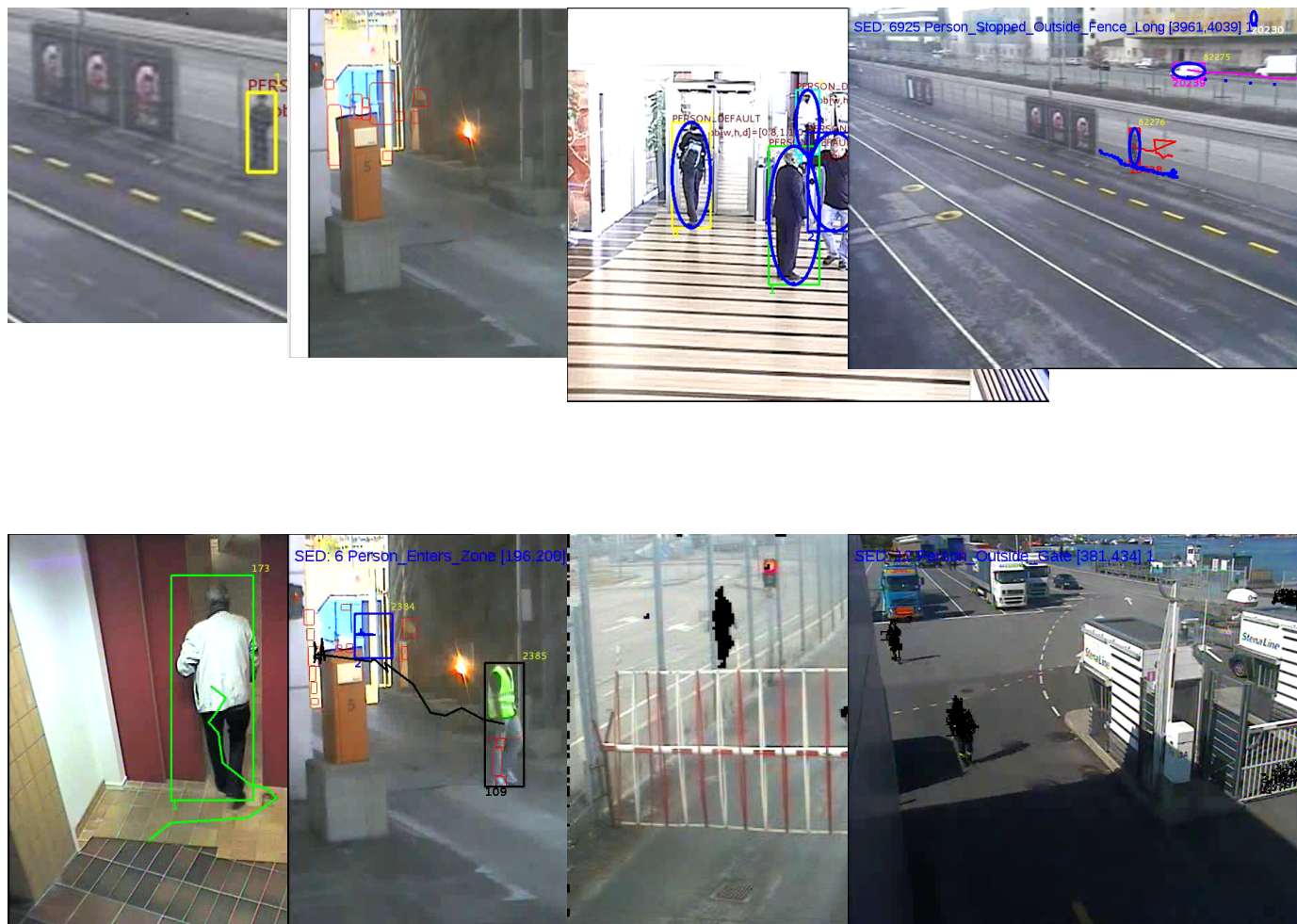


Figure 27. Detection of complex security threats from the tested dataset for Port surveillance

We have processed the data set from 1 month video surveillance camera in the Torino subway and the Minds eye data set. Recognizing several and different events such as: walking groups, standing still groups, running groups, calm groups (i.e. having a bounding box with stable size), active groups (i.e. with bounding box's size variations, meaning that group members move a lot).

## 6.19. Evaluation of an Event Detection Framework for Older People Monitoring: from Minute to Hour-scale Monitoring and Patients Autonomy and Dementia Assessment

**Participants:** Carlos F. Crispim-Junior, Alvaro Gomez Uria Covella, Carola Strumia, Baptiste Fosty, Duc Phu Chau, Anh-Tuan Nghiem, Alexandra Konig, Auriane Gros, Philippe Robert, François Brémond.

**keywords:** RGBD cameras, description-based activity recognition, older people,

Two main works are reported here: the continuous evaluation and extension of our event detection framework for older people monitoring, and the proposal of a behavioral classification model for the assessment of autonomy and cognitive health level of older people using automatically detected events. The evaluation of our event monitoring framework was extended from 29 to 49 recordings of senior participants undertaking physical tasks (7 min per participant, total : 5.71 hours) and instrumental activities of daily living (IADL, 15 minutes per participant, total: 12.25 hours). The recordings have taken place in a ecological observation room set in the Memory Center of Nice hospital. In the extended evaluation we employed a RGBD sensor as input instead of a standard RGB camera due to its advantages like invariance to illumination changes and real-time measurements of 3D information which foster better performance of the underlying algorithms for people detection and tracking. Table 6 presents the event monitoring performance of the present system for 49 participants. Event detection performance on physical task generalized to the larger dataset with a small performance increase of 1.4% (average F-Score). Concerning IADL detection although the global performance value (F-Score, 80.7 %) are the same, the new approach have made a trade-off between recall and precision to obtain more reliable detection of activities and their parameter estimations. Low precision values on preparing drink (e.g., making coffee) and watering plant are due to these activities being performed in very close - if not overlapping - locations (contextual zones). Low precision values in reading are due to the preferred reading location be close to image edges where most parts of person body are frequently outside the camera field of view.

Using the event monitoring system as input we have devised a behavioral classification model for the automatic assessment of participant cognitive health and autonomy level. Besides to event data the model also uses fine-grained data about person gait attributes (e.g., stride-length, cadence, etc), obtained by a RGBD-based algorithm for gait analysis also developed in STARS team. Briefly, the event monitoring system supports the doctor by automatically annotating the patient daily living activities and assessing his/her gait parameters in a quantitative way, and the behavioral model performs the classification of participant's dementia and autonomy levels as a complement for standard psychometric scales for autonomy. We achieved an average accuracy of 83.67 % at the prediction of patient autonomy (poor, mediocre, good), and of 73.46 % for cognitive level class (healthy, memory cognitive impairment - MCI, alzheimer's disease), all models using a Naïve Bayes classifier. The results suggest that the behavioral classification model using automatically detected events outperforms the same model using events manually annotated by domain experts (81 %). On the contrary, the model using annotated data still outperforms the automated detection at dementia classification (79.46 %). Results indicate it is easier to predict the autonomy level than the Dementia, since the latter may be seen as the cause /source and the first its consequences. Deciding whether a decay on cognitive abilities relates to normal aging or early MCI or a given mild cognitive decay is an early symptom of Alzheimer's disease or a severe case of MCI is also a open-problem for medical community. Future work will focus on investigating whether the remaining performance to achieve is related to the performance failures of the underling event monitoring system, to important behavioral aspects still not covered by the behavioral model, or even to the inherently ambiguous nature of the dementia classes.

Table 6. Event Monitoring Performance

Physical Tasks	Recall	Precision
Single Task	100%	88%
Dual Task	100%	98%
IADLs	Recall	Precision
Preparing drug box	87%	93%
Watering plant	80%	63%
Reading	60%	88%
Prepare drink	90%	68%
Talk on phone	89%	89%

We have also started the evaluation of the event monitoring system in Nursing home scenario passing from a minute time-scale to hours. A first participant was monitored with two RGBD sensors, one for bed-related events (sleep, bed exits) and one for living room and daily living activity events for 14 days. Preliminary results are 80 % for entering in bed and 100 % for bed exit in set of 6 events of each class in 13 hours monitoring (6 pm - 7am). Figure 28 illustrates the detection of restroom usage during the night. The automatic monitoring of participant activities during night is an important contribution to medical/nursing staff as wandering behavior at night is a common cause of accident in older people population. For instance, detecting whether a bed-exit during the night will be followed by a restroom visit or a bedroom exit plays a significant role at predicting a possibly dangerous situation.

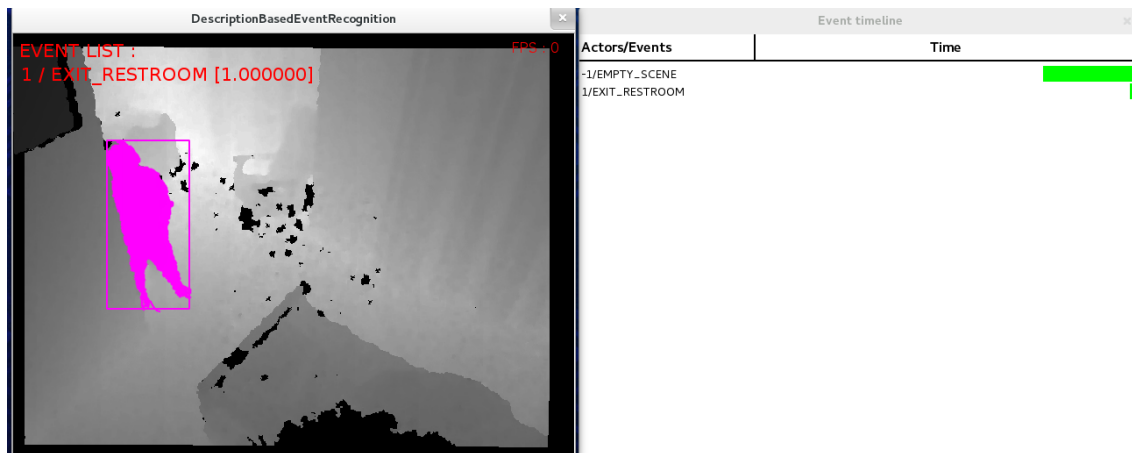


Figure 28. Long-term monitoring of Older People: Sleep-related events

Two papers are envisaged to report the results of this year research to scientific community, one describing the new version of the event monitoring system, and a second one for the developed behavioral classification model. As a publication of this year we highlight the paper in partnership with Alexandra Konig and Philippe Robert - entitled Validation of an automatic video monitoring system for the detection of instrumental activities of daily living in dementia patients - in the Journal of Alzheimer disease where we summarize the results of the validation of our event monitoring system for the recognition of activities of daily living of participants of Alzheimer's disease study.

## 6.20. Uncertainty Modeling Framework for Constraint-based Event Detection in Vision Systems

**Participants:** Carlos F. Crispim-Junior, François Brémond.

**keywords:** description-based activity recognition, uncertainty modeling, vision system, older people

Event detection has advanced significantly in the past decades relying on pixel- and feature-level representations of video-clips. Although effective, those representations have difficulty on incorporating scene semantics. Alternatively, ontology and description-based approaches for event modeling can explicitly embed scene semantics, but the deterministic nature of such languages is susceptible to noise from underlying components of vision systems. We have developed a probabilistic framework to handle uncertainty on our constraint-based ontology framework for event detection. This task spans from elementary scenarios uncertainty handling (from low-level data and event intra-class variance) to complex scenario semantic modeling, where time ordering in between event sub-components and the effect of missing components (for instance, due to miss-detection) plays a significant role.

Preliminary results of this work have been published in [40], where the presented formalism for elementary event (scenario) uncertainty handling is evaluated on the detection of activities of daily living of participants of the Alzheimer’s disease study of Nice hospital using the newest version of our vision system using a RGB-D sensor (Kinect®, Microsoft®) as input. Two evaluations have been carried out: the first one, (a 3-fold cross-validation) focuses on elementary scenario constraint modeling and recognition, and the second one was devoted to complex scenario recognition following a semi-probabilistic approach (n:45).

Table 7 presents the performance of the uncertainty modeling framework on elementary scenario (primitive state) detection for  $N : 10$  participants; 15 min. each; Total : 150 min. The 3-fold cross-validation scheme (n:10 participants) is employed for constraint probabilistic distribution learning and event detection evaluation on 10 RGB-D recordings of participants of the Nice hospital clinical protocol for Alzheimer’s disease study. “Crisp” term stands for our deterministic constraint-based ontology language for event modeling. Results are reported as the average performance on the crisp and uncertainty frameworks on the validation sets. Results confirm that the uncertainty modeling improves the detection of elementary scenarios in recall (e.g., In zone phone: 84 to 100 %) and precision indices (e.g., In zone Reading: 54.5 to 85.7%).

Table 7. Framework Performance on Elementary Scenario Detection on a 3-fold-cross-validation scheme

IADL	Crisp		Uncertainty	
	Rec.	Prec.	Rec.	Prec.
In zone Pharmacy	100.0	71.4	100	<b>83.3</b>
In zone Phone	84.0	95.45	<b>100.0</b>	<b>100.0</b>
In zone Plant	100.0	81.8	100.0	81.8
In zone Tea	93.3	<b>77.7</b>	93.3	73.7
In zone Read	75.0	54.5	75.0	<b>85.7</b>

Table 8 presents the performance of the proposed framework on Composite Event Detection for  $N : 45$  participants; 15 min. each; Total : 675min. Here a hybrid strategy is adopted where the uncertainty modeling is used for elementary scenarios and the crisp constraint-based framework is used for composite event modeling. Results show improvement on recall index of event detection performance, but the uncertainty framework performance on precision index is still worse than the crisp approach. The latter performance may be attributed to the crisp constraints that did not have their uncertainty addressed yet.

Future work will focus on modeling complex scenario constraints such as time ordering and missing components, and on extending the set of low-level uncertainties which are addressed. Moreover, we have been conducting a joint work with partners of Dem@care project to evaluate the uncertainty framework for multiple sensor fusion at decision level. Currently, processed data from different visual modalities (standard RGB, RGBD, and wearable cameras) have been gathered for 17 participants of Nice hospital pilot@lab, and preliminary results are expected for the first semester of 2015.

Table 8. Framework Performance on Composite Event Detection Level

IADL	Crisp		Uncertainty	
	Rec.	Prec.	Rec.	Prec.
Talk on Phone	88.76	89.77	<b>88.76</b>	85.86
Preparing Tea/Coffee	81.42	40.36	<b>92.85</b>	<b>55.08</b>
Using Pharmacy Basket	87.75	95.65	<b>89.79</b>	<b>97.77</b>
Watering plant	78.57	<b>84.61</b>	<b>100.0</b>	28.86

## 6.21. Assisted Serious Game for Older People

**Participants:** Minh Khue Phan Tran, François Brémond, Philippe Robert.

**keywords:** interactive system, elderly people, serious game

A system able to interact with older people has been recently devised. The system consists of two parts: Recognition and Interaction. Recognition part, requiring an Asus Xtion Pro Live Camera, consists in observing the scene to decide when is the best moment to interact with users. Afterwards, the Interactive system tries to engage the patient via an interface and through Microsoft Kinect Camera, the patient can interact with the interface using voice or gesture. The interface is designed with Unity 3D game engine (see figure 29).

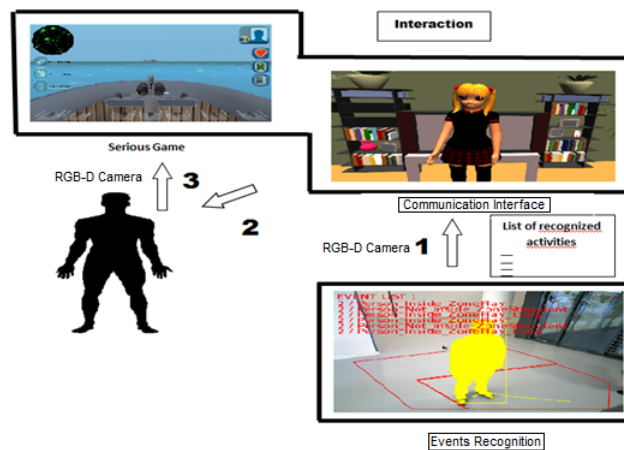


Figure 29. Functional diagram of our interactive system

An experiment was conducted in a memory center for older people, Institut Claude Pompidou in Nice, in order to test different functionalities of the system. Here, participants can experiment the system in a private room (see figure 30) equipped with a large screen and can start the game without having to use devices (mouse, keyboard). The "best moments" to interact with participants are defined when they stay more than 5 seconds in front of the screen. Once these moments are recognized, the interface of Interactive part is called. The avatar indicates the place to be for playing and starts the game.

19 older people have participated to the experiment. 16 succeeded to follow the indications of avatar up to the start of the game. Most of them have appreciated the interaction with the avatar. Even 13 of them have continued to play the second game after suggestion of the avatar.

Future work aims at looking at other indicators (behavior, gaze) that the system can rely on to improve user's interaction.



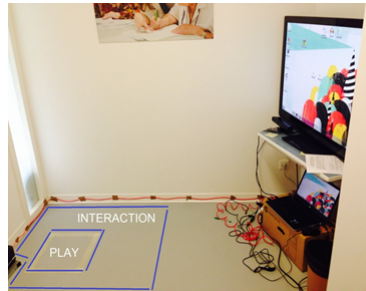


Figure 30. Experimental private room

## 6.22. Enhancing Pre-defined Event Models Using Unsupervised Learning

**Participants:** Serhan Coşar, François Brémond.

**keywords:** Pre-defined activity models, unsupervised learning, tailoring activity models

In this work, we have developed a new approach to recognize human activities from videos, given models that are learned in an unsupervised way and that can take advantage of *a priori* knowledge provided by an expert of the application domain. The description-based methods use pre-defined models and rules to recognize concrete events. But, if the data has unstructured nature, such as daily activities of people, the models cannot handle the variability in data (e.g., the way of preparing meal is person dependent).

In order to overcome this drawback, we have combined the description-based method in [66] with an unsupervised activity learning framework, as presented in Figure 31. We have created a mutual knowledge loop system, in which both frameworks are combined in a way to compensate their individual limitations. In [66], scene regions are pre-defined and the activity models are created via defining an expected duration value (e.g., 2 seconds) and a posture type (e.g., standing) by hand. Thus, these hand-crafted models fail to cover the variability in data and require an update by experts whenever the scene or person changes. To automatically define these parameters, we utilize the unsupervised activity recognition framework. The unsupervised approach first learns scene regions (zones) in the scene using trajectory information and then, it learns the duration and posture distribution for each zone. By matching the pre-defined zones with learned zones, we connect the learned parameter distributions with hand-crafted models.

The knowledge is passed in a loopy way from one framework to another one. By knowledge we mean: (i) the geometric information and scene semantics of the description-based system are used to label the zones that are learned in an unsupervised way, (ii) the activity models that are learned in an unsupervised way are used to tune the parameters (i.e. tailoring) in the activity models of the description-based framework. It is assumed that the person detection and tracking are already performed and we have the trajectory information of people in the scene beforehand.

We have tested the performance of the knowledge-loop based framework on two datasets: i) Hospital-RGB, ii) Hospital-RGBD. Each dataset contains one person performing everyday activities in a hospital room. The activities considered in the datasets are "watching TV", "preparing tea", "answering phone", "reading newspaper/magazine", "watering plant", "organizing the prescribed drugs", "writing a check at the office desk" and "checking bus routes in a bus map". Each person is recorded using RGB and RGBD cameras of  $640 \times 480$  pixels of resolution. RGB dataset consists of 41 videos and RGBD dataset contains 27 videos. For each person, video lasts approximately 15 minutes.



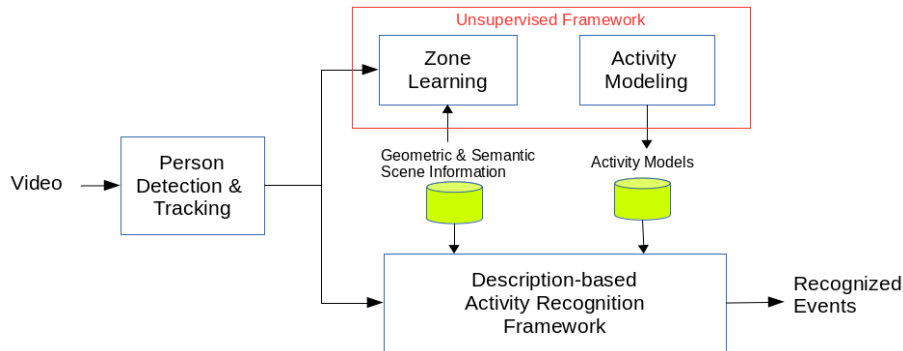


Figure 31. The flow diagram of the framework that combines pre-defined activity models with unsupervised models.

The performance of the approach in [66] with hand-crafted models and our approach with learned models for Hospital-RGB and Hospital-RGBD datasets are presented in Table 9 and in Table 10. The results have been partially presented in Ellomietcv2014 (**waiting hal acceptance**). It can be clearly seen that updating the constraints in activity models using data learned by the unsupervised approach enables us to detect activities missed by the pre-defined models. For "watching TV" and "using pharmacy basket" activities in RGB dataset and "answering phone" and "preparing tea" activities in RGBD dataset, there is increase in false positive rates. The reason is that, for some activities, the duration and posture distributions learned by the unsupervised approach can be inaccurate because of other actions occurring inside a zone (e.g., a person standing inside tea zone and reading). For this reason, the constraints updated in activity models are too wide and other activities that occur inside the zone are also detected. Despite the small increase of false positives in some activities, we have increased the true positive rates and obtained sensitivity rates around 90% and 87% on average in RGB and RGBD datasets, respectively, and precision rates around 81% on average in RGBD dataset. Thanks to the distributions learned for time duration and posture, we can enhance the activity models that are manually defined in the description-based, and thereby detect missed events.

Table 9. The activity recognition results of the description-based approach and our knowledge-loop based approach for the Hospital-RGB. The bold values represent the best result for each activity class.

ADLs	Hand-crafted Models		Unsupervised Models	
	Sensitivity (%)	Precision (%)	Sensitivity (%)	Precision (%)
Answering Phone	70	82.35	<b>95</b>	<b>90.47</b>
Watching TV	84.61	78.57	<b>100</b>	54.16
Using Office Desk	91.67	47.82	91.67	<b>52.38</b>
Preparing Tea	<b>80.95</b>	70.83	76.19	<b>80</b>
Using Phar. Basket	100	<b>90.90</b>	100	76.92
Watering Plant	100	81.81	88.89	<b>88.89</b>
Reading	45.46	83.34	<b>81.82</b>	<b>90</b>
TOTAL	81.81	<b>76.52</b>	<b>90.5</b>	76.11

### 6.23. Using Dense Trajectories to Enhance Unsupervised Action Discovery

**Participants:** Farhood Negin, Serhan Coşar, François Brémond.

**keywords:** zone learning, action descriptors, dense trajectories, supervised action recognition, unsupervised activity recognition

Table 10. The activity recognition results of the description-based approach and our knowledge-loop based approach for the Hospital-RGBD datasets. The bold values represent the best result for each activity class.

	Hand-crafted Models		Unsupervised Models	
	Sensitivity (%)	Precision (%)	Sensitivity (%)	Precision (%)
ADLs				
Answering Phone	80	<b>100</b>	<b>84.21</b>	88.89
Watching TV	55.56	45.46	<b>77.78</b>	<b>58.34</b>
Preparing Tea	100	<b>73.68</b>	92.85	65
Using Phar. Basket	100	90	100	<b>100</b>
Watering Plant	40	66.67	<b>83.34</b>	<b>71.42</b>
Reading	<b>100</b>	66.67	71.42	<b>83.34</b>
Using Bus Map	50	71.42	<b>100</b>	<b>100</b>
TOTAL	75.07	73.41	<b>87.08</b>	<b>80.99</b>

The main purpose in this work is to monitor older people in an unstructured scene (e.g., home) and to recognize the types of activities they perform. We have extended the work in Ellomietcv2014 that was basically an unsupervised method to learn behavioral patterns of individuals without restraining subjects to follow a predefined activity model. The main concern in previous work is to find different zones in the scene where activities take place (scene topology) by employing trajectory information provided by tracking algorithm. The previous work in Ellomietcv2014 (**waiting hal acceptance**) proposes a Hierarchical Activity learning Model (HAM) to learn activities based on previously identified topologies. The current work examines the same potential while first, incorporating image descriptors [93] in a bag-of-word representation to differentiate actions in a supervised manner and second, combining the two approaches (supervised and unsupervised) to provide clues about actions inside each zone by classifying retrieved descriptors using a classifier.

Recently, dense trajectories are widely used for action recognition and have been shown state-of-the-art performance [93]. For the purpose of the current work, we use HOG and HOF descriptors for supervised action recognition. Figure 32 shows a general description of the supervised framework. For the learning phase, the dense trajectories are extracted from input images coming from RGBD camera. Following Ellomietcv2014, three-level topology of the scene is constructed by trajectory information coming from tracking algorithm [62]. The topology is used to split input video stream into chunks by checking where the person is with respect to the learned zones. Then, for every video chunk, dense descriptors are extracted and stored. A codebook representation is obtained by applying a k-means clustering algorithm on the whole set of extracted features. Next, the action histograms are calculated by employing the codebook. A SVM classifier is trained and stored to use in test phase via calculated histograms.



Figure 32. Flow diagram for supervised action recognition.

In recognition phase, we similarly split the test videos by comparing each trajectory point with learned topologies, extract the descriptor for each split, and the histograms are calculated via k-NN using the codebook generated in learning phase. Then, the histograms are classified using the trained SVM classifier and resulting labels are evaluated by comparing with the ground truth.

We have assessed the performance of the supervised activity recognition framework using 183 video splits of 26 subjects. We divided the video dataset to training and testing groups. Training set includes 93 videos of 15 subjects and the test set includes 90 videos of 11 subjects. Notice that the number of videos is counted after splitting process has been done on input data. We used the videos recorded from CHU Nice hospital while real patients are visiting their doctors and are asked to perform several activities in specified locations of the room. The activities we considered in our tests include: “preparing tea”, “watching TV”, “using phone”, “reading on chair”, “using pharmacy”, and “using bus map”. For RGB-D camera, we have used the person detection algorithm in [79] and tracking algorithm in [62]. The classification results for using HOG and HOF descriptors and corresponding confusion matrices are depicted in Table 11 and in Table 12. For SVM classifier, we used RBF kernel.

Table 11. Confusion matrix for recognition results for HOG descriptor

Activity Names	1	2	3	4	5	6
1 Watching TV	11	0	0	0	0	0
2 Preparing Tea	0	18	0	0	0	0
3 Reading in Chair	1	0	10	0	0	0
4 Using Bus Map	0	0	0	14	0	0
5 Using Pharmacy Basket	0	0	0	0	10	0
6 Using Phone	0	0	0	0	0	25
Total	98.89%					

Table 12. Confusion matrix for recognition results for HOF descriptor

Activity Names	1	2	3	4	5	6
1 Watching TV	4	1	1	4	0	1
2 Preparing Tea	0	5	0	5	0	8
3 Reading in Chair	1	0	2	4	0	4
4 Using Bus Map	0	0	0	13	0	1
5 Using Pharmacy Basket	1	0	0	0	9	0
6 Using Phone	0	1	1	5	0	19
Total	57.78%					

As a future work, we are going to benefit from the action descriptors to discriminate different activities occurring in the same zone.

## 6.24. Abnormal Event Detection in Videos and Group Behavior Analysis

**Participants:** Giuseppe Donatiello, Vania Bogorny, Serhan Cosar, Luis Campos Alvares, Carolina Garate, François Brémond.

**keywords:** activity recognition, abnormal events, group behavior analysis, trajectory clustering

This work addresses two different issues: (i) abnormal event detection and (ii) group behavior analysis in videos.

### 6.24.1. Abnormal Event Detection

For abnormal event detection we are proposing a fused approach that combines trajectory-based and pixel-based analysis. In this work we first discover the activity zones based on object trajectories, and we investigate abnormal events considering objects that move in wrong direction and/or with abnormal speed. Second, inside each zone we extract dense tracklets and using the clustering technique we discover different types of actions, and are able to distinguish between normal and abnormal actions inside each zone.

While existing approaches for abnormal behavior detection do either use trajectory based or pixel based methods, we propose a fused solution which can detect simple abnormal behavior based on speed and direction, as well as more complex behavior as abnormal activities. In a first step we automatically learn the zones of the scene where most activities occur, by taking as input the trajectories of detected mobiles, analyzing then statistical information of each mobile in each zone (speed and direction), through the use of a scale-resolution analysis. This approach implies a considerable complexity decrease of having huge data set and then an extensive impact of the algorithm speed, without losing useful information. Figure 33 shows an example of this first part.



Figure 33. Example of Trajectories (left), Trajectories over the grid (center), that represents a given scale resolution, Zones discovering (right)

The next step concerns a pixel based analysis inside each zone. This step takes as input each zone computed in the previous step and the bounding box of the object trajectories, and extracts action descriptors inside the bounding box of each object trajectory inside the zone. With this step we obtain the different body movements of each detected mobile inside a zone. By clustering the body motions and using Bags of Words, we detect different types of abnormal activities inside each zone. Figure 34 shows an example of what just mentioned.

The last step of our approach is a clustering operation of all information gathered in the previous two steps, that is for each mobile, speed, direction, and body movements-actions in each zone are applied to discriminate between different types of abnormal behavior in the scene. A flow diagram of our approach is presented in Figure 35.

We have tested our approach on several real videos recorded. We show with experiments on two open datasets that our approach is able to detect several types of abnormal behavior.

### 6.24.2. Group Behavior Analysis

Group behavior analysis is focused on the extraction of groups based on object trajectories and the analysis is performed over the dense tracklets, computed for the groups bounding boxes. From the analysis of the dense tracklets we detect different levels of agitation. These works are ongoing and have not yet been published.



Figure 34. Tracked object

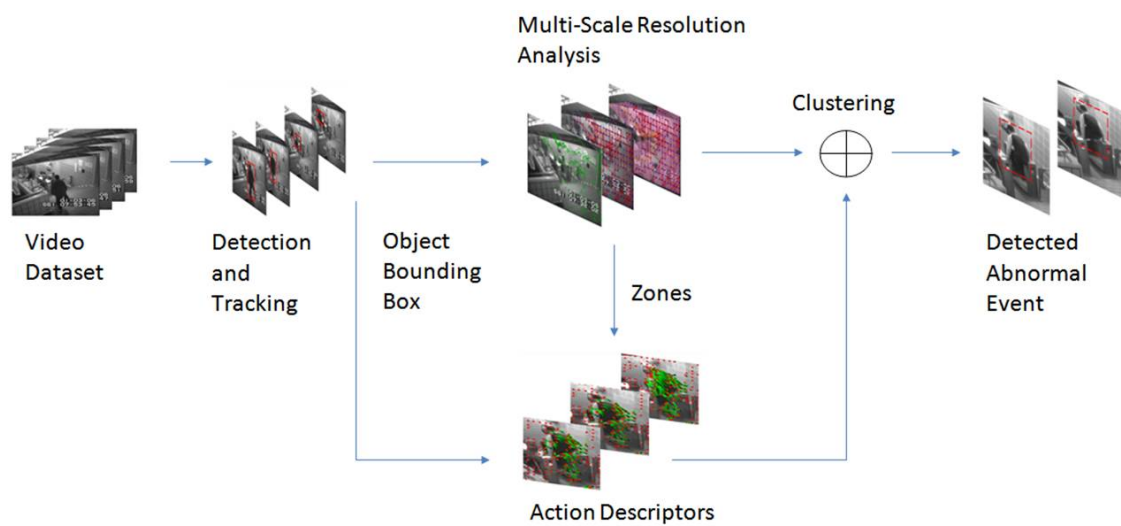


Figure 35. Flow Diagram of our approach

## 6.25. Model-Driven Engineering for Activity Recognition Systems

**Participants:** Sabine Moisan, Jean-Paul Rigault, Luis Emiliano Sanchez.

We continue to explore the applicability of model driven engineering (MDE) to activity recognition systems. Of course, setting up a complete methodology is a long term objective.

### 6.25.1. Feature Models

Features models are convenient representations of system variability but the drawback is a risk of combinatorial explosion of the number of possible configurations. Hence we have extended feature models with quality attributes and associated metrics to facilitate the choice of an optimal configuration, at deployment as well as at run time. We have proposed several strategies and heuristics offering different properties regarding optimality of results and execution efficiency [41].

This year we have conducted further experiments to evaluate the optimization algorithm and the metrics. In particular, we studied the prediction accuracy of the additive metrics for estimating two properties of interest: frame processing time and reconfiguration time. The goal was to compare predicted against measured properties of a running system. We used a simple video chain implemented with OpenCV components (acquisition, filtering, various detections, and visualisation) and we tested it on a video sample of 48s (i.e., about 1350 frames). We defined a feature model for this chain, which exhibits 14 valid configurations.

We first computed the properties of each component in isolation (based on a set of repetitive measurements), then we measured the actual frame processing time and reconfiguration time, and finally we compared the estimated and actual values.

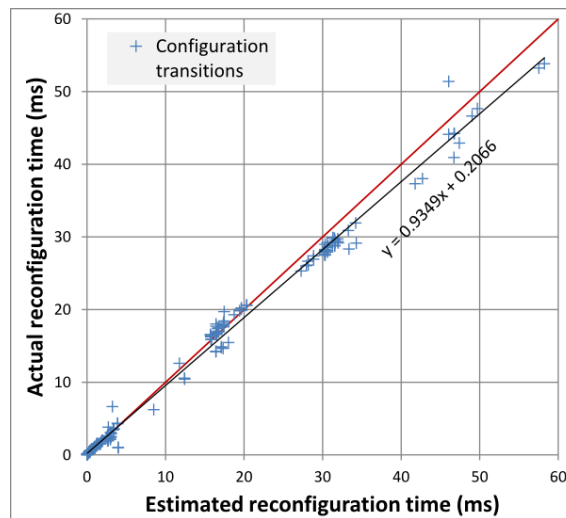


Figure 36. Measured and estimated reconfiguration time. Black line represents regression curve and red line the ideal trend

Figure 36 displays the results for reconfiguration time. In our experiment, we have a total of 210 transitions over a set of 15 valid configurations (14 system configuration + one for system shutdown). For frame processing time, we achieved an accuracy of 96.7% on average, and for reconfiguration time the accuracy was between 90.5% and 87.6%.



Introducing metrics in feature models is a precious help to reduce the huge set of valid configurations after a dynamic context change and to provide a real time selection of an appropriate running configuration. However, more evaluation remains to be done on other runtime properties and for other video chains.

### 6.25.2. Configuration Adaptation at Run Time

To react to environment changes we favor the “model at run-time” approach. Our current prototype resorts to tools and technologies which were readily available. This made possible a proof of concepts. However, this induces several redundant representations of the same data, consistency problems, coexistence of several formalisms or programming languages, and superfluous back and forth inter-module communications. This year we started to design a more homogeneous and better integrated prototype. The two key points are, first, a component management framework, second, a feature model management tool.

This year, we have continued to develop our OSGi-like component framework, but more adapted to real time and compatible with our extended feature models. Concerning feature model management, we started to study a FAMILIAR replacement that integrates smoothly into the C++ video analysis chain instead of being a separate (Java) tool. Moreover, the new tool should cope with our feature extensions (e.g., quality attributes).

## 6.26. Scenario Analysis Module

**Participants:** Annie Ressousche, Sabine Moisan, Jean-Paul Rigault, Daniel Gaffé, Omar Abdalla.

**Keywords:** Synchronous Modelling, Model checking, Mealy machine, Cognitive systems.

To generate activity recognition systems we supply a scenario analysis module (SAM) to express and recognize complex events from primitive events generated by SUP or other sensors. The purpose of this research axis is to offer a generic tool to express and recognize activities. Genericity means that the tool should accommodate any kind of activities and be easily specialized for a particular framework. In practice, we propose a concrete language to specify activities in the form of a set of scenarios with temporal constraints between scenarios. This language allows domain experts to describe their own scenario models. To recognize instances of these models, we consider the activity descriptions as synchronous reactive systems [80] and we adapt usual techniques of synchronous modelling approach to express scenario behaviours. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario model.

Setting up our tools on top of an existing language such as LUSTRE was convenient for rapid prototyping. However, it appeared delicate for efficiency reasons on the one hand, but also because it is a closed environment, difficult to customize. Hence we developed our own language LE and its environment CLEM (see section 6.27). This year, we focus on the expression of scenario models in CLEM through the internship of Omar Adballa [51] and we define in CLEM a specific back end to generate recognition engines (see figure 37). However, mastering all aspects of this environment will allow the user scenario description language to rely directly on the semantics of LE and not on its syntax. This reduces the number of necessary translations.

Currently, SAM implements an “exact” algorithm in the sense that it generates, at each instant, *all* possible scenario instances although many of them will freeze, still holding system resources. We have started scalability studies to evaluate the risk of combinatorial explosion. In parallel we enriched the synchronous scenario descriptions to reduce the number of generated scenario instances as well as the number of instances to awake at each instant. We are currently modifying our recognition engine generator to take advantage of this supplementary information.

## 6.27. The Clem Workflow

**Participants:** Annie Ressousche, Daniel Gaffé, Mohamed Bouatira, Ines Sarray.

**Keywords:** Synchronous languages, Synchronous Modelling, 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 37) 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 focus on the improvement of both LE language and compiler concerning data handling and in the generation of back-ends required by other research axis of the team (see 6.26 and 6.28). We also improve the design of a new simulator for LE programs which integrates our new approach.

First, synchronous language semantics usually characterizes each output and local signal status (as present or absent) according to input signal status. To reach our goal, we defined a semantics that translates LE programs into equation systems. This semantics bears and grows richer the knowledge about signals and is never in contradiction with previous deduction (this property is called constructiveness). In such an approach, causality turns out to be a scheduling evaluation problem. We need to determine all the partial orders of equation systems and to compute them, we consider a 4-valued algebra to characterize the knowledge of signal status (unknown, present, absent, overknown). Last year, we chose an algebra which is a bilattice and we showed that it is well suited to solve our problem. To compute the partial orders of equation systems, we introduced two ways : a CPM (Critical Path Method) like algorithm, efficient and a "fix point" approach that allows us to show that we can compute partial orders locally and in an incremental way, or globally (thanks to the uniqueness of fix points). We introduced "abstract" equation systems and our method allows us to compute their partial orders. We defined a new intermediate format LEA (see figure 37) to record these sorted abstract equation systems and they will be expanded later in concrete equation systems with a refinement operation. We apply this technique to the compilation of Grafset language. Our work is under publication in two journal papers.

In CLEM, we added types and data a few years ago and this year we complete this addition, we now can express automata where control and output signals are valued. From last year, we rely on CLEM both to design SAM (see section 6.26) and to perform validation in a component-based middleware (see section 6.28). To this aim, we generate now two specific output formats dedicated to these applications [54]

Finally, 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. This year, we complete a simulator begun last year but which did not integrate the data part of the language. The simulator GUI has been designed again in Qt and the simulator takes into account the values carried by signals. This work has been done by Mohamed Bouatira during his internship.

## 6.28. Multiple Services for Device Adaptive Platform for Scenario Recognition

**Participants:** Annie Ressouche, Daniel Gaffé, Ines Sarray, Jean-Yves Tigli.

**Keywords:** Synchronous Modelling, Model checking, Mealy machine, Ubiquitous Computing.

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 5.1) with a service oriented middleware approach dealing with dynamic evolutions of system infrastructure. The Rainbow team (Nice-Sophia Antipolis University) proposes a component-based adaptive middleware (WComp [91], [90], [73]) to dynamically adapt and recompose assemblies of components. These operations must obey the "usage contract" of components. The existing approaches don't really ensure that this usage contract is not violated during application design. Only a formal analysis of the component behaviour models associated with a well sound modelling of composition operation may guarantee the respect of the usage contract.

The approach we adopted introduces in a main assembly, a synchronous component for each sub assembly connected with a critical component. This additional component implements a behavioral model of the critical component and model checking techniques apply to verify safety properties concerning this critical component. Thus, we consider that the critical component is validated.

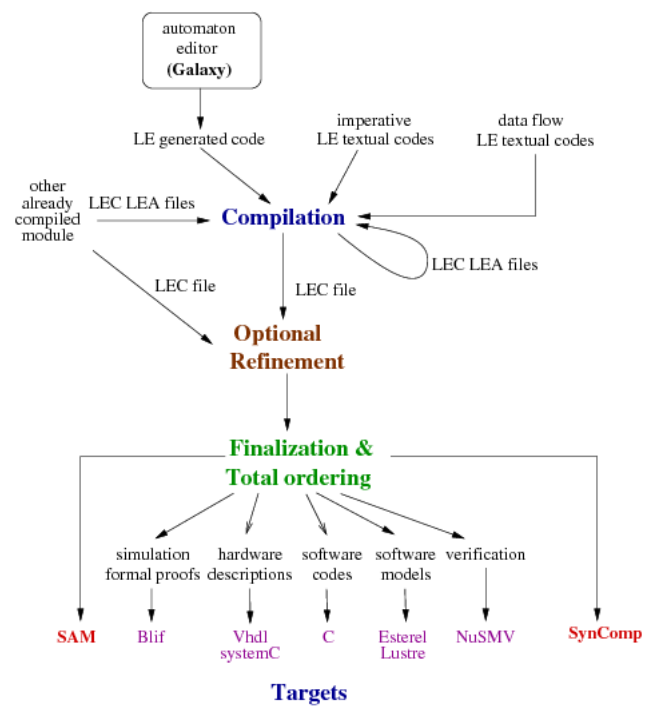


Figure 37. The CLEM toolkit

In [83], [82], we showed that an efficient means to define the synchronous components which allow to validate critical component behaviours, is to specify them with Mealy machines. Previously, we used a classical synchronous language (Lustre) to specify synchronous components, but the integration of the synchronous component code into WComp was not straightforward because Lustre compiler is not opened and cannot integrate new target code needed by WComp. This year, we rely on CLEM to design synchronous monitor as CLEM automata and we extend CLEM to generate automatically the internal code of WComp(C#).

On another hand, for some critical components, we can be led to introduce several synchronous monitors, each of them being related to a sub assembly. Then, some outputs of these synchronous monitors can be linked to the same input of a critical component. To face this problem, we introduced in [83], [82] a composition under constraints operation, which composed all the synchronous monitors linked to a critical component according to a set of constraints telling us how the respective outputs of monitors which access the same input are combined. We proved that this operation preserved safety properties, however it cannot ensure adaptivity and incrementality. This year, we have introduced a new way to compose synchronous monitors. We introduce an additional monitor (called constrain monitor) which express as a CLEM Mealy machine (set of equations) the constraints and we perform a usual parallel composition of all the monitors. Moreover, we define a syntactic mean (language DCL) to describe constraints in a generic way and we can derive automatically the constraint monitor for each specific application. In such a setting, we are able to manage the appearance and the disappearance of synchronous monitors.

Moreover, in middleware as WComp, communication is asynchronous while synchronous monitors respect the synchrony paradigm. Thus, we need to introduce in the middleware a means to transform asynchronous events into synchronous entries for synchronous monitors and the opposite to connect again the synchronous events into asynchronous design. To this aim, we introduced in WComp, specific components that receive asynchronous events and generate synchronous ones. Indeed, the part of this component is to decide how asynchronous events will be gather into a synchronous instant (logical time). To this aim, the user can choose between different politics. Then, all the events composing an instant are serialized and deserialized by the synchronous monitor. The desynchronisation operation performs the opposite way[54].

## WILLOW Project-Team

# 6. New Results

## 6.1. Highlights of the Year

- J. Sivic started ERC project LEAP (2014-2018).
- J. Sivic serves as a Program Chair for International Conference on Computer Vision, Santiago, Chile, 2015

## 6.2. 3D object and scene modeling, analysis, and retrieval

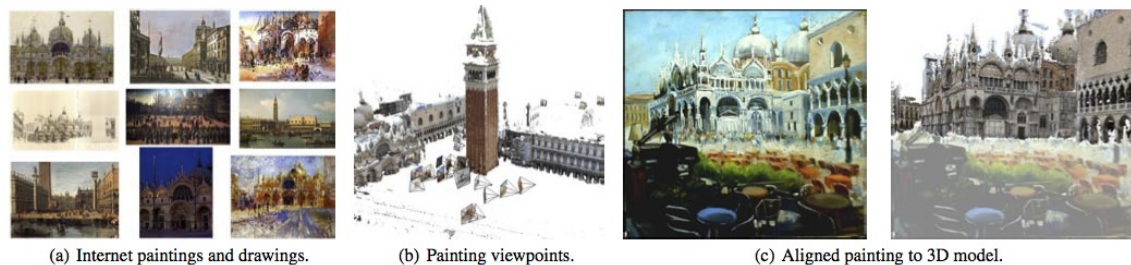


Figure 1. Our system automatically aligns and recovers the viewpoint of paintings, drawings, and historical photographs to a 3D model of an architectural site.

### 6.2.1. Painting-to-3D Model Alignment Via Discriminative Visual Elements

**Participants:** Mathieu Aubry, Bryan Russell [Intel Labs], Josef Sivic.

In this work we describe a technique that can reliably align arbitrary 2D depictions of an architectural site, including drawings, paintings and historical photographs, with a 3D model of the site. This is a tremendously difficult task as the appearance and scene structure in the 2D depictions can be very different from the appearance and geometry of the 3D model, e.g., due to the specific rendering style, drawing error, age, lighting or change of seasons. In addition, we face a hard search problem: the number of possible alignments of the painting to a large 3D model, such as a partial reconstruction of a city, is huge. To address these issues, we develop a new compact representation of complex 3D scenes. The 3D model of the scene is represented by a small set of discriminative visual elements that are automatically learnt from rendered views. Similar to object detection, the set of visual elements, as well as the weights of individual features for each element, are learnt in a discriminative fashion. We show that the learnt visual elements are reliably matched in 2D depictions of the scene despite large variations in rendering style (e.g. watercolor, sketch, historical photograph) and structural changes (e.g. missing scene parts, large occluders) of the scene. We demonstrate an application of the proposed approach to automatic re-photography to find an approximate viewpoint of historical paintings and photographs with respect to a 3D model of the site. The proposed alignment procedure is validated via a human user study on a new database of paintings and sketches spanning several sites. The results demonstrate that our algorithm produces significantly better alignments than several baseline methods. This work has been published at ACM Transactions on Graphics 2014 [3] and its extension has appeared at RFIA 2014 [17]. The problem addressed in this work is illustrated in Figure 1 and example results are shown in Figure 2.



Figure 2. Example alignments of non-photographic depictions to 3D models. Notice that we are able to align depictions rendered in different styles and having a variety of viewpoints with respect to the 3D models.

### 6.2.2. Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models

**Participants:** Mathieu Aubry, Bryan Russell [Intel labs], Alyosha Efros [UC Berkeley], Josef Sivic.

This work poses object category detection in images as a type of 2D-to-3D alignment problem, utilizing the large quantities of 3D CAD models that have been made publicly available online. Using the “chair” class as a running example, we propose an exemplar-based 3D category representation, which can explicitly model chairs of different styles as well as the large variation in viewpoint. We develop an approach to establish part-based correspondences between 3D CAD models and real photographs. This is achieved by (i) representing each 3D model using a set of view-dependent mid-level visual elements learned from synthesized views in a discriminative fashion, (ii) carefully calibrating the individual element detectors on a common dataset of negative images, and (iii) matching visual elements to the test image allowing for small mutual deformations but preserving the viewpoint and style constraints. We demonstrate the ability of our system to align 3D models with 2D objects in the challenging PASCAL VOC images, which depict a wide variety of chairs in complex scenes. This work has been published at CVPR 2014 [9].

### 6.2.3. Anisotropic Laplace-Beltrami Operators for Shape Analysis

**Participants:** Mathieu Andreux [TUM], Emanuele Rodola [TUM], Mathieu Aubry, Daniel Cremers [TUM].

This work introduces an anisotropic Laplace-Beltrami operator for shape analysis. While keeping useful properties of the standard Laplace-Beltrami operator, it introduces variability in the directions of principal curvature, giving rise to a more intuitive and semantically meaningful diffusion process. Although the benefits of anisotropic diffusion have already been noted in the area of mesh processing (e.g. surface regularization), focusing on the Laplacian itself, rather than on the diffusion process it induces, opens the possibility to effectively replace the omnipresent Laplace-Beltrami operator in many shape analysis methods. After providing a mathematical formulation and analysis of this new operator, we derive a practical implementation on discrete meshes. Further, we demonstrate the effectiveness of our new operator when employed in conjunction with different methods for shape segmentation and matching. This work has been published at the Sixth Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA) 2014 [8].

### 6.2.4. Trinocular Geometry Revisited

**Participants:** Jean Ponce, Martial Hebert [CMU].



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. This work has been published at CVPR 2014 [15].

### 6.2.5. On Image Contours of Projective Shapes

**Participants:** Jean Ponce, Martial Hebert [CMU].

This work revisits classical properties of the outlines of solid shapes bounded by smooth surfaces, and shows that they can be established in a purely projective setting, without appealing to Euclidean measurements such as normals or curvatures. In particular, we give new synthetic proofs of Koenderink's famous theorem on convexities and concavities of the image contour, and of the fact that the rim turns in the same direction as the viewpoint in the tangent plane at a convex point, and in the opposite direction at a hyperbolic point. This suggests that projective geometry should not be viewed merely as an analytical device for linearizing calculations (its main role in structure from motion), but as the proper framework for studying the relation between solid shape and its perspective projections. Unlike previous work in this area, the proposed approach does not require an oriented setting, nor does it rely on any choice of coordinate system or analytical considerations. This work has been published at ECCV 2014 [14].

## 6.3. Category-level object and scene recognition

### 6.3.1. Finding Matches in a Haystack: A Max-Pooling Strategy for Graph Matching in the Presence of Outliers

**Participants:** Minsu Cho, Jian Sun, Olivier Duchenne, Jean Ponce.

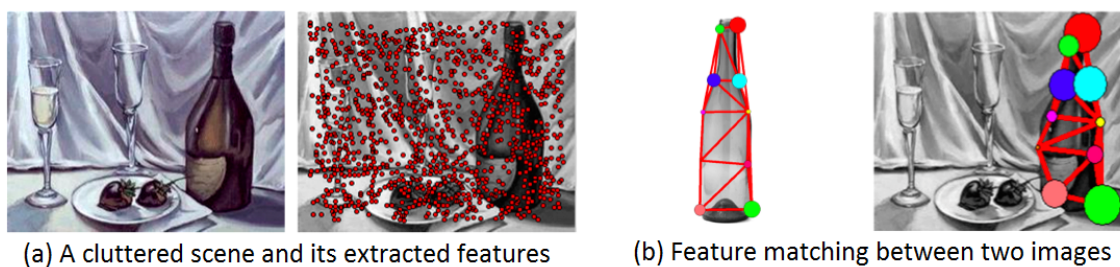


Figure 3. Feature matching in the presence of outliers. (a) In real-world scenes, background clutter often produces numerous outlier features, making it hard to find correspondences. (b) We address the issue with a max-pooling approach to graph matching. The proposed method is not only resilient to deformations but also remarkably tolerant to outliers. Each node on the left image corresponds to one with the same color on the right image, where bigger nodes represent more similar nodes. (Best viewed in color.)

A major challenge in real-world feature matching problems is to tolerate the numerous outliers arising in typical visual tasks. Variations in object appearance, shape, and structure within the same object class make it harder to distinguish inliers from outliers due to clutters. In this work, we propose a max-pooling approach to graph matching, which is not only resilient to deformations but also remarkably tolerant to outliers. The proposed algorithm evaluates each candidate match using its most promising neighbors, and gradually propagates the corresponding scores to update the neighbors. As final output, it assigns a reliable score to each match together with its supporting neighbors, thus providing contextual information for further verification. We demonstrate the robustness and utility of our method with synthetic and real image experiments. This work has been published at CVPR 2014 [11]. The proposed method and its qualitative results are illustrated in Figure 3 .

### 6.3.2. Unsupervised Object Discovery and Localization in the Wild: Part-based Matching with Bottom-up Region Proposals

**Participants:** Minsu Cho, Suha Kwak, Cordelia Schmid [Inria Lear], Jean Ponce.

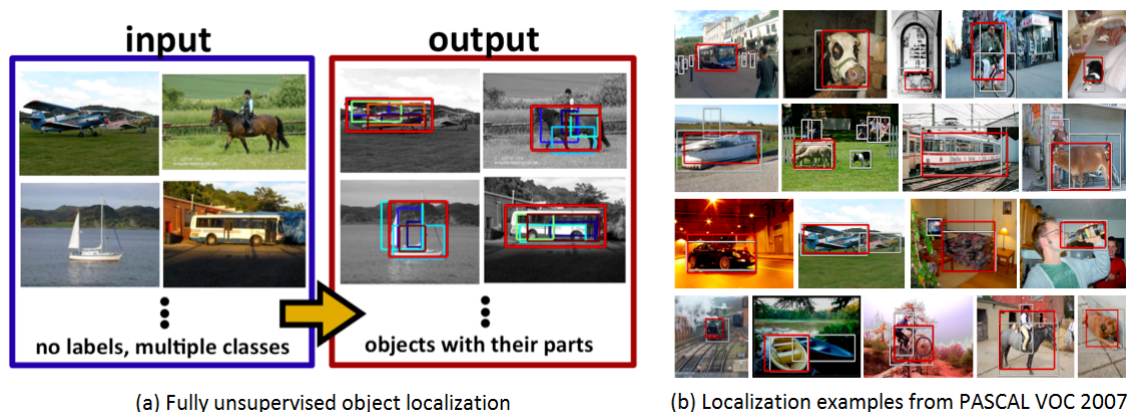


Figure 4. Unsupervised object discovery in the wild. (a) We tackle object localization in an unsupervised scenario without any types of annotations, where a given image collection may contain multiple dominant object classes and even outlier images. The proposed method discovers object instances (red bounding boxes) with their distinctive parts (smaller boxes). (b) Examples of localization on mixed-class PASCAL VOC 2007 train/val datasets are shown. Red boxes represent localized objects while white boxes are ground truth annotations. (Best viewed in color.)

This work addresses unsupervised discovery and localization of dominant objects from a noisy image collection of multiple object classes. The setting of this problem is fully unsupervised, 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 discovery and localization problem using a part-based 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 in each candidate region 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 experimental evaluations on standard benchmarks demonstrate that the proposed approach significantly outperforms the current state of the art in colocalization, and achieves robust object discovery in

challenging mixed-class datasets. This work has been submitted to CVPR 2015 [22]. The proposed method and its qualitative results are illustrated in Figure 4 .

### 6.3.3. Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks

**Participants:** Maxime Oquab, Leon Bottou [MSR New York], Ivan Laptev, Josef Sivic.

Convolutional neural networks (CNN) have recently shown outstanding image classification performance in the large-scale visual recognition challenge (ILSVRC2012). The success of CNNs is attributed to their ability to learn rich mid-level image representations as opposed to hand-designed low-level features used in other image classification methods. Learning CNNs, however, amounts to estimating millions of parameters and requires a very large number of annotated image samples. This property currently prevents application of CNNs to problems with limited training data. In this work we show how image representations learned with CNNs on large-scale annotated datasets can be efficiently transferred to other visual recognition tasks with limited amount of training data. We design a method to reuse layers trained on the ImageNet dataset to compute mid-level image representation for images in the PASCAL VOC dataset. We show that despite differences in image statistics and tasks in the two datasets, the transferred representation leads to significantly improved results for object and action classification, outperforming the current state of the art on Pascal VOC 2007 and 2012 datasets. We also show promising results for object and action localization. This work has been published at CVPR 2014 [13].

### 6.3.4. Weakly supervised object recognition with convolutional neural networks

**Participants:** Maxime Oquab, Leon Bottou [MSR New York], Ivan Laptev, Josef Sivic.

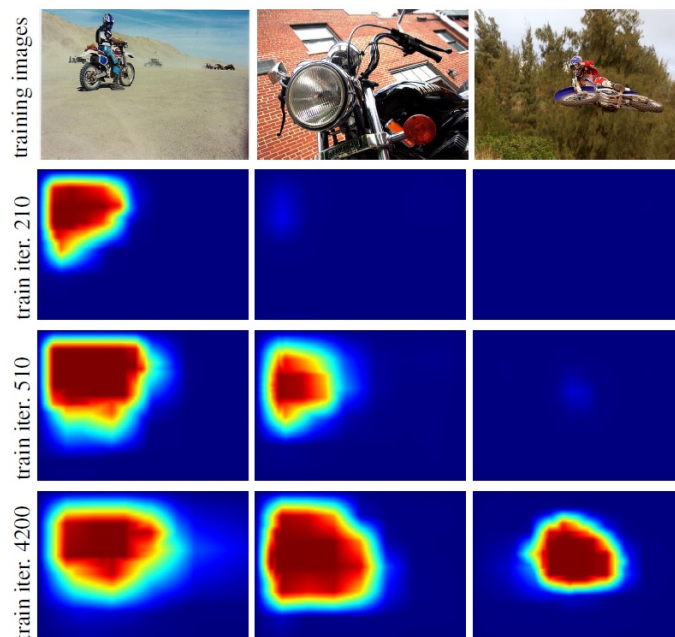


Figure 5. Evolution of localization score maps for the motorbike class over iterations of our weakly-supervised CNN training. Note that locations of objects with more usual appearance are discovered earlier during training.

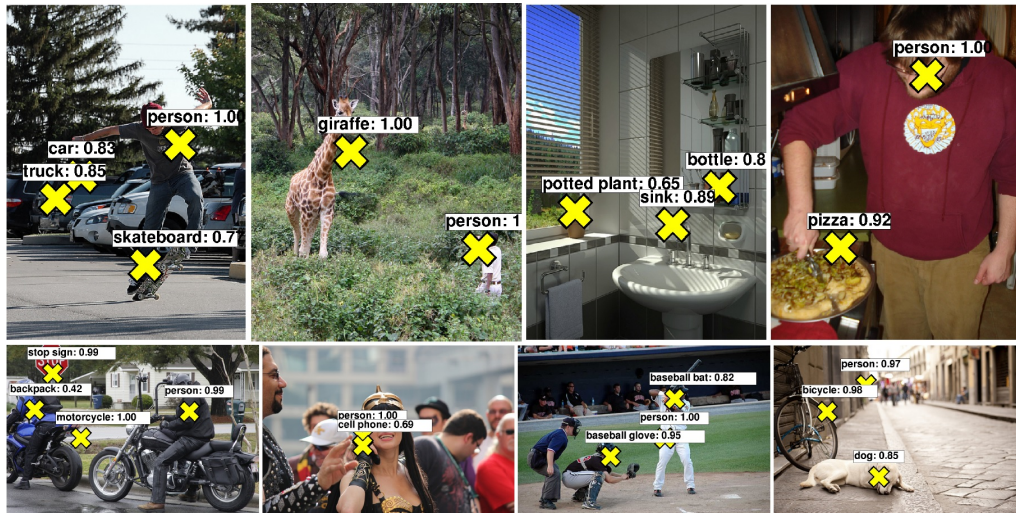


Figure 6. Example location predictions for images from the Microsoft COCO validation set obtained by our weakly-supervised method. Note that our method does not use object locations at training time, yet can predict locations of objects in test images (yellow crosses). The method outputs the most confident 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 5). 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 submitted to CVPR 2015 [23]. Illustration of localization results by our method in Microsoft COCO dataset is illustrated in Figure 6.

### 6.3.5. Learning Dictionary of Discriminative Part Detectors for Image Categorization and Cosegmentation

**Participants:** Jian Sun, Jean Ponce.

This work proposes a novel approach to learning mid-level image models for image categorization and cosegmentation. We represent each image class by a dictionary of discriminative part detectors that best discriminate that class from the background. We learn category-specific part detectors in a weakly supervised setting in which the training images are only labeled with category labels without part / object location labels. We use a latent SVM model regularized by  $l_{1,2}$  group sparsity to learn the discriminative part detectors. Starting from a large set of initial parts, the group sparsity regularizer forces the model to jointly select and optimize a set of discriminative part detectors in a max-margin framework. We propose a stochastic version of a proximal algorithm to solve the corresponding optimization problem. We apply the learned part detectors to image classification and cosegmentation, and quantitative experiments with standard benchmarks show that our approach matches or improves upon the state of the art. This work has been submitted to PAMI [24].

## 6.4. Image restoration, manipulation and enhancement



### 6.4.1. *Fast Local Laplacian Filters: Theory and Applications*

**Participants:** Mathieu Aubry, Sylvain Paris [Adobe], Samuel Hasinoff [Google], Jan Kautz [University College London], Fredo Durand [MIT].

Multi-scale manipulations are central to image editing but they are also prone to halos. Achieving artifact-free results requires sophisticated edge-aware techniques and careful parameter tuning. These shortcomings were recently addressed by the local Laplacian filters, which can achieve a broad range of effects using standard Laplacian pyramids. However, these filters are slow to evaluate and their relationship to other approaches is unclear. In this work, we show that they are closely related to anisotropic diffusion and to bilateral filtering. Our study also leads to a variant of the bilateral filter that produces cleaner edges while retaining its speed. Building upon this result, we describe an acceleration scheme for local Laplacian filters on gray-scale images that yields speed-ups on the order of 50x. Finally, we demonstrate how to use local Laplacian filters to alter the distribution of gradients in an image. We illustrate this property with a robust algorithm for photographic style transfer. This work has been published at ACM Transactions on Graphics 2014 [2].

### 6.4.2. *Learning a Convolutional Neural Network for Non-uniform Motion Blur Removal*

**Participants:** Jian Sun, Wenfei Cao, Zongben Xu, Jean Ponce.

In work 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 submitted to CVPR 2015.

## 6.5. Human activity capture and classification

### 6.5.1. *Weakly Supervised Action Labeling in Videos Under Ordering Constraints*

**Participants:** Piotr Bojanowski, Remi Lajugie [Inria Sierra], Francis Bach [Inria Sierra], Ivan Laptev, Jean Ponce, Cordelia Schmid [Inria Lear], Josef Sivic.

We are given a set of video clips, each one annotated with an ordered list of actions, such as “walk” then “sit” then “answer phone” extracted from, for example, the associated text script. We seek to temporally localize the individual actions in each clip as well as to learn a discriminative classifier for each action. We formulate the problem as a weakly supervised temporal assignment with ordering constraints. Each video clip is divided into small time intervals and each time interval of each video clip is assigned one action label, while respecting the order in which the action labels appear in the given annotations. We show that the action label assignment can be determined together with learning a classifier for each action in a discriminative manner. We evaluate the proposed model on a new and challenging dataset of 937 video clips with a total of 787720 frames containing sequences of 16 different actions from 69 Hollywood movies. This work has been published at ECCV 2014 [10].

### 6.5.2. *Predicting Actions from Static Scenes*

**Participants:** Tuan-Hung Vu, Catherine Olsson [MIT], Ivan Laptev, Aude Oliva [MIT], Josef Sivic.

Human actions naturally co-occur with scenes. In this work we aim to discover action-scene correlation for a large number of scene categories and to use such correlation for action prediction. Towards this goal, we collect a new SUN Action dataset with manual annotations of typical human actions for 397 scenes. We next discover action-scene associations and demonstrate that scene categories can be well identified from their associated actions. Using discovered associations, we address a new task of predicting human actions for images of static scenes. We evaluate prediction of 23 and 38 action classes for images of indoor and outdoor scenes respectively and show promising results, see Figure 7. We also propose a new application of geo-localized action prediction and demonstrate ability of our method to automatically answer queries such as “Where is a good place for a picnic?” or “Can I cycle along this path?”. This work has been published in ECCV 2014 [16].

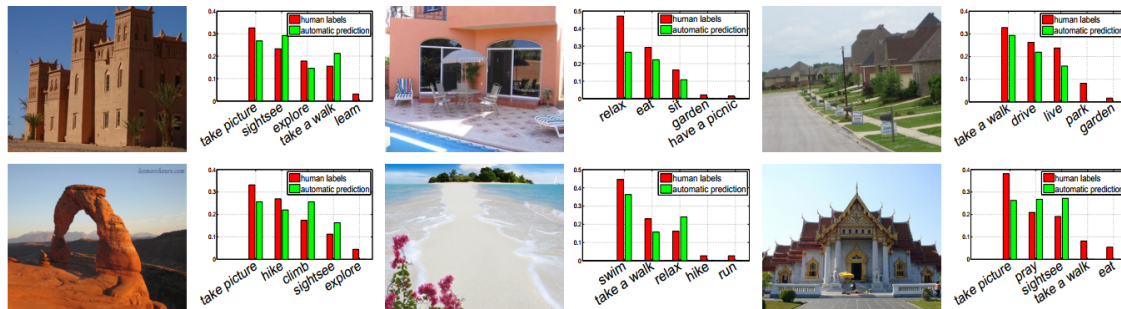


Figure 7. Automatic visual action prediction for test images in SUN Action dataset.

### 6.5.3. Efficient feature extraction, encoding and classification for action recognition

**Participants:** Vadim Kantorov, Ivan Laptev.

Local video features provide state-of-the-art performance for action recognition. While the accuracy of action recognition has been continuously improved over the recent years, the low speed of feature extraction and subsequent recognition prevents current methods from scaling up to real-size problems. We address this issue and first develop highly efficient video features using motion information in video compression. We next explore feature encoding by Fisher vectors and demonstrate accurate action recognition using fast linear classifiers. Our method improves the speed of video feature extraction, feature encoding and action classification by two orders of magnitude at the cost of minor reduction in recognition accuracy. We validate our approach and compare it to the state of the art on four recent action recognition datasets. This work has been published at CVPR 2014 [12].

### 6.5.4. On Pairwise Cost for Multi-Object Network Flow Tracking

**Participants:** Visesh Chari, Simon Lacoste-Julien [Inria Sierra], 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 submitted to CVPR 2015 [21].