



RESEARCH CENTER

FIELD

**Perception, Cognition and Interaction**

Activity Report 2013

# Section New Results

Edition: 2014-03-20



DATA AND KNOWLEDGE REPRESENTATION AND PROCESSING

1. DAHU Project-Team ..... 5  
2. DREAM Project-Team ..... 7  
3. EXMO Project-Team ..... 13  
4. GRAPHIK Project-Team ..... 16  
5. LINKS Team ..... 21  
6. MAGNET Team ..... 23  
7. MAIA Project-Team ..... 25  
8. OAK Project-Team ..... 36  
9. ORPAILLEUR Project-Team ..... 39  
10. SMIS Project-Team ..... 46  
11. TYREX Team ..... 48  
12. WIMMICS Project-Team ..... 52  
13. ZENITH Project-Team ..... 63

INTERACTION AND VISUALIZATION

14. ALICE Project-Team ..... 69  
15. AVIZ Project-Team ..... 82  
16. HYBRID Project-Team ..... 90  
17. IMAGINE Project-Team ..... 103  
18. IN-SITU Project-Team ..... 115  
19. MANAO Team ..... 122  
20. MAVERICK Project-Team ..... 129  
21. MIMETIC Project-Team ..... 145  
22. MINT Project-Team ..... 154  
23. POTIOC Team ..... 158  
24. REVES Project-Team ..... 166  
25. TITANE Team ..... 179

LANGUAGE, SPEECH AND AUDIO

26. ALPAGE Project-Team ..... 189  
27. PANAMA Project-Team ..... 198  
28. PAROLE Project-Team ..... 210  
29. SÉMAGRAMME Project-Team ..... 219

ROBOTICS

30. COPRIN Project-Team ..... 222  
31. IMARA Project-Team ..... 227

ROBOTICS AND SMART ENVIRONMENTS

32. E-MOTION Project-Team ..... 235  
33. FLOWERS Project-Team ..... 259  
34. LAGADIC Project-Team ..... 284

VISION, PERCEPTION AND MULTIMEDIA INTERPRETATION

35. AYIN Team ..... 291

36. LEAR Project-Team .....	309
37. MAGRIT Project-Team .....	326
38. MORPHEO Team .....	329
39. PERCEPTION Team .....	335
40. Prima Project-Team .....	338
41. SIROCCO Project-Team .....	342
42. STARS Project-Team .....	353
43. TEXMEX Project-Team .....	396
44. WILLOW Project-Team .....	406

## DAHU Project-Team

# 5. New Results

## 5.1. Specification and Verification of Database Driven Systems

**Participants:** Serge Abiteboul, Luc Segoufin, Victor Vianu.

We continued our investigation on the verification of database driven systems using an automata model with registers. We have exhibited new classes of decidable scenarios using nominal set theory [25]. These new classes contain the previously known relational cases but also the some semistructured ones.

We introduce in [24] and study a model of collaborative data-driven workflows. In a local-as-view style, each peer has a partial view of a global instance that remains purely virtual. Local updates have side effects on other peers' data, defined via the global instance. We also assume that the peers provide (an abstraction of) their specifications, so that each peer can actually see and reason on the specification of the entire system. We study the ability of a peer to carry out runtime reasoning about the global run of the system, and in particular about actions of other peers, based on its own local observations. A main contribution is to show that, under a reasonable restriction (namely, key-visibility), one can construct a finite symbolic representation of the infinite set of global runs consistent with given local observations. Using the symbolic representation, we show that we can evaluate in pspace a large class of properties over global runs, expressed in an extension of first-order logic with past linear-time temporal operators, PLTL-FO. We also provide a variant of the algorithm allowing to incrementally monitor a statically defined property, and then develop an extension allowing to monitor an infinite class of properties sharing the same temporal structure, defined dynamically as the run unfolds. Finally, we consider an extension of the language, augmenting work-flow control with PLTL-FO formulas. We prove that this does not increase the power of the workflow specification language, thereby showing that the language is closed under such introspective reasoning.

## 5.2. Distributed data management

**Participants:** Serge Abiteboul, Émilien Antoine, Cristina Sirangelo.

We have studied the feasibility of query answering in the presence of incomplete information in data. In particular we have investigated when it is the case that classical query evaluation techniques, which are commonly used over complete data, suffice to answer queries also in the presence of incompleteness [26]. These results allowed to find syntactic classes of queries that can be answered efficiently under many well known semantics of incompleteness, using query answering techniques which are already implemented (and optimized) in classical database systems.

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 [21], [20], we propose a novel access control model that operates within a distributed data management framework based on datalog. Using this model, users can control access to data they own and control applications they run. They can conveniently specify access control policies providing flexible tuple-level control derived using provenance information. We present a formal specification of the model, a theoretical analysis, and an implementation. We show that the computational cost of access control is acceptable.

## 5.3. Query Processing for the Web

**Participants:** Johann Brault-Baron, Arnaud Durand, Nadime Francis, Wojciech Kazana, 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 linear preprocessing time and constant enumeration delay for first-order queries over structures of bounded expansion [27] and for monadic second-order queries over structures of bounded tree-width [15]. We also presented a survey about this work at the Intl. Conf. on Database Theory (ICDT) [19].

Web data is often structured in the XML format. In [18] we have surveyed results about static analysis of pattern-based queries over XML documents. These queries are analogs of conjunctive queries, their unions and Boolean combinations, in which tree patterns play the role of atomic formulae. These can be viewed as both queries and incomplete documents, and thus static analysis problems can also be viewed as answering queries over such documents. We looked at satisfiability of patterns under schemas, containment of queries for various features of XML used in queries, query answering, and applications of pattern-based queries in reasoning about schema mappings for data exchange.

## DREAM Project-Team

## 6. New Results

### 6.1. Diagnosis of large scale discrete event systems

**Participants:** Marie-Odile Cordier, Christine Largouët, Sophie Robin, Laurence Rozé, Yulong Zhao.

The problem we deal with is monitoring complex and large discrete-event systems (DES) such as an orchestration of web services or a fleet of mobile phones. Two approaches have been studied in our research group. The first one consists in representing the system model as a discrete-event system by an automaton. In this case, the diagnostic task consists in determining the trajectories (a sequence of states and events) compatible with the sequence of observations. From these trajectories, it is then easy to determine (identify and localize) the possible faults. In the second approach, the model consists in a set of predefined characteristic patterns. We use temporal patterns, called chronicles, represented by a set of temporally constrained events. The diagnostic task consists in recognizing these patterns by analyzing the flow of observed events.

#### *6.1.1. Distributed monitoring with chronicles - Interleaving diagnosis and repair - Making web services more adaptive*

Our work addresses the problem of maintaining the quality of service (QoS) of an orchestration of Web services (WS), which can be affected by exogenous events (i.e., faults). The main challenge in dealing with this problem is that typically the service where a failure is detected is not the one where a fault has occurred: faults have cascade effects on the whole orchestration of services. We have proposed a novel methodology to treat the problem that is not based on Web service (re)composition, but on an adaptive re-execution of the original orchestration. The re-execution process is driven by an orchestrator Manager that takes advantage of an abstract representation of the whole orchestration and may call a diagnostic module to localize the source of the detected failure. It is in charge of deciding the service activities whose results can be reused and may be skipped, and those that must be re-executed.

This year, we have improved the prototype, adding the visualization of the roadmap and the activities that do not have to be reexecuted. This work has been published in ICWS2013 [15] and we are working on a journal paper that will be submitted in 2014.

#### *6.1.2. Scenario patterns for exploring qualitative ecosystems*

This work aims at giving means of exploring complex systems, in our case ecosystems, and more recently agrosystems, specifically herd management systems. We proposed to transform environmental questions about future evolution of ecosystems into formalized queries that can be submitted to a simulation model. The system behavior is represented as a discrete event system described by a set of interacting timed automata, the global model corresponding to their composition on shared events. To query the model, we have defined high-level generic query patterns associated to the most usual types of request scenarios. These patterns are then translated into temporal logic formulas. The answer is computed thanks to model-checking techniques that are efficient for analyzing large-scale systems. Five generic patterns have been defined using TCTL (Timed Computation Tree Logic) “WhichStates”, “WhichDate”, “Stability”, “Always”, “Safety”. Three of them have been implemented using the model-checker UPPAAL.

The approach has first been experimented on a marine ecosystem under fishing pressure. The model describes the trophodynamic interactions between fish trophic groups as well as interactions with the fishery activities and with an environmental context. A paper has been previously published in the Environmental Modelling Software Journal [65]. More recently, a similar approach has been experimented on agrosystems, specifically herd management systems, for which a hybrid model has been built using hierarchical timed automata. This later work has been achieved in the context of Yulong Zhao’s PhD thesis [6] and done in collaboration with our colleagues of INRA.

### 6.1.3. Controller synthesis for dealing with “How to” queries

We extended the approach to deal with “How to” queries. As before, we rely on a qualitative model in the form of timed automata and on model-checking tools to answer queries. We proposed and compared two approaches to answer questions such as “How to avoid a given situation ?”(safety query). The first one exploits controller synthesis and the second one is a “generate and test” approach. We evaluated these two approaches in the context of an application that motivates this work, i.e. the management of a marine ecosystem and the evaluation of fishery management policies. The results have been previously published in [88].

More recently, we used similar methodological tools to analyze in the context of herd management on a catchment. A hybrid model has been built using hierarchical timed automata and scenarios can be simulated and evaluated using the approach presented in the previous paragraph. In this context, the goal is to identify and analyse the best/optimal farming practices in order to reduce nitrate pollution due to livestock effluents. We proposed to use controller synthesis tools and to couple them with machine learning techniques in order to get the best strategies and to put them on easy-to-use form. This work has been made in the context of Yulong Zhao’s PhD thesis [6] and in collaboration with our colleagues of INRA (UMR PEGASE).

## 6.2. Machine learning for model acquisition

**Participants:** Sid Ahmed Benabderrahmane, Marie-Odile Cordier, Thomas Guyet, Simon Malinowski, René Quiniou.

Model acquisition is an important issue for model-based diagnosis, especially while modeling dynamic systems. We investigate machine learning methods for temporal data recorded by sensors or spatial data resulting from simulation processes. Our main objective is to extract knowledge, especially sequential and temporal patterns or prediction rules, from static or dynamic data (data streams). We are particularly interested in mining temporal patterns with numerical information and in incremental mining from sequences recorded by sensors.

### 6.2.1. Representing and mining time series

Time series are sequences of numerical values, e.g. recorded by sensors. Since these series can be huge and subject to noise, they are often transformed into sequences of symbols. The best known symbolic transformation method is SAX (Symbolic Aggregate approxXimation) [68]. SAX is based on a piecewise constant approximation method that does not take into account the slope of the time series values in successive windows. We have extended the SAX method by adding a symbolic slope information to the SAX symbols. We have experimented our new representation, 1d-SAX, on three mining tasks. In most of these experiments 1d-SAX leads to a better accuracy than SAX [19].

We have also investigated a probabilistic representation of temporal patterns based on the latent Dirichlet allocation model (LDA). Such patterns can approximate the dynamics of a set of similar multivariate time series. We have experimented the method on hydrological flood time series to extract temporal patterns [7]. The extracted patterns were considered relevant and easy to understand by experts of the domain.

### 6.2.2. Incremental sequential mining

Sequential pattern mining algorithms operating on data streams generally compile a summary of the data seen so far from which they compute the set of actual sequential patterns. We propose another solution where the set of actual sequential patterns are incrementally updated as soon as new data arrive on the input stream. Our work stands in the framework of mining an infinite unique sequence. Our method [60] provides an algorithm that maintains a tree representation (inspired by the PSP algorithm [71]) of frequent sequential patterns and their minimal occurrences [69] in a window that slides along the input data stream. It makes use of two operations: deletion of the itemset at the beginning of the window (obsolete data) and addition of an itemset at the end of the window (new data). The experiments were conducted on simulated data and on real data of instantaneous power consumption. The results show that our incremental algorithm significantly improves the computation time compared to a non-incremental approach [61].



Recently, we have worked on the adaptation of our algorithm to closed sequential patterns. A closed pattern is a local maximal pattern such there exists no extension of this pattern having the same support. Closed patterns are known to provide a condensed representation of the solution patterns and lead to more efficient algorithms without losing information or completeness on extracted patterns. The tree of closed-patterns is less deep than the pattern-tree but the transformations of the tree by addition or deletion of items are more complex. The algorithm is under evaluation. We plan to submit a paper in 2014.

### **6.2.3. 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 [58], we presented a method to segment an image through the characterization of the evolution of a vegetation index (NDVI) on two scales: annual and multi-year. The main issue of this approach was the required computation resources (time and memory). We first propose to adapt image segmentation algorithm to SITS. Segmented images reduces the number of time series to analyze and the computation time. We secondly applied 1D-SAX to reduce data dimensionality [20]. We evaluated 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.

### **6.2.4. Analysis of landscape based on spatial patterns**

Researchers in agro-environment need a great variety of landscapes to test the agro-ecological models of their scientific hypotheses. 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 realistic landscapes. This approach 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. We extract spatial patterns from a real geographic area and we use these patterns to generate new realistic landscapes. Using patterns preserves the interface properties between parcels.

We have begun the exploration of graph mining techniques, such as gSPAN [87], to discover the relevant spatial patterns present in a spatial-graph. But the graph-mining techniques are very time-consuming in comparison to sequence mining.

This year, we would like to test if using a path instead of a graph would be a faithful representation of the spatial organization of the landscape. In [17], we compare the potential expressivity of graphs and Hilbert-Peano curves [66] to characterize an agricultural landscape. The results show that mining frequent patterns in Hilbert-Peano curves would be as discriminant as mining frequent patterns in graphs.

The perception of the environment is an important dimension of the landscape we live in. One of our objectives is to study the relationships between the landscape patterns and their perception. We cope with this dimension by analysing the textual content of "atlas du paysage" (landscape atlas), that are produce by each french administrative regions. This year we worked on the construction of an ontology of landscape perception [21].

### **6.2.5. 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 require to discover relevant pattern-sequences that are recurrent over the overall time series sequences, and to find temporal associations among these frequently occurring patterns. Previous work on information retrieval and similarity search in time series has been performed in different contexts such as diagnosis or failure detection of industrial materials. In whole query matching, a time series given as query is entirely compared to every time series of a database. The series should have same length, and a similarity measure is used to retrieve either a most similar time series or the top-k ranked time series. However, theses 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. Moreover, in some real world applications, one can be interested in retrieving specific interesting subsequences that are frequently present at different instants.

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. During the preparation step, all the multi-valued time series present in the database are transformed into a multi-resolution symbolic representation thus ensuring a lower dimensionality. Then, to accelerate and enhance the similarity search and the retrieval over the database, our model creates an index over recurrent patterns in the time series collection. These patterns can be generated by different techniques. Finally, the extracted patterns are grouped by clustering and the resulting clusters are indexed in a table within their centroids. A paper presenting the preliminary results is under submission to an international journal.

### 6.2.6. 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. A two-way classification method was proposed, combining genes expression levels, represented as numerical data, and Gene Ontology (GO) annotations as symbolic data. The hopeful results obtained with genes clustering, through GO annotations, are an encouraging track to predict transcriptional regulatory networks, and for refining the existing sets of genes [11], [12].

We also introduced a new method for extracting enriched biological functions from transcriptomic databases using an integrative bi-classification approach. The initial gene datasets are firstly represented as a formal context (objects attributes), where objects are genes, and attributes are their expression profiles and complementary information of different knowledge bases. Formal Concept Analysis (FCA) is applied for extracting formal concepts regrouping genes having similar transcriptomic profiles and functional behaviors. An enrichment analysis is then performed in order to identify the relevant formal concepts from the generated Galois lattice, and to extract biological functions that could participate in the proliferation of cancers. Preliminary results seem very promising, and could help experts during the identification of degenerated biological functions [13].

## 6.3. Decision aiding with models and simulation data

**Participants:** Louis Bonneau de Beaufort, Tassadit Bouadi, Marie-Odile Cordier, Véronique Masson, René Quiniou.

Models can be very useful for decision aiding as they can be used to play different plausible scenarios for generating the data representing future states of the modeled process. However, the volume of simulation data may be huge. Thus, efficient tools must be investigated in order to store the simulation data, to focus on relevant parts of the data and to extract interesting knowledge from these data.

### 6.3.1. A datawarehouse for simulation data

The ACASSYA project 8.2.1 aims at providing experts or stakeholders or farmers with a tool to evaluate the impact of agricultural practices on water quality. As the simulations of the deep model TNT2 are time-consuming and generate huge data, we have proposed to store these simulation results in a datawarehouse and to extract relevant information, such as prediction rules, from the stored data. We have devised a general architecture for agro-environmental data on top of the framework Pentaho.

This year we have been working on the efficient computation of OLAP queries related to realistic scenarios proposed by experts in the domain. Precisely, we have devised indexing schemes to access the data in the OLAP cube. We have also worked on the visualization by a GIS (Geographical Information System) of the query results on maps of the geographical area under interest. A paper have been submitted to the COMPAG Journal. This work is detailed in Tassadit Bouadi's thesis [5].

### 6.3.2. Efficient computation of skyline queries in an interactive context

Skyline queries retrieve from a database the objects that maximize some criteria, related to user preferences for example, or objects that are the best compromises satisfying these criteria. When data are in huge volumes, such objects may shed light on interesting parts of the dataset. However, computing the skylines (i.e. retrieving the skyline points) may be time consuming because of many dominance tests. This is, especially the case in an interactive setting such as querying a data cube in the context of a datawarehouse.

We have worked at improving the formal setting of the partial materialization of skyline queries when dynamic preferences are refined online by the user. We have explicitated which parts of the skyline evolve (which point are added or removed) when a new dimension is introduced in the computation. This led to an efficient incremental method for the online computation of the skyline corresponding to new user preferences [46]. An extended version of this paper is published in Journal "Transactions on Large Scale Data and Knowledge Centered Systems" (TLDKS) [8] and in Tassadit Bouadi's thesis [5].

### 6.3.3. Hierarchical skylines

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. It allows the user to navigate along the dimensions hierarchies (i.e. specialize / generalize) while ensuring an online calculation of the associated skyline. The method is described in Tassadit Bouadi's thesis [5]. A paper describing this contribution is currently under submission to the "Very Large Data Bases (VLDB 2014)" conference.

### 6.3.4. Modeling influence propagation by Bayesian causal maps

The goal of this project is modeling shellfish fishing to assess the impact of management pollution scenarios on the *Rade de Brest*. Cognitive maps were built from interviews with fishermen. To represent and reason about these cognitive maps, we propose to use Bayesian Causal Maps making use of fishermen knowledge, particularly to perform influence propagation [82].

However, this model does not take into account the variety of influences asserted by the fishermen, but only the "mean" causal map. A report describing the project is available [28]. An approach that could combine individual knowledge with belief functions in the way of Philippe Smets's Transferable Belief Model [83] has been proposed. A report describing the project is available [28].

This work is done in the framework of the RADE2BREST project, involving Agrocampus Ouest and CNRS (GEOMER/LETG), funded by "Ministère de l'Ecologie" (This project is not mentioned in section 8.2 because DREAM is not an official partner of this project.).

### 6.3.5. Recommending actions from classification rules

In the framework of the SACADEAU project, a paper dedicated to building actions from classification rules has been published in the KAIS Journal [9]. Our goal is to burden of analysing a large set of classification rules when the user is confronted to an "unsatisfactory situation" and needs help to decide about the appropriate actions to remedy to this situation. The method consists in comparing the situation to a set of classification rules. For this purpose, we propose DAKAR, a new framework for learning action recommendations dealing with complex notion of feasibility and quality of actions.

Sacadeau-Software, which is the decision support tool implemented with F. Ployette (former Inria engineer in the EPI Dream, now retired) in the SACADEAU project, has been published in the RIA Journal [10]. Sacadeau-Software allows to run simulations throughout a watershed and obtain the transfer rate of pollution through the catchment. Classification rules, characterizing the sub-parts of the watershed with pollution and the sub-parts without pollution, are automatically learned from the simulations. A visualization tool enables to relate the learned rules to the examples characterized by these rules. Finally, a user can select a situation of pollution

and the action recommendation tool analyses the learned rules and proposes actions that improve this situation of pollution.

## 6.4. Diagnostic, causal reasoning and argumentation

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

Stemming on [38], [39], [40], [41], [42], we have designed an inference system based on causal statements. This is related to diagnosis (observed symptoms explained by faults). The aim is to produce possible explanations for some observed facts. Previously existing proposals were ad-hoc or, as in [45], [57], they were too close to standard logic to make a satisfactory diagnosis. A key issue for this kind of work is to distinguish logical implication from causal links and from ontological links. This is done by introducing a simple causal operator, and an *is-a* hierarchy. These two operators are added to a restricted first order logic of the Datalog kind (no function symbols). Then, our system produces elementary *explanations* for some set of observed facts. Each explanation links some facts to the considered observation, together with a set of atoms called the *justifications*: The observation is explained from these facts, provided the justifications are possible (not contradicted by the available data). This formalism has been translated into answer set programming [72], [73]). It is able to deal with complex problems such as finding explanations for the hurricane Xynthia (2010, February 28). In such situations, there are many data and many possible elementary explanations can be examined. This involves an extension of our formalism, in order to deal with more complex chains of causations and *is-A* links. Our formalism makes precise what all these possible explanations are. Then, in order to deal with so many possible complex explanations, we integrate this causal formalism into an argumentation framework. Logic-based formalizations of argumentation [43] take pros and cons for some conclusion into account. These formalizations assume a set of formulae and then exhaustively lay out arguments and counterarguments. This involves providing an initiating argument for the inference and then providing undercuts to this argument, and then undercuts to undercuts. So here our causal formalism provides a (rather large) set of explanations, and the argumentation part allows to select the best ones, under various criteria [22], [14].

Then, since answer set programming can easily deal with logical formalisms, the argumentation part will be incorporated into our already existing answers set programming translation of the causal formalism. Regarding this field of knowledge representation and reasoning, and more generally, artificial intelligence, we have participated to several chapters in the to be published "Panorama de l'intelligence artificielle. Ses bases méthodologiques, ses développements" [27], [26], [23], [24].

## EXMO Project-Team

# 6. New Results

## 6.1. Ontology matching and alignments

We pursue our work on ontology matching and alignment support [5], [12] with contributions to evaluation and alignment semantics.

### 6.1.1. 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 2013 evaluation campaign [7]. It offered 8 different test sets (7 of which under the SEALS platform). This issue brought the following results:

- Once again, more participants than ever (23);
- Most ontology matchers running on the SEALS platform (20);
- Increased performances in terms of precision and recall;
- Matchers are now faster and more scalable. There are also more matchers using networked resources.

We used again the our generator for generating new version of benchmarks [4]. The Alignment API was used for manipulating alignments and evaluating results.

A novelty of this year was the evaluation of interactive systems, included in the SEALS client. It brings interesting insight on the performances of such systems and should certainly be continued.

The participating systems and evaluation results were presented in the 8th Ontology Matching workshop, that was held in Sydney, Australia [13]. More information on OAEI can be found at <http://oaei.ontologymatching.org/>.

### 6.1.2. Algebras of relations in alignments

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

We had previously shown that algebras of relations between concepts can be used for expressing relations in alignments. We have worked this year as extending them in two ways.

We increased the expressiveness of relations between concepts, not restricting the algebra to necessarily non empty concepts. This describes all taxonomical (as opposed to mereological) relation algebras, i.e., all those relations that have been used by matchers so far.

We also dealt with relations among different kinds of entities – individuals or concepts. For this, relation algebra structures are considered in an arbitrary one- or many-sorted logical theory. We established a sufficient condition for a set of dyadic formulas in a first-order theory to generate a relation algebra. This result is extended to many-sorted theories by means of Schröder categories.

This work is part of the PhD of Armen Inants.

## 6.2. Data interlinking

The web of data uses semantic web technologies to publish data on the web in such a way that they can be interpreted and connected together. It is thus critical 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.

### 6.2.1. Interlinking cross-lingual RDF data sets

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

Data interlinking is a difficult task in a cross-lingual environment like the Web. Even systems based on graph structure, ultimately rely on anchors based on language fragments. If languages are different, fragments have to be compared by more sophisticated techniques. In that context, we are developing an approach which represents RDF entities as (virtual) text documents and compare them using different strategies [9], [10]. We investigate two directions: (1) a translation-based approach where the virtual documents are automatically translated; (2) a language-independent approach where important terms found in documents are mapped to a terminological resource like Wordnet to compute document similarity.

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

### 6.2.2. Data interlinking from expressive alignments

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

In the context of the DATALIFT project, we are further developing the data interlinking module. We have developed an algorithm able to determine potential attribute correspondences of two classes depending on their features. For that purpose, we use  $k$ -means or  $k$ -medoids clustering. These correspondences are then used to construct a SILK script which generates an initial link set. Some of the links are presented to the user who assesses their validity. We then use an improvement of the disjunctive version space supervised learning method to learn a better script from the assessed links. Such a technique can be iterated until satisfactory links are found.

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

### 6.2.3. Key and pseudo-key detection for web data set interlinking

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

Keys are sets of properties which uniquely identify individuals (instances of a class). We have refined the notion of database keys in a way which is more adapted to the context of description logics and the openness of the semantic web. We have also refined the weaker notion of a linkkey introduced in [12]. Then we have shown how such keys, together with ontology alignments, and linkkeys may be used for deducing equality statements (links) between individuals across data sources in the web of data.

However, ontologies do not necessarily come with key descriptions, and never with linkkey assertions (which would hold across ontologies). But, these can be extracted from data by assuming that keys holding for specific data sets, may hold universally. We have extended these classical key extraction techniques for extracting linkkeys.

This work is developed partly in the LINDICLE and DATALIFT projects. A proof of concept implementation is available at <http://rdfpkeys.inrialpes.fr/>.

## 6.3. Ontology networks

Dealing with the semantic web, we are interested in ontology networks, i.e., sets of distributed ontologies that have to work together. One way for these systems to interact consists of exchanging queries and answers. For that reason, we pay particular attention to query systems.

### 6.3.1. Path queries and $\mu$ -calculus

**Participant:** Jérôme Euzenat.

Querying the semantic web is mainly done through the SPARQL language or its extensions through paths and entailment regimes [14]. Query containment is the problem of deciding if the answers to a query are included in those of another query for any queried data sources. This problem is very important for query optimisation purposes. In the SPARQL context, it can be equally useful for distributing federated queries or for implementing schema-based access control. In order to experimentally assess implementation strengths and limitations, we provided a first SPARQL containment test benchmark. We studied the query demographics on DBPEDIA logs to design benchmarks for relevant query containment solvers. We tested available solvers on their domain of applicability on three different benchmark suites [6] and found that (i) tested solutions are overall functionally correct, (ii) in spite of its complexity, SPARQL query containment is practicable for acyclic queries, (iii) state-of-the-art solvers are at an early stage both in terms of capabilities and implementation.

This work has been developed in collaboration with the TYREX team and within the PhD thesis of Melisachew Wudage Chekol now in the ORPAILLEUR team. The benchmarks, results and software are available at <http://sparql-qc-bench.inrialpes.fr>.



## GRAPHIK Project-Team

# 6. New Results

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

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

*Note that for this section, as well as all sections in New Results, participants are given in alphabetical order.*

This year we continued to work on the existential rule framework in the context of Ontology-Based Query Answering (a.k.a. Ontology-Based Data Access, OBDA). See the 2011-2012 activity reports for details on this framework (a.k.a. Tuple-Generating Dependencies or Datalog+/-). The ontology-based query answering issue consists in querying data while taking into account inferences enabled by an ontology. This ontology is here described by existential rules, a very expressive formalism which generalizes the lightweight description logics used for OBDA (e.g. the tractable fragments of the Semantic Web language OWL 2).

From 2009 to 2011, we mainly investigated decidability and complexity issues. In 2012, we tackled the next step, which consists in developing algorithms with good theoretical properties (they should at least run in the “right” worst-case complexity class) and with good performance in practice. There are two main ways of processing rules, namely forward chaining and backward chaining, which are also known as “materialization” and “query rewriting”. In forward chaining, rules are applied to enrich the initial data and query answering can then be solved by evaluating the query against the “saturated” database (as in a classical database system, i.e., forgetting the rules). 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).

In 2013, on the one hand we focussed on the improvement of query rewriting algorithms, on the other hand we began to investigate extensions of our framework.

### 6.1.1. Improvement of Query Rewriting Algorithms

The advantage of the query rewriting approach is that the data are not modified (hence no write access permission is required and the data do not grow; moreover, there is no materialization that would need to be updated when data change). However, the practicability of this approach is questionable due to (1) the weak expressivity of classes for which efficient rewriters have been implemented, and (2) the large size of rewritings using UCQ.

With respect to the first point, we improved the algorithm designed in 2012. This algorithm accepts as input any set of existential rules and stops if this set of rules fulfills so-called finite unification set (fus) property, meaning that the set of rules allows to rewrite any query as a first-order query, e.g. a UCQ (this property is not true in general, where no finite rewriting may exist). We also studied properties of rewriting operators that ensure the correctness and the termination of a generic breadth-first rewriting algorithm and analyzed some operators with respect to these properties.

- *Work published in IJCAI 2013 [37] and RR 2013 (Rules and Web Reasoning) [36]*

With respect to the second point, we defined semi-conjunctive queries (SCQs), which are a syntactical extension of conjunctive queries. We designed and implemented an algorithm called Compact, which computes sound and complete rewritings of a conjunctive query in the form of a union of SCQs (USCQs). As in the above work, any kind of existential rules can be considered, however the algorithm is ensured to stop only for fus rules. First experiments show that USCQs are both very efficiently computable and more efficiently evaluable than their equivalent UCQs.

- *Work published in IJCAI 2013 [41]*



### 6.1.2. Ongoing Work: Extensions of the Framework

*Inconsistent-tolerant query answering.* It may be the case that the data are inconsistent with the ontology, specially when there are several data sources. The classic logical framework becomes inappropriate since an inconsistent logical theory entails everything. Therefore, inconsistency-tolerant semantics have been defined to get meaningful answers. These semantics are based on the notion of repairs, which are maximal subsets of the data consistent with the ontology. In the most natural semantics, a tuple is an answer to the query if it is an answer in each repair. This issue is relevant to Pagoda and Qualinca, two ANR projects respectively started in 2013 and 2012 (see Section 8.1). Swan Rocher’s master thesis was devoted to a query answering algorithm in this framework, where the ontology is described by existential rules and negative constraints.

*Existential Rules with non-monotonic negation.* Non-monotonic negation is very useful for modeling purposes. We added non-monotonic negation to existential rules, under stable model semantics. This brought us close to logic programs considered in the area called Answer Set Programming. First results were obtained on the semantics and decidability of query answering with these rules. This work is part of ASPIQ project started in 2013 (see Section 8.1).

- Paper curenly submitted to an international conference.

### 6.1.3. Others

Michael Thomazo defended his PhD thesis entitled “Conjunctive Query Answering Under Existential Rules —Decidability, Complexity, and Algorithms” (Oct. 2013). The main contributions of this thesis are the following: first, a unified view of the currently known existential rule classes ensuring decidability of query answering, together with a complexity analysis and a worst-case optimal algorithm for a new generic class, which generalizes a family of very expressive decidable classes (see the gbts class in 2012 activity report); second, a generic algorithm for query rewriting, which overcomes some causes of combinatorial explosion that make classical approaches inapplicable.

- See the PhD thesis [15] and the extended abstract published in IJCAI 2013 [42].

The journal version extending the papers at IJCAI 2011 and KR 2012, in collaboration with Sebastian Rudolph (TU Dresden), is still in preparation but almost finished (postponement due to the addition of complementary results).

## 6.2. Reasoning with Imperfect Information and Priorities

**Participants:** Madalina Croitoru, Jérôme Fortin, Souhila Kaci, Tjitze Rienstra, Rallou Thomopoulos.

### 6.2.1. Monotonic and Non-monotonic Inference for Abstract Argumentation

An argumentation framework (or framework, for short) consists of a set of arguments, whose content may be left unspecified, together with an attack relation encoding conflict between arguments. Given a framework, a semantics specifies which sets of arguments (called extensions) are rationally acceptable. This formalism captures many different types of reasoning considered in the area of AI. In many applications, a framework somehow represents (part of) an agent’s belief state. Beliefs are then formed on the basis of acceptable sets of arguments. For example, a ‘grounded reasoner’ forms beliefs on the basis of the framework’s grounded extension, a ‘preferred reasoner’ on the basis of the preferred extensions, and so on. There is a problem with this account, however. Two different argumentation frameworks may be considered equivalent as soon as they lead to the same extensions. A more appropriate notion of equivalence is strong equivalence. Given a semantics, two frameworks are said to be strongly equivalent if their extensions are the same given every possible addition of new arguments and attacks. But still, it leaves open the question of how to form beliefs on the basis of a framework, so that different frameworks can be meaningfully distinguished, even if their extensions are the same. We addressed this problem and presented a new approach to reasoning about the outcome of an argumentation framework, where an agent’s reasoning with a framework and semantics is represented by an inference relation defined over a logical labeling language. We first studied a monotonic type of inference which is, in a sense, more general than an acceptance function, but equally expressive. In

order to overcome the limitations of this expressiveness, we studied a non-monotonic type of inference which allows counterfactual inferences. We precisely characterized the classes of frameworks distinguishable by the non-monotonic inference relation for the admissible semantics.

- *Joint work with R. Booth and L. van der Torre (Univ. of Luxembourg), published in FLAIRS 2013 [27]*

### 6.2.2. Dynamics in Abstract Argumentation

Recent years have seen a considerable work on dynamics in argumentation framework (AF). We addressed dynamics in abstract argumentation using a logical theory where an agent's belief state consists of an argumentation framework and a constraint that encodes the outcome the agent believes the argumentation framework should have. Dynamics enters in two ways: (1) the constraint is strengthened upon learning that the AF should have a certain outcome and (2) the argumentation framework is expanded upon learning about new arguments/attacks. A problem faced in this setting is that a constraint may be inconsistent with the AF's outcome. We discussed two ways to address this problem: First, it is still possible to form consistent fallback beliefs, i.e., beliefs that are most plausible given the agent's argumentation framework and constraint. Second, we showed that it is always possible to find argumentation framework expansions to restore consistency. Our work combines various individual approaches in the literature on argumentation dynamics in a general setting.

- *Joint work with R. Booth and L. van der Torre (Univ. of Luxembourg), published in SUM 2013. [26]*

Preferences have been intensively studied in argumentation framework. Preference-based argumentation frameworks are instantiation of Dung's framework in which the defeat relation (in the sense of Dung) is computed from an attack relation and a preference relation over the set of arguments. We distinguish between different ways to derive preferences over arguments, e.g., from their relative specificity, relative strength or from values promoted by the arguments. However an underexposed aspect in these models is change of preferences. We proposed a dynamic model of preferences in argumentation, centering on what we call property-based AFs. It is based on Dietrich and List's model of property-based preference and it provides an account of how and why preferences in argumentation may change. The idea is that preferences over arguments are derived from preferences over properties of arguments, and change as the result of moving to different motivational states. We also provided a dialogical proof theory that establishes whether there exists some motivational state in which an argument is accepted.

- *Joint work with R. Booth (Univ. of Luxembourg), published in ADT 2013.*

### 6.2.3. Representing Synergy Among Arguments with Choquet Integral

Preference-based argumentation frameworks are instantiation of Dung's framework in which the defeat relation (in the sense of Dung) is computed from an attack relation and a preference relation over the set of arguments. Value-based argumentation framework is a preference-based argumentation framework where the preference relation over arguments is derived from a preference relation over values they promote. We extended value-based argumentation framework with collective defeats and arguments promoting values with various strengths. In the extended framework, we defined a function which computes the strength of a collective defeat. We also defined desired properties for the proposed function. Surprisingly, we showed that this function obeying the corresponding properties is Choquet integral, a well-known aggregation function at work in multiple criteria decision.

- *Joint work with C. Labreuche (Thales), published in EC-SQARU 2013 [35]*

### 6.2.4. Compiling Preference Queries in Qualitative Constraint Problems

Comparative preference statements are the basic ingredients of conditional logics for representing users' preferences in a compact way. These statements may be strict or not and obey different semantics. Algorithms have been developed in the literature to compute a preference relation over outcomes given a set of comparative preference statements and one or several semantics. These algorithms are based on insights from non-monotonic reasoning (more specifically, minimal and maximal specificity principles) enforcing the preference

relations to be a complete preorder. The main limitation of these logics however relies in preference queries when comparing two outcomes. Indeed given two outcomes having the same preference w.r.t. the preference relation, there is no indication whether this equality results from an equality between two preference statements or the outcomes are in fact incomparable and equality has been enforced by specificity principles. On the other hand, comparative preference statements and their associated semantics can be translated into qualitative constraint satisfaction problems in which one can have a precise ordering over two outcomes. We investigated this bridge and provided a compilation of conditional logics-based preference queries in qualitative constraint problems.

- *Joint work with J.-F. Condotta (CRIL), published in FLAIRS 2013 [31]*

### 6.2.5. Argumentation for Reasoning with Inconsistencies

We investigate the use of argumentation when reasoning over an inconsistent knowledge base. We use argumentation in this context given the explanation power that it may bring (and that is currently under investigation).

We have investigated logical based argumentation following two methods. First, we have defined our own argument and attack notion (given the logical language at hand) and showed that such instantiation respects desirable properties of consistency and maximality (called rationality postulates in the field). This work has showed that the ICR, AR, IAR semantics investigated by inconsistent query answering (see Pagoda, Section 8.1) are the same as skeptically preferred or stable semantics, grounded and universally stable or preferred. Such result is encouraging as it bridges the two communities (argumentation and inconsistent query answering) allowing to use results from one field in order to enrich the other. We have also investigated the practical applicability of such argument definition and approach in the selection of flour for bread.

- *Joint work with Srdjan Vesic (Univ. of Luxembourg), published in RIA 2013 [23] and SUM 2013 [32]*

On the other hand we have also looked at using a generic logical argumentation framework (ASPIC) in order to instantiate it with a simple logic in the EcoBioCap project (see Section 8.2). We have extended previous results to enrich bipolar queries. A software tool is under construction.

- *Work published in RIA 2013 [21]*

## 6.3. Semantic Data Integration

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

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. Since the seminal paper [59], this issue has been studied under various names: “record linking”, “entity resolution”, “reference resolution”, “de-duplication”, “object identification”, “data reconciliation”, etc., mostly in databases (cf. the bibliography by William E. Winckler [60]). 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.

Past years, 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 catalogue from the ABES agency). The prototype providing this service has been implemented on top of Cogui and experiments have been led in the context of the SudocAd project (jointly conducted by ABES and GraphIK).

Our proposed method can be stated as follows: first, enrich authority records with knowledge extracted from bibliographic records in which the authority is mentioned ; then, use logical rules which conclude on different levels of reconciliation, to compare the authors of a new bibliographic record with the enriched authority records ; finally, for each author of the new bibliographic record, order the authority identifiers by level of reconciliation.

- *Work published in [30].*

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. This year, this work has been pursued along three different axis:

1. We have built a formal framework allowing to evaluate the quality of links in a documents database. We propose two different “quality” notions, based upon an identification predicate *id* and a differentiation predicate *di* between pairs of authority notices identifiers (these predicates can be either given by an expert or computed using rules). We have first introduced the notion of a *well-founded* database, when *id* is an equivalence relation and *di* its complement. This property can be checked using logical inferences and combinatorial techniques. In the general case where a database is not necessarily well-founded, we have proposed different distances to a well-founded one. We have also introduced a more complex quality criterion that corresponds to *stability by substitution* (a fundamental property of logical equality that is not necessarily satisfied by *id*).
  - *A research report should lead to a publication in 2014.*
2. We developed a methodology for detecting linkage errors and fixing them, based upon a clustering method of authors in bibliographic records. Last year, the general schema of the methodology was defined. It is based upon a set of criteria which allows us to cluster “similar” authors together. Each criterion represents a point of view on the author: name, publication time span, publication domain, etc... This year, two aggregation semantics for such criteria have been developed, implemented and evaluated.
  - *Work published in AI-SGAI 2013 [34].*
3. We have studied methods allowing to automatically extract similarity criteria between named entities. This problem is very similar to the automatic discovery of composite key constraints in RDF data sources that conform to a given ontology. We have studied the different existing methods allowing to discover such keys, and have proposed logical semantics for these different keys. These semantics allow to understand and compare the results produced by these different methods. These methods have been evaluated against the documentary databases provided by our partners ABES and INA.
  - *Work described in a research report [48], at the moment, two papers are submitted.*

## LINKS Team

# 5. New Results

## 5.1. Querying Heterogeneous Linked Data

**Participants:** Guillaume Bagan, Iovka Boneva, Angela Bonifati, Pierre Bourhis, Radu Ciucanu, Tom Sebastian, Slawomir Staworko, Sophie Tison.

Staworko, Ciucanu and Boneva presented a new class of schemas for unordered XML trees, which are based on unordered regular expressions, also called multiplicity schemas. They show that many static analysis problems become feasible when removing disjunctions there [6].

Ciucanu and Staworko [8] investigated the case of unordered XML, where the relative order among siblings is ignored, and focused on the problem of learning schemas from examples given by the user. They considered disjunctive multiplicity schemas (DMS) and their restrictions, disjunction-free multiplicity schemas (MS). For both DMS and MS, they prove the learnable cases.

Regular path queries in graphs have found much recent interest in the context of SPARQL queries for linked open data in the RDF format. Bagan, Bonifati and Groz (former PhD student of Mostrare, now PostDoc at Tel-Aviv University) have obtained a precise characterization of those regular path queries that can be answered with polynomial data complexity [5] leading to a trichotomy (AC0, NL-complete, or else NP-complete). Thereby, they have solved an open question (raised by W. Martens in PODS'12).

XPath query evaluation over compressed trees has been studied in [12]. They focused on a fragment of XPath, which is the downward, navigational XPath and presented precise bounds on the time complexity of XPath query execution over grammar-compressed trees. In particular, they focused on counting the nodes selected by an XPath expression, extracting and materializing their pre-order numbers and serializing the obtained subtrees.

In [2], Groz, Staworko, Caron, Roos and Tison studied query rewriting with views when the classes used to define queries and views are Regular XPath and MSO. Next, they investigated problems of static analysis of security access specifications (SAS) by introducing the novel class of interval-bounded SAS and they defined three different manners to compare views (i.e. queries), with a security point of view. Finally, they provided a systematic study of the complexity for deciding these three comparisons.

## 5.2. Managing Dynamic Linked Data

**Participants:** Angela Bonifati, Denis Debarbieux, Joachim Niehren, Tom Sebastian.

Bonifati, Goodfellow (former PhD student at the University of Strathclyde, UK), Manolescu and Sileo (former PhD student at the University of Basilicata, Italy, directed by Bonifati) studied XML view maintenance in the presence of updates [1]. Their approach relies on algebraic operators for propagating source updates to the target XML view, e.g. in a typical scenario of GAV (global-as-view) schema mappings. Their algebraic approach is set-oriented as opposed to tuple-oriented methods presented in the literature. Moreover, it leverages structural identifiers and structural join algorithms. As such, it proved to be more efficient than existing methods for updating materialized XML views.

Debarbieux, Gauwin (former PhD student in the team, now Assistant Professor at the University of Bordeaux), Niehren, Sebastian and Zergaoui (CEO at Innovimax) focused on using early nested word automata in order to approximate earliest query answering algorithms for nested word automata in a highly efficient manner [9]. This approximation can be made tight in practice for automata obtained from XPath expressions. An XPath streaming algorithm based on early nested word automata has been implemented in the FXP tool. FXP outperforms most previous tools in efficiency, while covering more queries of the XPathMark benchmark.

### 5.3. Linking Data Graphs

**Participants:** Angela Bonifati, Radu Ciucanu, Joachim Niehren, Aurélien Lemay, Grégoire Laurence, Antoine Ndione, Slawomir Staworko.

In [7], Bonifati, Ciucanu and Staworko investigate the problem of inferring arbitrary n-ary join predicates across two relations via user interactions. The relations can be found on the Web, thus they lack integrity constraints. In such a scenario, the user is asked to label as positive or negative a few tuples depending on whether she would like them in the join result or not. Deciding whether the remaining tuples are uninformative, i.e. do not allow to infer the query goal, can be done in polynomial time.

The PhD thesis of Ndione focuses on probabilistic algorithms to decide approximate membership of words in a language by using property testing. In [3], Ndione, Lemay and Niehren presented an algorithm that tests the membership modulo the edit distance. Their algorithm runs in polynomial time, as opposed to other property testing algorithms, leveraging the Hamming distance or the edit distance with moves, that are exponential.

In [11], Laurence, Lemay, Niehren, Staworko and Tommasi (project leader of the Magnet team) studied the problem of learning sequential top-down tree-to-word transducers (STWs). 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.

In [4], Niehren, Champavère (former PhD student in the team), Gilleron and Lemay addressed the problem of learnability of regular queries in unranked trees. The idea is that tree pruning strategies and the schemas (DTD in the specific case) can guide the learning process and lead to a class of queries that are learnable according to those. The obtained learning algorithm adds pruning heuristics to the traditional learning algorithm based on tree automata and exploiting positive and negative examples.



## MAGNET Team

# 6. New Results

## 6.1. Probabilistic models for large graph

We have developed new approaches for the statistical analysis of large-scale undirected graphs. The main insight is to exploit the spectral decomposition of subgraph samples, and in particular their Fiedler eigenvalues, as basic features for density estimation and probabilistic inference. Our contributions are twofold. First, we develop a conditional random graph model for learning to predict links in information networks (such as scientific coauthorship and email communication). Second, we propose to apply the resulting model to graph generation and link prediction. This work is published in the *Journal of Machine Learning Research*, the top journal in the field of machine learning.

## 6.2. Learning in hypergraphs

In this work, we focus on the problem of learning from several sources of heterogeneous data represented as input graphs that encode different relations over the same set of nodes. Our goal is to merge those input graphs by embedding them into an Euclidean space related to the commute time distance in the original graphs. Our algorithm outputs a combined kernel that can be used for different graph learning tasks. This work has been published in [5].

The approach designed in that paper has raised a new definition of undirected hypergraphs with bipartite hyperedges. A bipartite hyperedge is a pair of disjoint sets of nodes in which every node is associated with a weight. A bipartite hyperedge can be viewed as a relation between two teams of nodes in which every node has a weighted contribution to its team. Undirected hypergraphs generalize over undirected graphs. Consistently with the case of graphs, we have studied the hypergraph spectral framework. We have defined the notions of hypergraph gradient, hypergraph Laplacian, and hypergraph kernel as the Moore-Penrose pseudoinverse of a hypergraph Laplacian. Therefore, smooth labeling of (teams of) nodes and hypergraph regularization methods can be performed. Contrary to the graph case, we show that the class of hypergraph Laplacians is closed by the pseudoinverse operation (thus it is also the class of hypergraphs kernels), and is closed by convex linear combination. Closure properties allow us to define (hyper)graph combinations and operations while keeping a hypergraph interpretation of the result. We exhibit a subclass of signed graphs that can be associated with hypergraphs in a constructive way. A hypergraph and its associated signed graph have the same Laplacian. This property allows us to define a distance between nodes in undirected hypergraphs as well as in the subclass of signed graphs. The distance coincides with the usual definition of commute-time distance when the equivalent signed graph turns out to be a graph. We claim that undirected hypergraphs open the way to solve new learning tasks and model new problems based on set similarity or dominance. We are currently exploring applications for modeling games between teams and for graph summarization. This work [8] has been submitted to *Journal of Machine Learning Research*.

## 6.3. Natural Language Processing

In [7] and [3], we develop a new algorithm for drastically improving a pairwise coreference classification system. Specifically, this algorithm works by learning the best partition over mention type pairs by training different pairwise coreference models for each pair type. In effect, our algorithm finds the optimal feature space (from a base feature set and set of types) for separating coreferential mention pairs, but it remains tractable by exploiting the structure of the hierarchies built from the pair types. In [6], we propose a new approach for the automatic identification of so-called implicit discourse relations. Our system combines hand-labeled examples and automatically annotated examples (based on explicit relations) using different methods inspired by work on domain adaptation. Our system is evaluated empirically and yields important performance gains compared to only using hand-labeled data. This paper has received the best paper award at the *TALN 2013* conference, the national NLP conference.

## 6.4. Query Induction

We have proposed a new algorithm for query learning that combines schema-guided pruning heuristics with the traditional learning algorithm for tree automata from positive and negative examples. We show that this algorithm is justified by a formal learning model, and that for stable queries it performs very well in practice of XML information extraction. This work [1] has also been published in *JMLR*.

## 6.5. Learning Transducers

We have pursued the work on learning finite state tree-to-word transducers. Tree-to-word transformations are ubiquitous in computer science. They are the core of many computation paradigms from the evaluation of abstract syntactic trees to modern programming languages *XSLT*. We have extended the results obtained last year on the study of a class of sequential top-down tree-to-word transducers, called *STWs*. Transducers in *STWs* are capable of: concatenation in the output, producing arbitrary context-free languages, deleting inner nodes, and verifying that the input tree belongs to the domain even when deleting parts of it. These features are often missing in tree-to-tree transducers, and for instance, make *STWs* incomparable with the class of top-down tree-to-tree transducers. The class of *STWs* has several interesting properties, in particular we proposed in 2011 a normal for *STWs*.

In [4], we present a Myhill-Nerode characterization of the corresponding class of sequential tree-to-word transformations. 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. Decision Making

#### 6.1.1. Searching for Information with MDPs

**Participants:** Mauricio Araya, Olivier Buffet, Vincent Thomas, François Charpillet.

In the context of Mauricio Araya’s PhD and PostDoc, we are working on how MDPs – or related models – can search for information. This has led to various research directions, such as extending POMDPs so as to optimize information-based rewards, or actively learning MDP models. This year has begun with the defense of Mauricio’s PhD thesis in February. Since then, we have kept extending Mauricio’s work and are preparing journal submissions.

While we have done some progress in this field, there are no concrete outcomes to present concerning optimistic approaches for model-based Bayesian Reinforcement Learning. Concerning POMDPs with information-based rewards, Mauricio’s PhD thesis presents strong theoretical results that allow – in principle – deriving efficient algorithms from state-of-the-art “point-based” POMDP solvers. This year we have put this idea into practice, implementing variants of PBVI, PERSEUS and HSVI.

Preliminary results have been published (in French) in JFPDA’13 [32]. A journal paper with complete theoretical and empirical results is under preparation.

#### 6.1.2. Adaptive Management with POMDPs

**Participant:** Olivier Buffet.

*Samuel Nicol, Iadine Chadès (CSIRO), Takuya Iwamura (Stanford University) 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.

This year, we have worked in the context of bird migratory pathways, in particular the East Asian-Australasian (EAA) flyway, which is modeled as a network whose nodes are land areas where birds need to stay for some time. An issue is that these land areas are threatened due to sea level rise. The adaptive management problem at hand is that of deciding in the protection of which land areas to invest money so as to preserve the migratory pathways as efficiently as possible.

The outcome of this work is a data challenge paper published at IJCAI’13 [27], which presents the problem at hand, describes its POMDP model, gives empirical results obtained with state-of-the-art solvers, and challenges POMDP practitioners to find better solution techniques.

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

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

*External collaborators: Christopher Amato (MIT), Arnaud Doniec (EMD), Charles Bessonnet (Telecom Nancy), Joni Pajarinen (Aalto University).*

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. In this work, we present an efficient methodology to solving decentralized stochastic control problems formalized as a DEC-POMDP or its subclasses. This methodology is three-fold: (1) it converts the original decentralized problem into a centralized problem from the perspective of a solution method that can take advantage of the total data about the original problem that is available during the online execution phase; (2) it shows that the original and transformed problems are equivalent; (3) it solves the transformed problem using a centralized method and transfers the solution back to the original problem. We applied this methodology in various different decentralized stochastic control problems.

Our results include the application of this methodology over DEC-POMDPs [20], [33]. We recast them into deterministic continuous-state MDPs, where states — called *occupancy states* — are probability distributions over states and action-observation histories of the original DEC-POMDPs. We also demonstrate the occupancy state is a sufficient statistic for optimally solving DEC-POMDPs. We further show the optimal value function is a piecewise-linear and convex function of the occupancy states. With these results as a background, we prove for the first time that POMDP (and more generally continuous-state MDP) solution methods can, at least in principle, apply in DEC-POMDPs. This work has been presented at IJCAI'2013 [20] and (in French) at JFPDA'2013 [33], and an in-depth journal article is currently under preparation. We have already extended the results we obtained for general DEC-POMDPs in the case of transition- and observation-independent DEC-MDPs. Of particular interest, we demonstrated that the occupancy states can be further compressed into a probability distribution over the states — the first sufficient statistic in decentralized stochastic control problems that is invariant with time. This work has been presented at AAMAS'2013 [21], and an in-depth journal article is currently under preparation.

We believe our methodology lays the foundation for further work on optimal as well as approximate solution methods for decentralized stochastic control problems in particular, and stochastic control problems in general.

#### 6.1.4. *Abstraction Pathologies in Markov Decision Processes*

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

*Jörg Hoffmann, former member of MAIA, is an external collaborator (from Saarland University).*

Abstraction is a common method to compute lower bounds in classical planning, imposing an equivalence relation on the state space and deriving the lower bound from the quotient system. It is a trivial and well-known fact that refined abstractions can only improve the lower bound. Thus, when we embarked on applying the same technique in the probabilistic setting, our firm belief was to find the same behavior there. We were wrong. Indeed, there are cases where every direct refinement step (splitting one equivalence class into two) yields strictly worse bounds. We give a comprehensive account of the issues involved, for two wide-spread methods to define and use abstract MDPs.

This work has been presented and published in the ICAPS-13 workshop on Heuristics and Search for Domain-Independent Planning (HSDIP) [29] and (in French) in JFPDA-13 [37].

#### 6.1.5. *Evolutionary programming for Policies Space exploration*

**Participants:** Amine Boumaza, Vincent Thomas.

Evolutionary Programming proposed by Fogel (initially introduced in 1966) is an approach to build an automaton optimizing a fitness function. Like other evolutionary algorithms, an initial population of automata is given, and the evolutionary programming algorithm will make this population evolve by progressively modifying automata (mutations) and keeping the most efficient ones in the next generation.

This process is close to the progressive construction by a policy iteration algorithm in a POMDP and we are currently investigating the links between these approaches.

This work has begun this year through an internship (Benjamin Bibler) and preliminary development has been made to solve the Santa Fe trail problem proposed by Koza (1992) which has become a benchmark to compare genetic and evolutionary programming approaches.

#### 6.1.6. *Evolutionary Learning of Tetris Policies*

**Participant:** Amine Boumaza.

Learning Tetris controllers is an interesting and challenging problem due to the fact of the size of its search space where traditional machine learning methods do not work and the use of approximate methods is necessary (see 6.1.10). In this work we study the performance of a direct policy search algorithm namely the Covariance Matrix Adaptation Evolution Strategy (CMAES). We also proposed different techniques to reduce the learning time, one of which is racing. This approach concentrates the computation effort on promising policies and quickly disregards bad ones in order to reduce the computation time. This approach allowed to

obtain policies of the same performance as those obtained without but at the fifth of the computation cost. The learned strategies are among the best performing players at this time scoring several millions of lines on average.

### 6.1.7. Evolutionary behavior learning

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

Evolutionary Robotics (ER) deals with the design of agent behaviors using artificial evolution. Within this framework, the problem of learning optimal policies (or controllers) is treated as a policy search problem in the parameterized space of candidate policies. The search for the optimal policies in this context is driven by a fitness function that associates a value to the candidate policy by measuring its performance on the given task.

The work shown here describes the results of the master's thesis of Iñaki Fernandèz which will be extended during a Ph.D. thesis started on october 2014.

- **Incremental policy learning with shaping.** Several methods have been proposed to accelerate the search for optimal policy in evolutionary robotics. In this work, we investigated the use of incremental learning and, more precisely, shaping, a well-known technique in behavioral psychology. The main idea is to learn to solve simple tasks and then exploit the learned behaviors to tackle increasingly harder tasks.

Our preliminary results show that the best performances are obtained either in the setups with shaping or in the control experiment where the task difficulty is maximal. Nevertheless, a closer look at the results indicates that the best controllers for the shaping setups are not obtained at the end of the evolution, but rather at an earlier stage. This means that, for these shaping techniques, the best controllers have learned to solve the task when its difficulty was at an easy level and their performance is maintained later when the task difficulty increases. Although this was unforeseen, the results seem promising and deserve further investigation.

- **Online evolutionary learning.** As opposed to traditional evolutionary robotics which treat the learning problem as an off-line, centralized process, online onboard distributed evolutionary algorithms [67], [55] consider the learning process as executed at the agent level in a decentralized way. In this sense, each agent has its own controller or genome which is locally broadcasted from agent to agent and the best performing ones survive and spread. This gene-centered view of evolution is inspired from the theory introduced by Richard Dawkins: The selfish gene.

The online aspect of the algorithms means that the agents are learning at the same time they are performing the task at hand. Another property that derives is that the agents are continuously learning which allows them to adapt to dynamically changing conditions and tasks. This is in opposition to the traditional view of evolutionary robotics (offline) where the outcome of evolution is tailored toward single task. Many challenging problems are raised in this framework and this thesis will address the problem of defining fitness functions that drive a swarm of agents to learn to solve a task. One other question is to study the dynamics of these algorithms both experimentally and theoretically using tools from distributed systems. Some promising work in this direction has been proposed [54].

### 6.1.8. Learning Bad Actions

**Participant:** Olivier Buffet.

*Jörg Hoffmann, former member of MAIA, and Michal Krajčanský are external collaborators from Saarland University.*

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?

This research work is conducted as part of Michal Krajňanský's master of science (to be defended in early 2014). Early experiments show encouraging results, and we consider participating in the learning track of the international planning competition in 2014.

### 6.1.9. Complexity of the Policy Iteration algorithm

**Participant:** Bruno Scherrer.

We have this year improved the state-of-the-art upper bounds for the complexity of a standard algorithm for solving Markov Decision Processes: Policy Iteration.

Given a Markov Decision Process with  $n$  states and  $m$  actions per state, we study the number of iterations needed by Policy Iteration (PI) algorithms to converge to the optimal  $\gamma$ -discounted optimal policy. We consider two variations of PI: Howard's PI that changes the actions in all states with a positive advantage, and Simplex-PI that only changes the action in the state with maximal advantage. We show that Howard's PI terminates after at most  $O\left(\frac{nm}{1-\gamma} \log\left(\frac{1}{1-\gamma}\right)\right)$  iterations, improving by a factor  $O(\log n)$  a result by Hansen et al. (2013), while Simplex-PI terminates after at most  $O\left(\frac{n^2m}{1-\gamma} \log\left(\frac{1}{1-\gamma}\right)\right)$  iterations, improving by a factor  $O(\log n)$  a result by Ye (2011). Under some structural assumptions of the MDP, we then consider bounds that are independent of the discount factor  $\gamma$ : given a measure of the maximal transient time  $\tau_t$  and the maximal time  $\tau_r$  to revisit states in recurrent classes under all policies, we show that Simplex-PI terminates after at most  $\tilde{O}(n^3m^2\tau_t\tau_r)$  iterations. This generalizes a recent result for deterministic MDPs by Post & Ye (2012), in which  $\tau_t \leq n$  and  $\tau_r \leq n$ . We explain why similar results seem hard to derive for Howard's PI. Finally, under the additional (restrictive) assumption that the state space is partitioned in two sets, respectively states that are transient and recurrent for all policies, we show that Simplex-PI and Howard's PI terminate after at most  $\tilde{O}(nm(\tau_t + \tau_r))$  iterations.

These results were presented at the JFPDA national workshop [36] and at the NIPS 2013 international conference [28].

### 6.1.10. Approximate Dynamic Programming and Application to the Game of Tetris

**Participant:** Bruno Scherrer.

*Victor Gabillon and Mohammad Ghavamzadeh are external collaborators (from the Inria Sequel EPI). Matthieu Geist is an external collaborator (from Supélec Metz).*

We present here three results: the first is a unified review of algorithms that are used to estimate a linear approximation of the value of some policy in a Markov Decision Process; the second concerns the analysis of a class of approximate dynamic algorithms for large scale Markov Decision Processes; the last is the successful application of similar dynamic programming algorithms on the Tetris domain.

In the framework of Markov Decision Processes, we have considered linear *off-policy* learning, that is the problem of learning a linear approximation of the value function of some fixed policy from one trajectory possibly generated by some other policy. We have made a review of *on-policy* learning algorithms of the literature (gradient-based and least-squares-based), adopting a unified algorithmic view. We have highlighted a systematic approach for adapting them to *off-policy* learning with *eligibility traces*. This lead to some known algorithms and suggested new extensions. This work has recently been accepted to JMLR and should be published at the beginning of 2014 [6].

We have revisited the work of Bertsekas and Ioffe (1996), that introduced  $\lambda$  policy iteration—a family of algorithms parametrized by a parameter  $\lambda$ —that generalizes the standard algorithms value and policy iteration, and has some deep connections with the temporal-difference algorithms described by Sutton and Barto (1998). We deepen the original theory developed by the authors by providing convergence rate bounds which generalize standard bounds for value iteration. We develop the theory of this algorithm when it is used in an approximate form. This work was published in JMLR [7].

Tetris is a video game that has been widely used as a benchmark for various optimization techniques including approximate dynamic programming (ADP) algorithms. A look at the literature of this game shows that while ADP algorithms that have been (almost) entirely based on approximating the value function (value function based) have performed poorly in Tetris, the methods that search directly in the space of policies by learning the policy parameters using an optimization black box, such as the cross entropy (CE) method, have achieved the best reported results. We have applied an algorithm we proposed in the past, called classification-based modified policy iteration (CBMPI), to the game of Tetris. Our experimental results show that for the first time an ADP algorithm, namely CBMPI, obtains the best results reported in the literature for Tetris in both small  $10 \times 10$  and large  $10 \times 20$  boards. Although the CBMPI's results are similar to those of the CE method in the large board, CBMPI uses considerably fewer (almost 1/6) samples (calls to the generative model) than CE. This work was presented at NIPS 2013 [26].

## 6.2. Ambient Intelligence And Robotic Systems

### 6.2.1. Robotic systems : autonomy, cooperation, exploration, robustness, assistance

#### 6.2.1.1. Local control based platooning

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

We consider decentralized control methods to operate autonomous vehicles at close spacings to form a platoon. We study models inspired by the flocking approach, where each vehicle computes its control from its local perceptions. We investigate different decentralized models in order to provide robust and callable solutions. Open questions concern collision avoidance, stability and multi-platoon navigation.

In order to reduce the tracking error (*i.e.* the distance between each follower's path and the path of its predecessor), we developed both an innovative approach [68] and a new lateral control law. This lateral control law reduces the tracking error faster than other existing control laws. An article, presenting this control law, its integration with a previously defined secure longitudinal control law [64] and the experimental results obtained with it, has been accepted to 2014 IEEE International Conference on Robotics and Automation.

#### 6.2.1.2. Map Matching

**Participants:** François Charpillat, Maan El-Badaoui-El-Najjar.

We addressed 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 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 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 [8].

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

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

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.

After working on a simple decentralized strategy to allow two queues of vehicles to share a single lane (presented in 2012, and this year in AATMO-13 [30]), 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. Preliminary results have been presented (in French) at RJCIA-13 [38], and more advanced work is under submission.

The next step is to coordinate the controller agents located in each of the network's intersections so as to create "green waves" that would improve the flows not just locally, but globally.

#### 6.2.1.4. *Living assistant Robot*

**Participants:** François Charpillat, Antoine Bautin, Abdallah Dib, Olivier Simonin.

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.

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

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

This work has been realized during the ANR Cart-O-matic project. Antoine Bautin has been hired by the Maia team for this project for a PhD. The main objective of the project was to design and build a multi-robot system able to autonomously map an unknown building. This work has been done in the framework of a French robotics contest called Defi CAROTTE organized by the General Delegation for Armaments (DGA) and the French National Research Agency (ANR). The scientific issues of this project deal with Simultaneous Localization And Mapping (SLAM), multi-robot collaboration and object recognition. The Maia Team has been mainly involved in multi-robot collaboration and navigation [13], [1], [11].

Nassim Kaldé, a new PhD student started last year in order to carry out the work done by Antoine Bautin. The new direction aims at addressing similar problems as the one we addressed in Cart-O-matic project but with dynamical environment, i.e. environment in which people are evolving with robots. An other point that Nassim Kaldé will address is social navigation, which is important for robot and human to coexist in a smart manner.

#### 6.2.1.6. *Features extraction for the control of redundant system with continuous sensori-motor space*

**Participants:** Alain Dutech, Thomas Moinel.

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

In collaboration with the CORTEX team and supported by a M2R internship, many questions related to learning the control of a complex (mono)-agent system with a continuous sensori-motor space are explored. For several reasons, the classical framework of Reinforcement Learning is not easily used in that context:

- the value function to be learned has to be encoded using features that are not known at start,
- because of the richness of the sensori-motor space, a random exploration scheme is unlikely to find the rewarded states that are needed by the learning process,
- exploiting what is learned is difficult as one would need to find the maximum of the value function *while* it is learned.



*Figure 1. Model of 2-joint human arm with 6 muscles.*

Our work is focused on a planar model of the human arm with 2 joints and 6 muscles (see figure 1 ). Control signals are the activity of the motor-neurons that alter the length of the muscles, and thus the forces applied on the joints. This system is redundant but also highly non linear as many aspects of the model are described by non-linear differential equations (our model is a slight improvement over the one of Li [59]). The task to learn is to reach different positions from given starting points.

We have studied a developmental learning process with a simple muscle activation pattern. The idea is to start the learning process in an artificially reduced sensori-motor space (using rough perception and motor capacities) and slowly increase the size and complexity of this space when interesting behaviors are learned. Our approach [60] gives results comparable to other developmental techniques and raises several important research questions. Our work showed that we need an abstraction mechanism in order to define or refine the features used in actions but also in perceptions. This is a very difficult challenge that is one of the keys to the understanding (and design) of cognition. There is also a need for stronger generalization capabilities in the function approximation used in the process.

In parallel, we are taking inspiration from the field of neurosciences, and particularly on the coupling between the cortex and the cerebellum in motor control. Models based on the work of Kaladjian [58] should help us understand what control signals are used by the brain apparatus and how the learning of gestures is organized between these two regions. Our long term goal is to design mechanisms for learning features abstraction in the sensori-motor space while being guided by the improvement in behavior performances.

## 6.2.2. Ambient intelligence

### 6.2.2.1. Personally Assisted Living

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

This action is supported by the Inria IPL Personally Assisted Living (PAL) which gathers 9 Inria teams associated with 6 research partners (technological, medical or social) which work together on three main issue guidelines: mobility assistance, assessing the degree of frailty of the persons, home activities analysis. The MAIA team is currently mainly involved in the 2 later topics, plus fall detection.

- **Evaluation of the degree of frailty of the elderly.** As argued in the famous paper of Fried *et al* [56] the estimation of frailty is highly significant to evaluate the risk of falls, disability, hospitalization and mortality. This issue is considered in Maia Team with different sensing devices: single RGB-D cameras [34], network of RGB-D cameras, sensing intelligent floor. One simple idea which is currently developed in the team is to determine either the center of mass of a person using one or several kinects, or the center of pressure and footsteps localization using an intelligent floor. The idea is to induce from these simple measures, the walking speed, the length of the steps and the position of the monitored persons.
- **People activity analysis.** The follow-up of the activity of elderly people over long period of time can be a good indicator of their well-being, but the evaluation of the behavior of a person at home is an open challenge.

To address this issue, we proposed this year a HMM based model capable of following simple activities such as sitting, walking, etc. An evaluation of this model has been conducted within a real smart environment with 26 subjects which were performing any of eight activities (sitting, walking, going up, squatting, lying on a couch, falling, bending and lying down). Seven out of these eight activities were correctly detected among which falling which was detected without false positives [24].

- **Fall detection.** Elderly fall is one of the major health issues affecting elderly people, especially at home. One of the objectives of the PhD work of Amandine Dubois 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. A simple and robust method based on the identification and tracking of the center of mass of people evolving in an indoor environment has been developed. 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



can surprisingly monitor the activity of a person with high accuracy and thus detect falls with very good accuracy without false positives [22], [23]. An experimental study, that is reported here, 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.

#### 6.2.2.2. *Interconnected intelligent tiles*

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

We are also involved in the development of a new innovative sensing device: a Pressure-Sensing Floor with LED lighting making possible to provide a new way for people to interact with their environment. Sensitive or intelligent floors have attracted a lot of attention during the last two decades for different applications going from interaction capture in immersive virtual environments to robotics or human tracking, fall detection or activity recognition. Different technologies have been proposed so far either based on optical fiber sensing, pressure sensing or electrical near field. In the Maia Team, we have developed a more sophisticated approach in which both computation and sensing is distributed within the floor. This floor is made up of interconnected intelligent tiles which can communicate with each other, have internal computation power, sense the environment activity (through four weight sensors, an accelerometer and a magnetometer) and can interact with users, robots or other sensor networks either by wireless/wire communication or through visual communication (each tile being equipped with 16 leds).

Several scientific challenges are open to us in the fields of decentralized spatial computing and in designing real application for assisting people suffering from loss of autonomy.

Some of these issues have been addressed this year. Mihai Andries, a PhD student, proposed two contributions demonstrating the relevancy of an intelligent floor such as the one we have developed. First contribution is about controlling a mobile robot through its interactions throughout the floor [10]. The second, least developed is about activity recognition of a person through its physical interaction on the floor. This approach has an important advantage compared to video based activity recognition: the privacy of people is without any doubt guaranteed. Let us mention too, the work of an internship student who developed a gait evaluation algorithm using the variation over time of the center of pressure that is sensed by the floor when one or several person walk over the floor.

#### 6.2.2.3. *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 cover simultaneously 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) POMDP formalisms to represent the state of the system (person and their internal states) and possible actions for the cameras, (2) a simulator provided and developed by Thales ThereSIS and (3) particle filtering approaches based on this simulator.

From a theoretical point of view, we are currently investigating how to use a deterministic simulator and to generate new particles in order to keep a good approximation of the posterior distribution.

## 6.3. Understanding and mastering complex systems

### 6.3.1. *Robustness of Cellular Automata and Reactive Multi-Agent Systems*

**Participants:** Olivier Bouré, Vincent Chevrier, 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 new definitions of asynchronism in lattice-gas cellular automata [3]. An experimental work was carried out and it was shown that the observation of an asynchronous version of a discrete model of swarm formation could help us gain insight on this well-studied model. The PhD thesis of O. Bouré [2] provides a detailed view of this work.

A study on the density classification problem, a well-studied problem of consensus in cellular automata, was carried out for infinite systems in 1D and 2D and for infinite trees [5], [4]. Positive results were provided and important conjectures were raised.

We proposed a survey on asynchronous cellular automata [25] and explained some of the difficulties in the classification of these objects [9].

In collaboration with colleagues from India, we proposed a complete characterisation of the reversibility of the set of the 256 Elementary Cellular Automata, which are known to be difficult to study in all generality [53]. We also proposed a mathematical analysis of the second-order phase transitions that are observed in the most simple asynchronous cellular automata [48]. We also coordinated a special issue on asynchronous cellular automata in the *Natural Computing* journal [41].

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

**Participants:** Vincent Chevrier, Tomas Navarrete.

*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. The thesis work of Tomas Navarrete 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 a heterogeneous system (different behaviors arise from different levels of interactions), how to evaluate the quality of the estimations obtained through 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.

Our architecture has been prototypically implemented in order to confront the basic ideas of 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, even when the initial conditions are supposed to drive the system to a state where no peer shares. We have also performed 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 [16] 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.

### 6.3.3. Multi-Modeling and multi-simulation

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

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

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

Models of Complex systems generally require different points of view (abstraction levels) at the same time in order to capture and to understand all the dynamics and the complexity. Consisting of different interacting parts, a model of a complex system also requires the joint and simultaneous use of modeling and simulation tools from different scientific fields.

We proposed the AA4MM meta-model [65] that solves the core challenges of multi-modelling 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 have made progress in the definition of multi-level modeling [15], [42]. We identified several facets of multi-level modeling and implemented them as different kinds of interactions in the AA4MM framework. A demonstration of these different multi-level couplings has been developed on a collective motion phenomenon.

In February started the MS4SG projet which involes MAIA, Madynes and EDF R&D on smart-grid simulation. A Phd thesis started on october 2013 by Julien Vaubourg in the MAIA team on the confrontation of the AA4MM principles against the specificities of smart-grid domain as a kind of complex system.

## OAK Project-Team

### 6. New Results

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

The team has continued developing expressive models and scalable algorithms for exploiting Semantic Web data, in particular RDF graphs, as well as rich corpora consisting of Web documents with semantic annotations.

We have studied efficient algorithms for answering RDF queries in the presence of schema (or semantic) constraints such as described through the RDF Schema language. The difficulty here consists of efficiently taking into account the data that is implicitly present in the RDF database due to semantic constraints, and which needs to be reflected in query results. We have identified the expressive database fragment of RDF, which extends previously identified fragments of the RDF specification by allowing more expressive schema and queries, and provided novel efficient algorithms for answering Basic Graph Pattern queries (a popular dialect of the standard SPARQL query language) over RDF graphs pertaining to the RDF Database Fragment. Our query answering algorithms take advantage of the processing power of a relational database management system while also reflecting RDF semantics [25].

The ability to exploit large corpora of heterogeneous RDF data requires tools for analyzing RDF content through the lenses of a specific user perspective, or user need. Such tools are commonplace in the context of relational data management, where data warehousing is a well-developed area, but lack completely in the realm of RDF. We have proposed a novel framework for building and exploiting all-RDF data warehouses [33] and have implemented this framework in a proof-of-concept platform [32]. A main contribution of this work is to preserve RDF graph structure, heterogeneity, and rich semantics from the base data to the analytical schema and analytical schema instance. Thus, our proposal is the first to allow the analysis of rich Semantic Web (RDF) data while preserving its rich content and semantics. For more information on this project, see <https://team.inria.fr/oak/warg/>.

We have investigated the usage of semantics as a way to enrich, interconnect, and interpret rich corpora of Web data. In particular, within the XR project, we had proposed in prior work the XR (XML+RDF) data model which integrates XML documents and RDF triples treating both as first-class citizens. One particular use of XR is to annotate nodes in XML documents, by RDF triples which may for instance describe their properties or state how nodes are semantically related to some concept or to each other. In [18] we describe the data model and core query language, make a comprehensive analysis of query evaluation algorithms, and describe extensive experiments carried within a fully implemented platform, as part of the PhD thesis of J. Leblay [12]. The XR platform was put to task in an application context related to digital journalism, where an XR content warehouse is continuously enriched through document analysis and annotation. This scenario has led to a software demonstration [24], [35] and a keynote tutorial [38]. In collaboration with A. Deutsch, we have extended the XR query language and provided query-view composition algorithms in [41].

#### 6.2. Massively Distributed Data Management Systems

Our work on the AMADA platform has shown how the different sub-systems of a popular cloud platform (namely, Amazon Web Services, or AWS in short) can be harnessed to build scalable stores and query evaluation engines for XML and RDF data. In [23], we propose and compare several storage and indexing strategies within AWS, and show that they help reduce not only query evaluation time but also the monetary costs associated to the exploitation of the AWS-based store, since the index helps direct queries only to the subsets of the data likely to have results for the query. Thus, the total effort (and the costs charged by AWS) in relation to the processing of a given query are reduced. A similar study focused mostly on RDF data management appears as a book chapter [40]. More information can be found at <http://cloak.saclay.inria.fr/research/amada/>.

Semantic Web data collections, that is, RDF graphs, may be very voluminous since RDF natively enables connections between different RDF databases (which may have been produced independently and in ignorance of each other) through the usage of common URIs (resource identifiers) in two or more databases. To scale up to such large volumes, we have developed CliqueSquare, a novel platform for storing and querying RDF graphs in a MapReduce-based architecture such as Hadoop. We have described the storage and query algorithm in [34]. Our analysis of existing frameworks and algorithms for managing large RDF graphs in a highly distributed environment has led to the tutorial [27].

Large-scale distributed processing of complex data was considered from a different perspective in our Delta project. Here, we considered the setting where one data source publishes new data items at a very high rate, and numerous clients subscribe to some of the updates by means of queries that must be matched by the published items. In this setting, the source may quickly become the bottleneck due to limitations in its capacity to match the published item against the subscription and/or to send the matching updates. We propose a fully automated approach for distributing the data dissemination effort across the network of subscribers, by identifying some which act as secondary data sources for others, in a peer-to-peer fashion. This distributed dissemination network is chosen so as to optimize a combination of overall dissemination costs and data propagation latency; since the space of options has daunting complexity, approximate algorithms involving Binary Integer Programming techniques were proposed in [20], [37], [42], and concluded in the PhD thesis of A. Katsifodimos [11].

### 6.3. Advanced Algorithms for Efficient XML processing

In 2013, several research works of the team focusing on advanced algorithms for processing XML data have been finalized and concluded through prestigious journal publications.

A first line of work concerned the usage of materialized views to speed up the evaluation of complex XML queries. In our previous work we had demonstrated that such views may bring up very significant speed-up factors of several orders of magnitude. However, materialized views need to be kept up to date when the underlying database changes. In [14] we have described efficient algorithms for updating materialized views expressed in a rich dialect of XQuery, the standard query language for XML.

A second class of work was concerned with XML static type analysis, in particular with the crucial problem of deciding XML type inclusion, that is: whether any XML tree of type  $\tau_1$  is also of type  $\tau_2$  where  $\tau_1, \tau_2$  are XML types with interleaving and counting (currently adopted by main stream schema languages). For these types, inclusion is EXPSPACE-complete. We have defined and formally studied a quadratic subtype-checking algorithm for the case where the right-hand side type  $\tau_2$  meets some restrictions on symbol occurrences and the use of counting. These restrictions are often met by human-designed types, so our technique perfectly fits the needs of typical XML type-checking algorithms, which frequently require to check for inclusion a machine-generated subtype  $\tau_1$  against a human-defined supertype  $\tau_2$ . Our approach has been validated by extensive experimental results [16]. In addition, we have devised and formally studied an alternative algorithm, still for the asymmetric case where  $\tau_2$  is restricted, based on structural, top-down analysis of types expression. This algorithm is almost linear: it has a linear-time backbone, and resorts to the above quadratic approach for some specific parts of the compared types. Our experiments show that this new algorithm is much faster than the quadratic one and that it typically runs in linear time, hence it can be used as a building block for a practical type-checking compiler for XML programs and queries [15].

Third, we have completed and concluded our work on type-based document projection for efficient XML data management. The idea here is to restrict XML documents, prior to evaluating a query over them, to only those parts of the document that the query actually needs to consult. We provide algorithms for determining such document parts and experimentally demonstrate the benefits of such techniques, in [13].

Finally, we have devised a system that is able to process both queries and updates on very large XML documents [22]. As observed in recent works, such very large documents are generated and processed in several contexts, in particular in those involving scientific data and logs. Our system supports a large fragment of XQuery and XUF (XQuery Update Facility). The system exploits dynamic and static partitioning to

distribute the processing load among the machines of a MapReduce cluster. The proposed technique applies when queries and updates are iterative, i.e., they iterate the same query/update operations on a sequence of subtrees of the input document. From our experience many real world queries and updates actually meet this property. Our partitioning technique is schema-less, as the presence of a user-supplied schema is not required; indeed, this technique only relies on path information extracted from the input query/update. Experiments conducted on a 8-machine Hadoop cluster have demonstrated that the system is able run both iterative queries and updates on quite large documents.

## **6.4. Data Transformation Management**

With the increasing complexity of data processing queries, for instance in applications such as relational data analysis or integration of Web data (e.g., XML or RDF) comes the need to better manage complex data transformations. This includes systematically verifying, maintaining, and testing the transformations an application relies on. In this context, Oak has focused on verifying the semantic correctness of a declarative program that specifies a data transformation query, e.g., an SQL query. To this end, we have investigated how to leverage data provenance (the information of the origin of data and the query operators) for query debugging. More specifically, we developed and implemented novel algorithms to explain why data is missing from the result of a relational query. As opposed to our previous work, which produced explanations based on the available source data, our new algorithms return explanations based on query operators [31] or both [26].

## **6.5. Social Data Management**

We considered top-k query answering in social tagging systems, also known as folksonomies, a problem that requires a significant departure from existing, socially agnostic techniques. In a network-aware context, one can and should exploit the social links, which can indicate how users relate to the seeker and how much weight their tagging actions should have in the result build-up. Our solutions addressed the main drawbacks of previous approaches. With respect to applicability and scalability, we avoid expensive and hardly updatable pre-computations of proximity values. With respect to efficiency, we show that our algorithm is instance optimal in the existing techniques. Our main results in this direction have been presented recently in [29], [28], [21].

## ORPAILLEUR Project-Team

# 6. New Results

## 6.1. 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, Mohsen Sayed, Yannick Toussaint.

**Keywords:** formal concept analysis, relational concept analysis, pattern structures, frequent itemset, association rule, graph mining, sequence mining, skyline

Formal Concept Analysis, together with itemset search and association rule extraction, 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 biological or chemical data or textual documents, involving objects with multi-valued attributes (e.g. domains or intervals), n-ary relations, sequences, trees and graphs.

### 6.1.1. FCA and variations: RCA and Pattern Structures

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 [10]. 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 has an important role in KDDK, especially in text mining [86], [85].

Another extension of FCA is based on Pattern Structures (PS) [92], which allows to build a concept lattice from complex data, e.g. nominal, numerical, and interval data. In [100], pattern structures are used for building a concept lattice from interval data. Since then, we worked on a some experiments involving pattern structures, namely sequence mining [41], information retrieval [48] and functional dependencies [38]. 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 [99]: 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.

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 [37], [47], [46], [44]. 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.

### 6.1.2. Advances in mining complex data: sequences and healthcare trajectories

Sequence data is widely used in many applications. Consequently, mining sequential patterns and other types of knowledge from sequence data has become an important data mining task. The main emphasis has been on developing efficient mining algorithms and effective pattern representation. The most frequent sequences generally provide a 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, the general idea is to extract patterns whose characteristic on a given measure such as the support strongly deviates from its expected value under a null model. The frequency of a pattern is considered as a random variable, whose distribution under the null model has to be calculated or approximated. Then, the significance of the pattern is assessed through a statistical test that compares the expected frequency under the null model to the observed frequency. One of



the key-points of this family of approaches is to choose an appropriate null model. It will ideally be a trade-off between adjustment to the data and simplicity: the model should capture some characteristics of the data, to integrate prior knowledge, without overfitting, to allow for relevant patterns discovery. We introduced a rigorous and efficient approach to mine statistically significant, unexpected patterns in sequences of itemsets. Experiments on sequences of replays of a video game demonstrated the scalability and the efficiency of the method to discover unexpected game strategies. This work was successfully published as an international conference paper [8].

Other work on sequences is in concern with patient trajectories, i.e. the “path” of a patient during its illness. 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 a set of recent works with Elias Egho, a PhD candidate, we present new algorithms to extract patient trajectory patterns with different levels of granularity by relying on external taxonomies [52]. Our algorithms rely on the general FCA framework to formalize the general notion of multidimensional healthcare trajectories. We also continued working on the complex notion of sequences or trajectory similarity measures. We show the interest of our approaches with the analysis of trajectories of care for colorectal cancer using data from the French healthcare information system (see also [41]).

### 6.1.3. 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 and the implementation of a system for ontology engineering from heterogeneous ontological resources [58]. This methodology is based on both FCA and RCA, and was previously successfully applied in contexts 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.

## 6.2. KDDK in Life Sciences

**Participants:** Yasmine Assess, Emmanuel Bresso, Adrien Coulet, Marie-Dominique Devignes, Anisah Ghoorah, Bernard Maigret, Amedeo Napoli, Gabin Personeni, David Ritchie, Mohsen Sayed, Malika Smaïl-Tabbone, My Thao Tang, Mohsen Sayed, Yannick Toussaint.

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 the Life Sciences is developed in one main direction which is in concern with the use of bio-ontologies to improve KDDK but also information retrieval, access to the so-called “Linked Open Data” and data integration.

### 6.2.1. Using ILP for the characterization and prediction of drug side-effect profiles

Inductive Logic Programming (ILP) is a learning method which allows expressive representation of the data and produces explicit first-order logic rules [89]. We applied ILP for understanding drug side-effects. Indeed, late appearance of adverse side effects during clinical trials constitute the main reason for stopping the drug development process which is very costly [1]. Improving our ability to understand drug side effects is necessary to reduce this inconvenience. Moreover, it can contribute to design safer drugs and anticipate the appearance of yet unreported side effects of approved drugs. Today, most investigations deal with prediction of single side effects and overlook possible combinations.



In our study, drug annotations are collected from the SIDER and DrugBank databases. Terms describing individual side effects reported in SIDER are clustered with the IntelliGO semantic similarity measure into term clusters (TCs) [83]. Maximal frequent itemsets are extracted from the resulting  $drug \times TC$  binary table, leading to the identification of what we call side-effect profiles (SEPs). A SEP is defined as the longest combination of TCs which are shared by a significant number of drugs. Frequent SEPs are explored on the basis of integrated drug and target descriptors using two machine learning methods: decision-trees and ILP. Learning efficiency is evaluated by cross-validation and direct testing with new molecules. Comparison of the two methods shows that the ILP displays a greater sensitivity than decision trees. Although both methods yield explicit models, ILP is able to exploit not only drug properties but also background knowledge, thereby producing rich and expressive rules.

### 6.2.2. Functional classification of genes

The IntelliGO measure computes semantic similarity between genes in taking into account domain knowledge in Gene Ontology (GO) [83]. IntelliGO is used for functional clustering of a set of genes, i.e. based on functional annotations of these genes. For example, a gene set of interest may include genes showing the same expression profile.

A functional clustering method based on IntelliGO was tested on four benchmarking datasets consisting of biological pathways (KEGG database) and functional domains (Pfam database) [90]. A follow-up of this study was motivated by the fact that the IntelliGO measure, like most of the biological similarity measures, does not verify “triangle inequality” and thus is not a mathematical distance. Interestingly, specific spectral clustering techniques can be used for improving the clustering of the objects for which exists a pairwise (dis-)similarity matrix [115], [125]. Spectral clustering techniques make use of the eigenvalues of this (dis-)similarity matrix to perform dimension reduction before clustering in fewer dimensions. We have conducted a comparative and large-scale gene clustering evaluation using the IntelliGO measure and reference sets. Our results showed an improvement of the clustering quality with “constant-shift spectral clustering” [63].

### 6.2.3. 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 [98]. This repository contains annotations of 36 biomedical databases annotated with concepts of more than 200 ontologies of the BioPortal<sup>3</sup>. In 2012, we compared the annotations of a database of biomedical publications (Medline) with two databases of scientific funding (Crisp and ResearchCrossroads) to profile disease research [105]. 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?

In 2013, we proposed an adaptation of FCA techniques, namely pattern structures, to explore the annotations of biomedical databases [2]. 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. Then, 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.

In another context, a related work was carried out in the Kolflow project (see 8.2.1.4). We proposed an interactive environment based on Formal Concept Analysis which makes possible a simultaneous enrichment of semantic annotations of medical texts and of the ontology of medical domain [66], [59].

### 6.2.4. Analysis and interpretation of sequential patterns with Linked Open Data

---

<sup>3</sup><http://biportal.bioontology.org/>

Linked Data is a set of principles and technologies that rely on the architecture of the Web (URIs and links) to share, model and integrate data. The basic idea is that data objects (e.g., a surgical procedure) are identified by web addresses (URIs), and the information attached to these objects are represented through links to values or other URIs representing other objects.

Considering the potential development and availability of biomedical Linked Data, we investigated it as a source of additional information to support the interpretation of the results of a data mining process, such as sequential pattern discovery. We developed a system using several linked data endpoints to collect descriptive dimensions about the items that constitute sequential patterns. These dimensions are used to automatically classify with Formal Concept Analysis the extracted patterns, thus generating a structure that can support exploration and navigation into the results of the data mining step [55].

### 6.3. Structural Systems Biology

**Participants:** Marie-Dominique Devignes, Anisah Ghoorah, Van-Thai Hoang, 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.

#### 6.3.1. Accelerating protein docking calculations using graphics processors

We have recently adapted the *Hex* protein docking software [113] to use modern graphics processors (GPUs) to carry out the expensive FFT part of a docking calculation [114]. Compared to using a single conventional central processor (CPU), a high-end GPU gives a speed-up of 45 or more. This software is publicly available at <http://hex.loria.fr>. A public GPU-powered server has also been created (<http://hexserver.loria.fr>) [106]. The docking server has performed some 14,000 docking runs during 2013.

Our docking work has facilitated further developments on modeling the assembly of multi-component molecular structures using a particle swarm optimization technique [123], and on modeling protein flexibility during docking [122]. In 2013, in collaboration with the Nano-D team at Inria Grenoble, we developed a new docking algorithm called “DockTrina” [31], which can rapidly model trimers of protein structures by combining multiple pair-wise docking results from *Hex*. We also used *Hex* successfully to model a challenging protein complex containing water molecules at the protein-protein interface [29].

#### 6.3.2. 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 [95]. KBDOCK is available at <http://kbdock.loria.fr>. KBDOCK combines coordinate data from the Protein Data Bank [87] with the Pfam protein domain family classification [91] 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 [94] for template-based docking using 73 complexes from the Protein Docking Benchmark [97]. We recently presented results obtained using KBDOCK at the CAPRI conference on protein docking in Utrecht [21]. In 2013, we updated KBDOCK with the latest data from Pfam and the Protein Data Bank. An article describing the new version of KBDOCK was accepted by the Database Issue of Nucleic Acids Research [6].

#### 6.3.3. Kpax: A new algorithm for protein structure alignment

We have developed a new protein structure alignment approach called Kpax [112]. 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 [124]. The Kpax program is available at <http://kpax.loria.fr/>. The Kpax program is now used heavily behind the scenes in the new KBDock web server [6] to find structural templates for docking which might be beyond the reach of sequence-based homology modeling approaches.

#### 6.3.4. *gEMpicker and gEMfitter: GPU-accelerated tools for cryo-electron microscopy*

Solving the structures of large protein assemblies is a difficult and computationally intensive task. Multiple two-dimensional (2D) images must be processed and classified to identify protein particles in different orientations. These images may then be averaged and stacked to deduce the three-dimensional (3D) structure of a protein. In order to help accelerate the first of these tasks we have recently developed a novel and highly parallel algorithm called “gEMpicker” which uses multiple graphics processors to detecting 2D particles in cryo-electron microscopy images [112]. We have also developed a 3D shape matching algorithm called “gEMfitter” which also exploits graphics processors, and which will provide a useful tool for the final 3D assembly step [112]. Both programs have been made publicly available at <http://gem.loria.fr/>.

#### 6.3.5. *DOVSA: Developing new algorithms for virtual screening*

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 [110]. As a continuation of this project, and in collaboration with colleagues from the University of Bari in Italy, we recently published a review on drug discovery relating to the GPCR receptor proteins [15]. We also published a book chapter describing the ParaFit program for fast spherical harmonic shape matching [70].

### 6.4. Around the Taaable research project

**Participants:** Valmi Dufour-Lussier, Emmanuelle Gaillard, Laura Infante Blanco, 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 [69] (<http://taaable.fr>) has been originally created as a challenger of the Computer Cooking Contest (ICCB Conference). A candidate to this contest is a system whose goal is to solve cooking problems on the basis of a recipe book (common to all candidates), where each recipe is a shallow XML document with an important plain text part. The size of the recipe book (about 1500 recipes) prevents from a manual indexing of recipes: this indexing is performed using semi-automatic techniques.

Beyond its participation to the CCCs, 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 is used to perform adaptation of recipe to user constraints. The reasoning process uses a cooking domain ontology (especially hierarchies of classes) and adaptation rules. The knowledge base used by the inference engine is encoded within a semantic wiki, which contains 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 for the adaptation of recipe preparations [3], using one of the engines included in the library Revisor (cf. § 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 [68] [5]. 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, ensuring the quality of results.

## 6.5. Some results in graph theory

**Participants:** Amedeo Napoli, Chedy Raissi, Jean-Sébastien Sereni, Mario Valencia.

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

### 6.5.1. 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 vertices, 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 vertices, 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  $n$  vertices, there is a color  $c$  and a set  $X$  of 4-vertices such that at least  $2n/3$  vertices 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.

### 6.5.2. 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.

### 6.5.3. Other aspects on graph coloring and clustering

Since September 2013, Mario Valencia has obtained a one year 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. More precisely, this study relies on families of graphs having a “good” decomposition as cographs and chordal graphs, and then, and on the analysis of the adaptation of these two families of graphs within a clustering activity.

Other research dimensions are dealing with algorithmic aspects of some variations of the classical graph coloring problem.

- Packing colorings of graphs where we need to color the vertices of a graph in such a way that vertices having a same color  $c$  should be at a distance at least equal to  $c + 1$  in the graph. With P. Torres, a postdoc student, we have obtained some upper bounds for the packing chromatic number of hypercubes graphs of dimension  $n$ , denoted by  $Q_n$ , and we have computed exactly this parameter for this family of graph for  $n = 6, 7, 8$ , extending previous results known for  $n = 2, 3, 4, 5$  [35].
- $(k, i)$ -coloring of graphs, which is a generalization of a  $k$ -tuple coloring of graphs: given positive integers  $k$  and  $i$ , we want to affect to each vertex a  $k$ -set of colors such that the intersection of the  $k$ -sets affected to adjacent vertices has cardinality at most equal to  $i$ . With F. Bonomo, I. Koch, and G. Duran, we have found a linear time algorithm for this problem on cycles and cacti graphs. Moreover, we have obtained an interesting equivalence between this problem on complete graphs and a problem on weighted binary codes.
- $b$ -coloring of graphs, where we need to color the vertices of a graph in such a way that in each color class  $j$  there exists at least one vertex  $x_j$  adjacent to at least one vertex in all the other color classes. The goal of this problem is to maximize the number of colors under such a constraint (i.e. the  $b$ -chromatic number of a graph). With F. Bonomo, O. Schaudt and M. Stein, we have shown that  $b$ -coloring is NP-hard on co-bipartite graphs and polytime solvable on tree-cographs [77].

## SMIS Project-Team

# 6. New Results

## 6.1. Minimum Exposure

**Participants:** Nicolas AnCIAUX, Marouane Fazouane, Benjamin Nguyen [correspondent], Michalis Vazirgianis.

When users request a service, the service provider usually asks for personal documents to tailor its service to the specific situation of the applicant. For example, the rate and duration of consumer's loans are usually adapted depending on the risk based on the income, assets or past lines of credits of the borrower. In practice, an excessive amount of personal data is collected and stored. Indeed, a paradox is at the root of this problem: service providers require users to expose data in order to determine whether that data is needed or not to achieve the purpose of the service. We explore a reverse approach, where service providers would publicly describe the data they require to complete their task, and where software (placed, depending on the context, on the client, on the server, or in a trusted hardware component) would use those descriptions to determine a minimum subset of information to expose.

Following our 2012 seminal works on the general Minimum Exposure framework, we have pursued its general study in 2013 [15], [29]. We have also developed a prototype system, using a low powered and highly secure smartcard [21], which is used to support hidden decision rules.

## 6.2. Flash-Based Data Management

**Participants:** Nicolas AnCIAUX, Matias Bjørling, Philippe Bonnet, Luc Bouganim [correspondent], Niv Dayan, Philippe Pucheral.

Mass-storage secure portable tokens are emerging and provide a real breakthrough in the management of sensitive data. They can embed personal data and/or metadata referencing documents stored encrypted in the Cloud and can manage them under holder's control. Mass on-board storage requires efficient embedded database techniques. These techniques are however very challenging to design due to a combination of conflicting NAND Flash constraints and scarce RAM constraint, disqualifying known state of the art solutions. To tackle this challenge, we proposed a log-only based storage organization and an appropriate indexing scheme, which (1) produce only sequential writes compatible with the Flash constraints and (2) consume a tiny amount of RAM, independent of the database size [13].

Solid State Drives (SSDs) are a moving target for system designers: they are black boxes, their internals are undocumented, and their performance characteristics vary across models. There is no appropriate analytical model and experimenting with commercial SSDs is cumbersome, as it requires a careful experimental methodology to ensure repeatability. Worse, performance results obtained on a given SSD cannot be generalized. Overall, it is impossible to explore how a given algorithm, say a hash join or LSM-tree insertions, leverages the intrinsic parallelism of a modern SSD, or how a slight change in the internals of an SSD would impact its overall performance. In 2013, we worked on a new SSD simulation framework, named EagleTree, which addresses these problems, and enables a principled study of SSD-Based algorithms. We published a demonstration on EagleTree at VLDB'13 [20]. The demonstration scenario illustrates the design space for algorithms based on an SSD-based IO stack, and shows how researchers and practitioners can use EagleTree to perform tractable explorations of this complex design space.

## 6.3. Secure Global Computing on Asymmetric Architecture

**Participants:** Benjamin Nguyen [correspondent], Philippe Pucheral, Cuong Quoc 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 primarily studied 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]. This work is an extension of [26]. A vulgarization paper on the scientific and societal challenges related to PPDP techniques has been published in a newspaper [24]. We are now trying to support general SQL queries in this same execution context. We concentrate first 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. Cost models and experiments demonstrate that this approach can scale to nationwide infrastructures [23], [42]. This work is part of Cuong Quoc To's Ph.D. thesis started in sept. 2012.

## 6.4. Trusted Cells

**Participants:** Nicolas AnCIAUX, Philippe Bonnet, Luc Bouganim, Benjamin Nguyen, Pilippe Pucheral [correspondent], Iulian Sandu Popa.

With the convergence of mobile communications, sensors and online social networks technologies, we are witnessing an exponential increase in the creation and consumption of personal data. Such data is volunteered by users, automatically captured by sensors or inferred from existing data. Today, there is a wide consensus that individuals should have increased control on how their personal data is collected, managed and shared. Yet there is no appropriate technical solution to implement such personal data services: centralized solutions sacrifice security for innovative applications, while decentralized solutions sacrifice innovative applications for security. In this work, we argue that the advent of secure hardware in all personal IT devices, at the edges of the Internet, could trigger a sea change. We propose the vision of trusted cells: personal data servers running on secure smart phones, set-top boxes, secure portable tokens or smart cards to form a global, decentralized data platform that provides security yet enables innovative applications. We motivate our approach, describe the trusted cells architecture and define a range of challenges for future research in a paper published at CIDR'13 (Int. Conf on Innovative Data Systems Research). This work was based on a thorough analysis of existing and potential threats on personal data, which led to a tutorial on data privacy [18], [30].

In parallel, we revisited the Trusted Cells vision to the context of Least Developed Countries (LDCs). 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. We propose, Folk-enabled Information System (Folk-IS), a new paradigm based on a fully decentralized and participatory approach, where each individual implements a small subset of a complete information system without the need for infrastructure. As trusted cells, Folk-IS builds upon the emergence of highly secure, portable, low-cost storage and computing devices, called hereafter Smart Tokens. Here, however, the focus is on the low cost of ownership, deployment and maintenance, and on the absence of a networked infrastructure. With Folk-IS, and thanks to 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 [17].



## TYREX Team

# 6. New Results

## 6.1. Multimedia Models and Formats

Modeling and authoring web content including rich media and interactions is still an open problem. We have drawn a reference state of the art of this area in [17]. We have also contributed to the cultural heritage domain through the experimentation of a multimedia production chain for digitized theatre performances based on semantic annotations [12].

In the context of the Claire project (see section 7.1.1 ), the results we obtained in 2013 in this area are:

- providing a data model which is format agnostic to cope with existing and future rendering systems. More specifically, we specified a chaptering component that includes the structuration and navigation features for continuous media such as video.
- prototyping a web environment for authoring such rich media (see section 5.2 . This authoring services are built as a contribution of the Mozilla Popcorn Maker project (**Popcorn Maker**).
- experimenting this environment for the production of the multimedia part of the first MOOC developed by OpenClassrooms, our main partner in this project.

## 6.2. XML Processing

In the area of XML processing, we obtained new results in several directions:

- We showed how to translate Schematron descriptions into the tree logic [15];
- We built the first IDE equipped with path reasoning capabilities [13];
- We showed that a whole class of logical combinators (or “macros”) can be used as an intermediate language between the query language and the logical language [20]. This provides a gain in terms of succinctness for the logical formalism.
- We continued our work on a novel technique and a tool for the static type-checking of XQuery programs, using backward type inference.
- We made preliminary investigations on how to support backward navigation axes in the static type checking for XQuery [18].
- In a joint work with the Exmo team, we benchmarked solvers for deciding the problem of query containment for fragments of SPARQL [14].

We briefly review these results below.

### 6.2.1. Rule-Based Validation à la Schematron

One major concept in web development using XML is validation: checking whether some document instance fulfills structural constraints described by some schema. Over the last few years, there has been a growing debate about XML validation, and two main schools of thought emerged about the way it should be done. On the one hand, some advocate the use of validation with respect to complete grammar-based descriptions such as DTDs and XML Schemas. On the other hand, motivated by a need for greater flexibility, others argue for no validation at all, or prefer the use of lightweight constraint languages such as Schematron with the aim of validating only required constraints, while making schema descriptions more compositional and more reusable.

We built a compiler for Schematron [15]. This compiler takes a Schematron description as input and generates the corresponding constraints as a logical formula. We showed that validators used in each of these approaches share the same theoretical foundations, meaning that the two approaches are far from being incompatible. Our findings include that modal logic can be seen as a unifying formal ground for the construction of robust and efficient validators and static analyzers using any of these schema description techniques. This reconciles the two approaches from both a theoretical and a practical perspective, therefore facilitating any combination of them.

### 6.2.2. *Integrated Development Environments with Path Reasoning Capabilities*

One of the challenges in web development is to help achieving a good level of quality in terms of code size and runtime performance, for popular domain-specific languages such as XQuery, XSLT, and XML Schema. We presented the first IDE augmented with static detection of inconsistent XPath expressions that assists the programmer for simplifying the development and debugging of any application involving XPath expressions [13]. The tool is based on newly developed formal verification techniques based on expressive modal logics, which are now mature enough to be introduced in the process of software development. We further develop this idea in the context of XQuery for which we introduce an analysis for identifying and eliminating dead code automatically. This proof of concept aims at illustrating the benefits of equipping modern IDEs with reasoning capabilities.

### 6.2.3. *Logical Combinators for Rich Type Systems*

A popular technique in the static analysis for query languages relies on the construction of compilers that effectively translate queries into logical formulas. These formulas are then solved for satisfiability using an off-the-shelf satisfiability solver. A critical aspect in this approach is the size of the obtained logical formula, since it constitutes a factor that affects the combined complexity of the global approach.

We showed that a whole class of logical combinators (or “macros”) can be used as an intermediate language between the query language and the logical language [20]. Those logical combinators provide an exponential gain in succinctness over the corresponding explicit logical representation, yet preserve the typical exponential time complexity of the subsequent logical decision procedure. This opens the way for solving a wide range of problems such as satisfiability and containment for expressive query languages in exponential-time, even though their direct formulation into the underlying logic results in an exponential blowup of the formula size, yielding an incorrectly presumed two-exponential time complexity. We illustrated this from a very practical point of view on a few examples such as numerical occurrence constraints and tree frontier properties, which are concrete problems found in the XML world.

### 6.2.4. *Backward type inference for XQuery*

We have continued our work on the design of a novel technique for static type-checking of XQuery programs based on backward type inference. The tool looks for errors in the program by jointly analyzing the source code of the program, input and output schemas that respectively describe the sets of documents admissible as input and as output of the program. The crux and the novelty of our results reside in the joint use of backward type inference and a two-way logic to represent inferred tree type portions. This allowed us to design and implement a type-checker for XQuery which is more precise and supports a larger fragment of XQuery than the approaches previously proposed in the literature; in particular compared to the only few actually implemented static type-checkers such as the one in Galax. The whole system uses compilers and a satisfiability solver for deciding containment for two-way regular tree expressions. Our tool takes an XQuery program and two schemas  $S_{in}$  and  $S_{out}$  as input. If the program is found incorrect, then it automatically generates a counter-example valid w.r.t.  $S_{in}$  and such that the program produces an invalid output w.r.t.  $S_{out}$ . This counter-example can be used by the programmer to fix the program.

### 6.2.5. *XQuery and Static Typing: Tackling the Problem of Backward Axes*

XQuery is a functional language dedicated to XML data querying and manipulation. As opposed to other W3C-standardized languages for XML (e.g. XSLT), it has been intended to feature strong static typing.

Currently, however, some expressions of the language cannot be statically typed with any precision. We argue that this is due to a discrepancy between the semantics of the language and its type algebra: namely, the values of the language are (possibly inner) tree nodes, which may have siblings and ancestors in the data. The types on the other hand are regular tree types, as usual in the XML world: they describe sets of trees. The type associated to a node then corresponds to the subtree whose root is that node and contains no information about the rest of the data. This makes navigational expressions using ‘backward axes,’ which return e.g. the siblings of a node, impossible to type.

In [18], we discussed how to solve this discrepancy and proposed a compromise: to use extended types representing possibly inner tree nodes in some key parts of a program, and to cut out the subtrees from their original context in the rest.

### 6.2.6. Semantic Web queries and $\mu$ -calculus

Querying the semantic web is mainly done through the SPARQL language or its extensions through paths and entailment regimes. Query containment is the problem of deciding if the answers to a query are included in those of another query for any queried database [4], [3]. This problem is very important for query optimization purposes. In the SPARQL context, it can be equally useful for distributing federated queries or for implementing schema-based access control. In order to experimentally assess implementation strengths and limitations, we provided a first SPARQL containment test benchmark. We studied the query demographics on DBPEDIA logs to design benchmarks for relevant query containment solvers. We tested available solvers on their domain of applicability on three different benchmark suites [14]. (i) tested solutions are overall functionally correct, (ii) in spite of its complexity, SPARQL query containment is practicable for acyclic queries, (iii) state-of-the-art solvers are at an early stage both in terms of capability and implementation.

This work has been developed in collaboration with the EXMO team. The benchmarks, results and software are available at <http://sparql-qc-bench.inrialpes.fr>.

## 6.3. Mixed Reality Environment

The concept of Mixed Reality comes from the fact that the real-virtual dichotomy is not sharp. Augmented Reality (AR) mode refers to all cases in which the auditory or visual display of a real environment is augmented by virtual sound or graphic objects. Pedestrian navigation is one of the numerous applications that fit into this field. Depending on the real speed of the user and on the real environment in which he moves (inside a building, ...), the system is augmented with synthetic audio instructions and points of interest. **OpenStreetMap** format has been extended to support navigation authoring and information related to the various passive or active location providers supported by IXE such as PDR, GPS and NFC.

### 6.3.1. Navigation Authoring

We defined a cue-based XML language (A2ML, for Advanced Audio Markup Language) using SMIL for internal and external synchronization of sound objects. A2ML is specified by a RELAX-NG grammar. A rule-based selector mechanism allows defining style sheets for OpenStreetMap (OSM) elements. This auditory display together with TTS makes our IXE browser accessible to visually impaired people. Format and Delivery for Mixed Reality Content IXE is based on an extended OSM data format with triggering zones, relations or groups with specific semantics and nodes or POIs whose URIs refer to content expressed in HTML5 and A2ML. Content delivery can be of two types, push or pull. Push content is coming from POIs which trigger when the user enters a new zone. This kind of content is very useful for navigation. We support it through a triggering specification that is inserted in the OpenStreetMap document. We use style sheets with rules to specify both the audio and visual rendering of the various types of OSM nodes. Pull content allows users to search detailed information about the artifacts that are located in the content referenced by the POI. Most of the time, this content is described using HTML5 and A2ML.

### **6.3.2. Location Provider Fusion**

Pedestrian navigation can be done with several sensors. GPS locations are better for outside locations, PDR is useful to guide people indoor, but we can also use NFC tags, user proprioception, wifi... Our researches focus on a smart fusion of providers depending on sensor accuracy and on the context in which the person moves. We start by using Kalman Filter to smooth locations and disable jumps during the walk. These algorithms have been successfully tested during Venturi Y2 demo.

### **6.3.3. Map Rendering**

We worked on offline map rendering around two solutions. The first one is based on an open source Android project called [Mapsforge](#); it provides a tile generation mechanism from a given OpenStreetMap file and a tile caching system for fast rendering on mobile devices. We mainly enhanced the open source project by increasing the zoom level limitation (21 by default) to 24 for displaying indoor maps. The other solution on which we worked is SVG-oriented and based on [OpenLayers](#) (dedicated to web browsers). As the rendering uses SVG we are no longer limited by a maximum zoom level. On the other hand, the SVG drawing has to be fully designed by the author, as we don't support, for the time being, SVG file generation from an OpenStreetMap document. These two approaches are different and their uses depend on the desired level of customization of the rendering (generated automatically or manually).

## WIMMICS Project-Team

## 6. New Results

### 6.1. QAKiS: Question Answering wiKiframework-based System

**Participants:** Elena Cabrio, Julien Cojan, Amine Hallili, Serena Villata.

We 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. QAKiS exploits the alignment between properties carried out by DBpedia contributors as a mapping from Wikipedia terms to a common ontology, to exploit information coming from DBpedia multilingual chapters (English, French and German), broadening therefore its coverage. We also worked on proposing an argumentation theory model to reason over the inconsistent information sets obtained from DBpedia multilingual chapters, and provide nevertheless a unique and motivated answer to the user.

A demo of the system is available online <sup>6</sup>. The results of this research have been published in [26], [27], [54], [34].

### 6.2. Combining Argumentation Theory and Natural Language Processing

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

With the growing use of the Social Web, an increasing number of applications for exchanging opinions with other people are becoming available online. To cut in on a debate, the participants need first to evaluate the opinions of the other users to detect whether they are in favor or against the debated issue. An automated framework to detect the relations among the arguments represented by the natural language formulation of the users opinions is therefore needed. The work in this area proposes the use of natural language techniques to identify the arguments and their relations. In particular, the textual entailment approach is adopted, i.e. a generic framework for applied semantics, where linguistic objects are mapped by means of semantic inferences at a textual level. Textual entailment is then coupled with an abstract bipolar argumentation system which allows to identify the arguments that are accepted in the considered online debate.

The same framework is also experimented to support the management of argumentative discussions in wiki-like platforms. The results of this research have been published in [16], [28], [29].

### 6.3. Understanding Query Behavior and Explaining Linked Data

**Participants:** Fabien Gandon, Rakebul Hasan.

Our main research is to understand how to assist users in querying [63] and consuming [64] Linked Data. In querying Linked Data, we help users by providing information on how a query may behave. In addition, we provide information about the behavior of similar queries executed in the past. Users can use these information for query construction and refinement. Accurately predicting query behavior is also important for workload management, query scheduling, query optimization. In consuming Linked Data, we explain why a given piece of data exists and how the data was derived. Users can use these explanations to understand and debug Linked Data. Overall, we address the followings research questions:

- i. *How to predict query behavior prior to executing the query?*
- ii. *How to explain Linked Data?*

---

<sup>6</sup><http://qakis.org/qakis2/>

### 6.3.1. Predicting query behavior

To predict query behavior prior to query execution, we apply machine learning techniques on the logs of executed queries. We work with SPARQL queries and predict how long a query would take to execute. We use the frequencies and the cardinalities of SPARQL algebra operators of a query as its features. We also extract a compact set of features from the basic graph patterns belonging to the query. We achieve high accuracy ( $R^2 = 0.837$ ) using the k-nearest neighbors regression. We also suggest similar queries from the query log using an efficient neighbors search. Users can use these suggestions to understand behaviors of similar past queries, and construct and refine their queries accordingly.

### 6.3.2. Explaining Linked Data

The diverse and distributed nature of Linked Data presents opportunities for large-scale data integration and reasoning over cross-domain data. In this scenario, consumers of Linked Data may need explanations for debugging or understanding ontologies. A consumer may also want a short explanation to have an overview of the reasoning. We propose to publish the explanation related metadata as Linked Data. This enables us to explain derived data in the distributed setting of Linked Data. We present the *Ratio4TA*<sup>7</sup> vocabulary to describe explanation metadata and guidelines to publish these metadata as Linked Data. In addition, we summarize explanations using four measures: centrality, coherence, abstractness, and similarity. Users can specify their explanation filtering criteria - types of information they are interested in. We evaluate our summarization approach by comparing the summarized explanations generated by our approach and ground truth summarized explanations generated by humans. Our explanation summarization approach performs roughly with 60% to 70% accuracy for small summaries.

## 6.4. Linguistic Knowledge Representation: the Unit Graphs Formalism

**Participants:** Fabien Gandon, Maxime Lefrançois.

As any community of interest, linguists produce knowledge. Generic needs arise with such produced knowledge: how to represent it, how to manipulate it, how to share it, how to query it, and how to reason with it. To answer these needs is the goal of the knowledge representation (KR) domain. Existing KR formalisms such as the Semantic Web formalisms are standard solutions, and their specialization to the linguistic domain is under active development. Yet, the description logic behind the OWL formalism fails to represent how the meaning of words combine to build up the meaning of sentences. To tackle this specific problem, we introduced the new so-called Unit Graphs KR framework that is portable to existing KR standards but that introduces its own formal logic. UGs are defined over a UG-support that contains: i) a hierarchy of unit types which is strongly driven by the actantial (from action) structure of unit types, ii) a hierarchy of circumstantial symbols, and iii) a set of unit identifiers. On these foundational concepts, we defined UGs, justified the introduction of a deep-semantic representation level for the Meaning–Text Theory, we represented lexicographic definitions of lexical units, and we introduced two formal semantics: one based on UGs closure and homomorphism, and one based on model theoretic semantics. The UGs formalism has been the object of 6 publications in [42], [43], [44], [45], [46], [47].

### 6.4.1. Editor of Formal Lexicographic Definitions

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

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. The development of the GUI was preceded by the elaboration of scenarios of how users would interact with Units Graphs objects. It was followed by user tests of the GUI with actual lexicographers unfolding the scenarios. This work is reported in [46].

## 6.5. Access Control and Presentation for Linked Data

**Participants:** Luca Costabello, Fabien Gandon, Serena Villata.

---

<sup>7</sup><http://ns.inria.fr/ratio4ta/>



PRISSMA is an *adaptive* rendering engine for Linked Data resources. PRISSMA tweaks RDF visualization to the mobile context in which the resource consumption is performed. The work in 2013 has been focused on designing the algorithm that selects the best RDF visualization according to the real, sensed context. Such *selection algorithm* finds optimal error-tolerant subgraph isomorphisms between RDF graphs using the notion of graph edit distance and is sublinear in the number of context declarations in the system. The PRISSMA selection algorithm has been implemented as an Android library, and a test campaign assessed response time and memory consumption. A proof-of-concept, PRISSMA-equipped, mobile RDF browser has been developed to test PRISSMA in a real-world application.

We proposed an extension of Shi3ld, a context-aware access control framework for the Web of Data, developed last year. In particular, we showed how the Shi3ld attribute-based authorization framework for SPARQL endpoints has been progressively converted to protect HTTP operations on RDF. We started by supporting the SPARQL 1.1 Graph Store Protocol and shifted towards a solution without SPARQL for the Linked Data Platform. The resulting authorization framework provides the same functionalities of its SPARQL-based counterpart, including the adoption of Semantic Web languages only. Moreover, a user-friendly interface allowing non expert users to create Shi3ld access policies through a GUI has been designed and developed. These results have been published in [22], [36], [35].

Luca Costabello co-supervised a six-month master student internship related to the Shi3d project. The student, Iacopo Vagliano, from Politecnico di Torino (Italy) developed a Web application to manage Shi3ld access control policies.

## 6.6. Reasoning about Data Licensing in the Web of Data

**Participants:** Fabien Gandon, 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), we 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 AND-composition and OR-composition heuristics have been proposed to support the data publisher in choosing the licenses composition strategy which better suits her needs w.r.t. the data she is publishing. The approach has been evaluated using the SPINdle defeasible reasoner, where the proposed heuristics have been hard coded in the reasoner. The results of this research line have been published in [50], [38].

## 6.7. Semantic and Temporal Analysis of Online Communities

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

This work is done in the PhD of Zide Meng in the OCKTOPUS ANR project.

**Data Formalization:** We use *FOAF* and *SIOC* schema to formalize a dataset from the popular question-answer site *StackOverflow* into RDF format. For some mis-matched vocabulary, we introduce *ugc* schema, which refer to user generated content. Moreover, in order to enrich the dataset, we link tag entity of our dataset to the corresponding entity in DBpedia by using cosine distance of two entities description to solve the disambiguation problem.

**Analysis:** After formalizing the dataset, we begin to exploit some graph mining algorithms, such as community detection algorithm, to analyse the dataset. We extract different kinds of graph from the RDF dataset, such as question-answer graph, co-answer graph, tag co-occurrence graph etc. We aim at finding useful information such as interest groups, experts and tag groups from this kind of question-answer site. By studying the state of the art of community detection algorithm, we analyse the advantage and disadvantage of different approaches, then try to introduce a better algorithm which could outperform others in this scenario.



Plan: During our analysis, we find out some difficult problems which haven't been well solved, such as question intent understanding and community evolution. We will use semantic technology, combining with social network analysis to solve this problem. In the future, we would develop an information management system for such dataset by using analysis algorithms we introduced to improve the performance of information retrieval on user generated content sites.

## 6.8. RDF Mining

**Participant:** Andrea Tettamanzi.

We started investigating an approach to RDF mining based on grammatical evolution and possibility theory: the aim is to mine large RDF graphs by automatically generating and testing OWL 2 axioms based on the known facts. This research effort brings together expertise on metaheuristics for machine learning and data mining, fuzzy logic and possibility theory for representing and handling uncertainty, and the core interests of the Wimmics team, namely, knowledge graphs and the Semantic Web.

Finally, an article describing work on the automatic design of multilayer feedforward neural networks with evolutionary algorithms carried out while still at the University of Milan, got published in [75].

## 6.9. Combination of Evolutionary and Semantic Web Techniques for Protein Design

**Participants:** David Simoncini, Andrea Tettamanzi.

Proteins are fundamental components of all living cells and are among the most studied biological molecules. They are involved in numerous diseases and being able to determine their 3D structures and interactions is essential to understand the mechanisms of cell functions. *De novo* computational protein design refers to the problem of finding a sequence of amino acids corresponding to a protein with the desired three-dimensional structure, or the desired biological function. It is a longstanding goal in computational structural biology and only a few examples of successful *de novo* computational protein designs can be found in the literature. Computational protein design has many industrial applications, such as biofuels, drug synthesis and food processing (through computational design of enzymes) or targetted drug delivery systems (through bio-nanotechnologies).

In this context, our research focuses on knowledge extraction from protein structure databases for the development of new computational protein design frameworks. Whereas most of the current methods ignore available structural information, our algorithm takes into account known profitable interactions between amino acids and uses this information to guide the energy minimization process and propose more realistic sequences of proteins.

## 6.10. Logical Foundations of Cognitive Agents

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

We carried on work on the logical foundations of cognitive agents in collaboration with Salem Benferhat of CRIL and Célia da Costa Pereira of I3S [25] and on the application of such theoretical framework to the problem of exploiting untrustworthy communication in vehicular ad-hoc networks, in collaboration with Ana L. Bazzan and Andrew Koster of the Federal University of Rio Grande do Sul in Brasil and Célia da Costa Pereira of I3S [41]; still related to the issue of trust in multi-agent systems, we took part, with Serena Villata and Célia da Costa Pereira of I3S in a joint investigation with a research team, led by Cristiano Castelfranchi, of the CNR-ISTC in Rome [19].

## 6.11. Requirement Engineering

**Participants:** Isabelle Mirbel, Zeina Azmeh.

The participation of stakeholders (and especially end-users) in requirement engineering is recognized as a key element in the development of useful and usable systems. But in practice, the involvement of end-users is often difficult to implement. 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. An important limitation of these new approaches resides in the information overload, lacking structure and semantics.

In this context, we proposed an approach based on Semantic Web languages as well as concept lattices to identify relevant groups of stakeholders depending on their past participation. We also developed a tool supporting this approach. This work relies on Semantic Web languages and formal concept analysis. Semantic Web languages are used to annotate the data extracted from the platform and to reason about it. Formal Concept Analysis is a theory of data analysis which identifies conceptual structures among data sets. We use it to classify users as well as requirements into lattices which can then be exploited as road maps to examine new requirements. The results of this research have been published in [24].

## **6.12. Management of Technical and Regulatory Knowledge**

**Participants:** Khalil Bouzidi, Michel Buffa, Catherine Faron Zucker, Nhan Le Than.

In the framework of a long-term collaboration with CSTB (Centre Scientifique et Technique du Bâtiment) on the management of technical and regulatory knowledge based on Semantic Web models and techniques, Catherine Faron Zucker and Nhan Le Than co-supervised the PhD thesis of Khalil Riad Bouzidi which has been defended on September 2013.

In the continuation of this work, Catherine Faron Zucker and Michel Buffa got involved with CSTB and three other partners on a project proposal submitted to ANR on the recommendation of technical documents in a social network of building professionals, based on the capitalization and sharing of best search practices.

## **6.13. Co-Construction of Community Ontologies and Corpus in a Limited Technological Environment**

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

In this thesis, we study the implementation of an online platform to build and share the collective memory of citizens in Senegal and revive stories by using a semantic layer. During the first year of this thesis, the first step has been to describe some Use Cases about the platform we would like to develop. We started to define what community means in our work which are group of people with a shared history, culture, ethnicity or interest and want to exchange or collaborate via the Web to share their knowledge of this area.

Our communities are characterized by three components: 1) a common socio-cultural interest, 2) exchange, collaboration and sharing among members and 3) use of the Internet to interact. Thanks to the use cases, we define two main types of users. A community member who is an user who participates in the construction of information and who has interactions with other users. The second type is a simple user who visits the platform for having information, he can be a tourist who want to have information about Senegalese communities activities.

With these use cases, we determine some features that the platform should have. Community members should have, among other thing, a place where they interact to collaborate. To have a "living" community, the system must notify the members of the community about new entries on their focus. Also, to have a catchy presentation, we plan to use maps with different kinds of information.

The second step has been to do a state of the art of online communities. This review allows us to find different definitions and typologies which differ from the study domain – anthropologist, sociologist, psychologist – or the objective – demographic, technological environment, members characteristic. The broadest definition takes into account our concept of community in the context of a knowledge-sharing platform is that of Porter [80]. Despite the fact that numerous typologies are proposed, none is completely consistent with our vision of community. However, the "Toronto School" proposes a category in the classification based on the knowledge transmission called "knowledge-building community" applied in the education area. We think that this type of community could be generalized in the field of socio-constructivisme development, which our communities belong to, for sharing socio-cultural knowledge.

Then, the second phase of this review has been to present the WestAfricapedia project which takes place in this thesis. The main objective is to enhance and sustain the socio-cultural heritage of Senegalese communities through a framework of sharing and co-construction of sociocultural knowledge. Thus we distinguish two main types of communities: knowledge-building community extended in the culture area and exchange information community that has sub-categories such as sports community, commercial community, etc.

## 6.14. Semantic Wiki

**Participants:** Pavel Arapov, Michel Buffa.

We worked on Semantic Web tools, more particularly on WikiNEXT, a semantic application wiki. WikiNEXT lies on the border between application wikis and modern Web based IDEs like jsbin<sup>8</sup>, JSFIDDLE<sup>9</sup>, cloud9 IDE<sup>10</sup>, etc. It has been initially created for writing documents that integrate data from external data sources of the Web of Data, such as DBpedia.org or FreeBase.com, or for writing interactive tutorials (e.g. an HTML5 tutorial, a semantic Web programming tutorial) that mix text and interactive examples in the same page. The system combines some powerful aspects from (i) wikis, such as ease of use, collaboration and openness, (ii) semantic Web/wikis such as making information processable by machines and (iii) Web-based IDEs such as instant development and code testing in a Web browser.

WikiNEXT can be used for writing documents/pages as well as for writing Web applications that manipulate semantic data, either locally or coming from the Web of Data. These applications can be created, edited or cloned in the browser and can be used for integrating data visualizations in wiki pages, for annotating content with metadata, or for any kind of processing. WikiNEXT is particularly suited for teaching Web technologies or for writing documents that integrate data from the Web of data.

## 6.15. Semantic Aggregation

**Participant:** Christophe Desclaux.

Christophe spent one year in the Wimmics team (October 2012 to October 2013) as an invited engineer funded by the BoostYourCode contest he won in 2012. The aim of the BoostYourCode contest (organized by Inria) is to offer to a junior engineer a one year full time contract to work on an innovating OpenSource project.

We worked on an RSS feed aggregation tool using Named Entities Recognition. Rador.NET<sup>11</sup> provides a specialized tool for monitoring news from various sources like RSS, twitter or facebook feeds. Rador.NET semantically increases news for a better classification for the user. We worked on document clustering, natural language processing, RDF datastores and building efficient SPARQL queries.

## 6.16. Semantic Mappings

**Participants:** Thi Hoa Hue Nguyen, Nhan Le Thanh.

---

<sup>8</sup><http://www.jsbin.com>

<sup>9</sup><http://www.jsfiddle.net>

<sup>10</sup><http://www.cloud9ide.com>

<sup>11</sup><http://www.rador.net>

This PhD thesis is about semantic mappings with a control flow-based business workflow: an approach to develop control flow applications using knowledge-based systems.

Although software systems employed to create and execute automatically business processes have been becoming more and more available and advanced, each system is built to deal with a particular workflow type. In addition, these systems require a great deal of time and effort of expert programmers as well as the knowledge of domain experts to set up. Therefore, it is desirable to develop an alternative approach.

Our objective is to represent control flow-based business workflow patterns (CBWPs) in knowledge base by a declarative approach. We first propose an ontological model to represent Coloured Petri Nets (CPNs) with OWL DL. On this basis, we define a meta-knowledge base for CBWPs management. We then develop a graphical interface to design and simulate CBWPs. Our ongoing work is to develop a middleware prototype for mapping and using a CBWP with a user's knowledge base in order to illustrate the feasibility of our approach [49].

## **6.17. Emotional and Social Web**

### ***6.17.1. Modeling, Detection and Annotation of Emotional States using an Algebraic Multidimensional Vector Space***

**Participants:** Nhan Le Thanh, Imen Tayari.

In this research work, we presents a generic solution of emotional data exchange between heterogeneous multi-modal applications. This proposal is based on a new algebraic representation of emotions and it is composed of three distinct layers: the psychological layer, the formal computational layer and the language layer. Moreover, our proposal provides powerful mathematical tools for the analysis and the processing of these emotions and it enables the exchange of the emotional states regardless to the modalities and sensors used in the detection step. The validation of the proposed solution is done with K-nearest neighbor classification algorithm for detecting and evaluating emotion from Eight-Emotion Sentics Data.

### ***6.17.2. Social radio: a Case Studies of Social Network Services***

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

In this project, we carry out some case studies of social radio that is an information service on social networking. Two case studies are conducted on the topics of traffic incidents and geo-epidemiologies. These case studies allow us to study a formal model of spatiotemporal annotations on social network.

### ***6.17.3. Participatory Mapping and Social Bookmarking***

**Participants:** Michel Buffa, Alain Giboin.

In continuation of ISICIL, collaboration began this year between the ITCS-HSS research teams Wimmics and Tech-CICO (UTT), in association with Mnemotix and Wannago startups, in order to design a platform of participatory mapping in the field of sustainable tourism. This platform will enable the various actors in the field (tourists, tourism service providers, scientific experts of fauna, flora and geology, associations, and so on) exchange knowledge about the site and thus enhance the site attraction. This platform is called "socio-semantic" because it offers a unique combination of Semantic Web and Social Web technologies.

The article [51] details one of the planned scenarios of use of the platform and illustrates some proposed functionalities such as Webmarks (Wimmics and Mnemotix) and multiple viewpoints (Tech-CICO). The article also shows how ICT and HSS researchers will collaborate to analyze the innovative uses of the platform on the first fields of application (in Provence- Alpes-Côte d'Azur Region).

### ***6.17.4. Modeling Team Processes***

**Participants:** Pierre Robillard, Isabelle Mirbel, Zeina Azmeh, Alain Giboin, Mathieu Lavallée.

Recent studies outline the importance of software development teams' interactions, suggesting that poor team dynamics can lead to poor software. The relationship between "soft" issues like team dynamics and "hard" issues like software quality is difficult to observe, however. To bridge the gap between these two kinds of issues, and to help development teams prevent quality issues through the planning of relevant team activities, 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 [83] – the CoDyMA (Collaborative Dynamics Measurement and Analysis) method. We proposed an analysis procedure of episodes based on the Formal Concept Analysis (FCA) approach. This procedure uses as input the data (namely the accounts of face-to-face interactions) reported by the developers in their activity diary. The entries are coded in terms of interaction episodes and artifact types to produce a FCA lattice. The observed lattice is compared to a prescribed lattice, and adjustments can be proposed to the team if necessary. The procedure was applied to data from a case study. This work is described in a paper submitted for publication.

### **6.17.5. Modeling Users and Groups of Users**

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

#### *6.17.5.1. Emphasizing Dysfunctional Group Dynamics in Collaboration Personas: Specification of an Approach*

Comparing Collaboration Personas and Individual Personas for the design and evaluation of collaboration software, Judge, Matthews, and Whittaker [79] found that practitioners preferred collaboration personas, but required that the method put more emphasis on problematic or dysfunctional group dynamics. Because Judge et al. only outlined a possible approach to meet this requirement, we decided to contribute to the specification of the approach. In [37] we report the first steps of this specification work.

#### *6.17.5.2. Using Formal Concept Analysis to elicit Personas*

Personas are built from a clustering of behavioral variables common to a set of users. Behavioral variables are ways in which users behavior differ (e.g., goals and attitudes); it is important to elicit them because they have an impact on the system to be designed. Today, the clustering is mainly performed manually. To automate it, we started this year to explore the use of Formal Concept Analysis tools.

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

**Participants:** Aurore Defays, Alain Giboin.

Grounding is the process used by participants to a collective activity to coordinate both the content and process of their communication to be successful [71]. 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 [72]. So far, grounding has been studied mainly from a unimodal point of view, i.e., from the point of view of the verbal modality (oral or written). Some authors have begun to study grounding from a bimodal viewpoint; for example [78] have studied the use of verbal and gestural modalities necessary to ensure mutual understanding in interactions between Japanese airline pilots and an American flight instructor. In the context of her PhD thesis in Ergonomics applied to architectural design digital tools, Aurore Defays extended the study of grounding to actual multimodality (with n modality: oral, written, gestures, gazes, etc.).

With Aurore Defays, we focused this year on improving the methodology of analysis of multimodal grounding proposed in [74]. To do this, we relied on the data of an existing study by Defays on remote collaboration between dyads and triads of architects interacting through a collaborative digital studio (the Distributed Collaborative Digital Studio, DSDC). Our initial research question was: Is a multimodal shared representation preferable to a unimodal representation to collaborate effectively? Analyzing the data, this question was gradually transformed into: Which modalities are relevant to build the common ground necessary for a particular type of collaboration to succeed?

### **6.17.7. The "Design Thinking" Toolset: Application to Discovery Hub**

**Participants:** Gessica Puri, Alain Giboin, Nicolas Marie, Damien Legrand.

Last year was developed a "design thinking" toolset (including a framework) for helping developers think in terms of a user's point of view when they design and evaluate link visualization and manipulation applications such as graph visualization applications [81]. The toolset was used this year to perform a qualitative evaluation of Discovery Hub, so contributing to the development of a new version of the discovery engine.

The ShowCaseMachine project led by Damien Legrand won the 11th Challenge *Jeune Pousses* at Telecom Valley in Sophia Antipolis.

## 6.18. Graph-based Knowledge Representation

### 6.18.1. SPARQL Based Pretty Printing Language

**Participant:** Olivier Corby.

We have designed SPARQL Template, a pretty-printing rule language for RDF graphs. It enables to pretty print RDF graphs representing Abstract Syntax Trees of languages such as SPIN or OWL RDF syntax. We have implemented a pretty printing engine that interprets SPARQL Template.

An example of template for a OWL "someValuesFrom" statement is shown below. The SPARQL 1.1 "where" part specifies the conditions to apply the rule on a focus node "?in". The template part specifies the result of the pretty print of the focus node. Variables in the template part are recursively replaced by the result of their pretty print.

```
template {
  "someValuesFrom(" ?p " " ?c ") "
}
where {
  ?in a owl:Restriction ;
  owl:onProperty ?p ;
  owl:someValuesFrom ?c
}
```

We have introduced named templates that are called explicitly using a "kg:template" extension function.

The pretty printing language and engine have been validated on five RDF AST<sup>12</sup>: SPIN, OWL 2, SQL, Turtle and a mockup of mathematical expressions pretty printed into Latex. The SPIN pretty printer is used in the PhD Thesis of Oumy Seye on "Rules for the Web of Data" and the SQL pretty printer is used in the PhD Thesis of Corentin Follenfant on "Usage semantics of analytics and Business Intelligence tools".

### 6.18.2. Federated Semantic Data Query

**Participants:** Olivier Corby, Alban Gaignard.

Another activity of the team addresses the data explosion challenges faced in e-Science. Semantic Web technologies are well adopted to represent the knowledge associated to both e-Science data and processing tools. A PhD thesis [76], addressing the distributed knowledge production and sharing in collaborative e-Science platforms, has successfully been defended this year. Moreover, we have been participating in the organization of the second edition of the CrEDIBLE workshop<sup>13</sup>, gathering international experts to discuss the challenges of federating distributed biomedical imaging data and knowledge.

In this area, the main scientific results are (i) a software architecture for transparently querying multiple data sources through the SPARQL language [73], (ii) a set of querying strategies and optimizations dedicated to limit the cost of distributed query processing, while still considering enough expressivity (full SPARQL 1.1 support, including named graphs, property path expressions, optional, aggregates, etc.).

<sup>12</sup><ftp://ftp-sop.inria.fr/wimmics/soft/pprint>

<sup>13</sup><http://credible.i3s.unice.fr>



Performance-oriented experiments have been conducted on the Grid'5000 distributed computing infrastructure to compare our approach with state-of-the-art engines such as FedX [85], Splendid [77], or DARQ [82]. Experiments, based on the FedBench benchmark [84] show performances between DARQ, Splendid, and FedX, while still high expressivity.

Since distributed query processing lead to complex and costly processes, we started to collect provenance information which opens interesting perspectives towards enhanced trust and reproducibility in Linked Data querying and reasoning.

These distributed query processing strategies have been implemented and integrated into Corese through two main components, namely a data source federator, and a data source endpoint. A prototype Web application has also been developed to demonstrate our approach. End-users can configure and launch distributed SPARQL querying and finally visualize SPARQL results and their associated provenance.

### **6.18.3. Rules for the Web of Data**

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

This work takes place in the PhD Thesis of Oumy Seye.

The objective of this year is to foster knowledge reuse on the Web based on the principles of Linked Data. Our approach is to consider rule bases like data sources that can be published, shared and queried as Linked Data, thus enabling the selection and reuse of relevant and useful shared rules in any particular context or application. We propose to select rules by querying either metadata annotating rules, rules content or both. To make rules content queryable, we use RDF representations of SPARQL rules with the SPIN format <sup>14</sup>.

This idea joins the principles of the Semantic Web that encourages the sharing and reuse of knowledge. We used the SPIN syntax (which allows the representation of a SPARQL query in RDF) obtained with the SPIN pretty printer of Corese. We have subsequently been able to select rules of interest with Corese. The proposal enables to search rules based on their content. This allows us to help users extract relevant set of rules for their data, and thus leverage more easily shared rules. This idea can be used to build a search engine for rules on the Web or a tool for automatically connect rules with semantic data.

In the remainder of this work, we will focus on updating harvested rules. A poster on this work was presented for the GLC pole day July 8, and at the summer school ESWC September 2.

### **6.18.4. Semantic Web and Business Intelligence**

**Participants:** Corentin Follenfant, Olivier Corby, Fabien Gandon.

This PhD Thesis is done with a CIFRE industrial grant from SAP Research.

The bilateral contract with SAP aims at converging Semantic Web and Business Intelligence through a framework applying the read/write Web principles to the business knowledge carried within Business Intelligence reports. These reports often provide a dynamic view upon numerical data from various enterprise sources, mainly relational databases. Reports are authored with a complex process that can be reduced to writing, directly or through different layers of user interfaces, SQL queries that will query the sources and feed the dynamic reports. In order to simplify the query authoring process, complementary approaches are envisioned.

Our approach proposes to model the queries as knowledge through their abstract syntax trees (ASTs) with Semantic Web tools, query and manipulate them through appropriate standards, respectively RDF/S and SPARQL. Indeed RDF enables us to model the actual structure of the ASTs by integrating the knowledge related to syntax and semantics of the SQL queries: types can be captured with XML Schema Datatypes, while more specific business knowledge can also be designed according to the source business models and annotate various entities referenced within the SQL queries. Regarding the query and manipulation part, a library of SPARQL queries was designed to perform generic AST manipulation (generic from a DSL perspective), and is usable to search, extract, edit, prune or graft parts of RDF-modelled ASTs.

---

<sup>14</sup><http://www.w3.org/Submission/spin-overview/>



While this year was mostly dedicated to manuscript writing, additional experiments were run to demonstrate the validity of our model: a large set of ANSI SQL queries generated with a TPC-DS benchmark was converted to its RDF representation. Inversely, a generic pretty printer system developed into the Corese engine was validated by the internship of Abdoul Macina who developed a set of rules to have the pretty printer turn RDF-modelled SQL queries back to their concrete syntactic form. This enables iterative query design by leveraging AST patterns rather than manually editing brute syntax.

## ZENITH Project-Team

# 6. New Results

## 6.1. Big Data Integration

### 6.1.1. 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 [20], we proposed FD (Fully Distributed), a decentralized algorithm for dealing with the ERPD problem over distributed data, with the goal of minimizing bandwidth usage and reducing processing time. It has the following salient features. First, it uses the novel concepts of *Potential* and *essential-set* to prune data at local nodes. This leads to a significant reduction of bandwidth usage compared to the baseline approaches. Second, its execution is completely distributed and does not depend on the existence of certain nodes. We validated FD through implementation over a 75-node cluster and simulation using both synthetic and real-world data. The results show very good performance, in terms of bandwidth usage and response time.

### 6.1.2. Open Data Integration

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

Working with open data sources can yield high value information but raises major problems in terms of metadata extraction, data source integration and visualization. For instance, Data Publica provides more than 12 000 files of public data. However, even though data formats become richer and richer in terms of semantics and expressivity (e.g. RDF), most data producers do not use them much in practice, because they require too much upfront work, and keep using simpler tools like Excel. Unfortunately, no integration tool is able to deal in an effective way with spreadsheets. Only few initiatives (OpenII and Google Refine) deal with Excel files. However, their importers are very simple and make some strict restrictions over the input spread-sheets.

In [31], we describe a demonstration of WebSmatch, a flexible environment for Web data integration. WebSmatch supports the full process of importing, refining and integrating data sources and uses third party tools for high quality visualization. We use a typical scenario of public data integration which involves problems not solved by current tools: poorly structured input data sources (XLS files) and rich visualization of integrated data.

### 6.1.3. Pricing Integrated Data

**Participant:** Patrick Valduriez.

Data is a modern commodity, being bought and sold. Electronic data market places and independent vendors integrate data and organize their online distribution. Yet the pricing models in use either focus on the usage of computing resources, or are proprietary, opaque, most likely ad hoc, and not conducive of a healthy commodity market dynamics. In [39], we propose a generic data pricing model that is based on minimal provenance, i.e. minimal sets of tuples contributing to the result of a query. We show that the proposed model fulfills desirable properties such as contribution monotonicity, bounded-price and contribution arbitrage-freedom. We present a baseline algorithm to compute the exact price of a query based on our pricing model. We show that the problem is NP-hard. We therefore devise, present and compare several heuristics. We conduct a comprehensive experimental study to show their effectiveness and efficiency.

In most data markets, prices are prescribed and accuracy is determined by the data. Instead, we consider a model in which accuracy can be traded for discounted prices: “what you pay for is what you get”. The data market model consists of data consumers, data providers and data market owners. The data market owners are brokers between the data providers and data consumers. A data consumer proposes a price for the data that she requests. If the price is less than the price set by the data provider, then she gets an approximate value. The data market owners negotiate the pricing schemes with the data providers. They implement these schemes for the computation of the discounted approximate values. In [38], we propose a theoretical and practical pricing framework with its algorithms for the above mechanism. In this framework, the value published is randomly determined from a probability distribution. The distribution is computed such that its distance to the actual value is commensurate to the discount. The published value comes with a guarantee on the probability to be the exact value. The probability is also commensurate to the discount. We present and formalize the principles that a healthy data market should meet for such a transaction. We define two ancillary functions and describe the algorithms that compute the approximate value from the proposed price using these functions. We prove that the functions and the algorithm meet the required principles.

## 6.2. Distributed Indexing and Searching

### 6.2.1. P2P Search and Recommendation

**Participants:** Esther Pacitti, Maximilien Servajean.

In crossdiscipline domains, users belonging to different communities produce various scientific material that they own, share, or endorse. In that context, we are interested in querying and recommending scientific material in the form of documents. Such documents cover various topics such as models for plant phenotyping, statistics on specific kinds of plants, or biological experiments.

In [40], we investigate profile diversity, a novel idea in searching scientific documents. Combining keyword relevance with popularity in a scoring function has been the subject of different forms of social relevance. On the other hand, content diversity has been thoroughly studied in search and advertising, database queries, and recommendations.

We introduce profile diversity for scientific document search as a complement to traditional content diversity. Profile diversity combines the discipline and communities to which a user belongs. We propose an adaptation of Fagin’s threshold-based algorithms to return the most relevant and most popular documents that satisfy content and profile diversities. To validate our scoring function, DivRSci, we ran experiments that use two benchmarks: a realistic benchmark with scientists and TREC’09. We show that DivRSci presents the best compromise between all requirements we have identified. DivRSci also shows to be the best generating list of inter-disciplinary and inter-community documents. Finally, it yields very good gains (by a factor of 6), suited for profile diversification

### 6.2.2. Spatial Queries in Wireless Data Broadcasting

**Participant:** Patrick Valduriez.

The main requirements for spatial query processing via mobile terminals include rapid and accurate searching and low energy consumption. Most location-based services (LBSs) are provided using an on-demand method, which is suitable for light-loaded systems where contention for wireless channels and server processing is not severe. However, as the number of users of SBSs increases, performance deteriorates rapidly since the servers' capability to process queries is limited. Furthermore, the response time of a query may significantly increase with the concentration of users' queries in a server at the same time. That is because the server has to check the locations of users and potential objects for the final result and then individually send answers to clients via a point-to-point channel. At this time, an inefficient structure of spatial index and searching algorithm may incur an extremely large access latency.

To address this problem, we propose in [27] the Hierarchical Grid Index (HGI), which provides a light-weight sequential location-based index structure for efficient SBSs. We minimize the index size through the use of hierarchical location-based identifications. And we support efficient query processing in broadcasting environments through sequential data transfer and search based on the object locations. We also propose Top-Down Search and Reduction-Counter Search algorithms for efficient searching and query processing. HGI has a simple structure through elimination of replication pointers and is therefore suitable for broadcasting environments with one-dimensional characteristics, thus enabling rapid and accurate spatial search by reducing redundant data. Our performance evaluation shows that our proposed index and algorithms are accurate and fast and support efficient spatial query processing.

## 6.3. Big Data Analysis

### 6.3.1. Big Data Analysis using Algebraic Workflows

**Participants:** Jonas Dias, Patrick Valduriez.

Analyzing big data requires the support of dataflows with many activities to extract and explore relevant information from the data. Recent approaches such as Pig Latin propose a high-level language to model such dataflows. However, the dataflow execution is typically delegated to a MapReduce implementation such as Hadoop, which does not follow an algebraic approach, thus it cannot take advantage of the optimization opportunities of PigLatin algebra.

In [35], we propose an approach for big data analysis based on algebraic workflows, which yields optimization and parallel execution of activities and supports user steering using provenance queries. We illustrate how a big data processing dataflow can be modeled using the algebra. Through an experimental evaluation using real datasets and the execution of the dataflow with Chiron, an engine that supports our algebra, we show that our approach yields performance gains of up to 19.6% using algebraic optimizations in the dataflow and up to 39% of time saved on a user steering scenario.

This work was done in the context of the CNPq-Inria Hoscarr project and FAPERJ-Inria P2Pcloud project .

### 6.3.2. 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 our team, we study the problem of data partitioning in two different contexts: (1) in scientific databases that are continuously growing and (2) in the MapReduce framework. In both cases, we propose automatic approaches, which are performed transparently to the users, in order to free them from the burden of complex partitioning.

In [25], 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 [25]. 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.

In [37] and [43], we address the problem of high data transfers in MapReduce, and propose a technique that repartitions tuples of the input datasets. Our technique optimizes the distribution of key-values over mappers, and increases the data locality in reduce tasks. It captures the relationships between input tuples and intermediate keys by monitoring the execution of a set of MapReduce jobs which are representative of the workload. Then, based on those relationships, it assigns input tuples to the appropriate chunks. With this data repartitioning and a smart scheduling of reducer tasks, our approach significantly contributes to the reduction of transferred data between mappers and reducers in job executions. We evaluate our approach through experimentation in a Hadoop deployment on top of Grid5000 using standard benchmarks. The results show high reduction in data transfer during the shuffle phase compared to Native Hadoop.

## 6.4. Data Stream Mining

### 6.4.1. Mining Uncertain Data Streams

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

Discovering Probabilistic Frequent Itemsets (PFI) is very challenging since algorithms designed for deterministic data are not applicable in probabilistic data. The problem is even more difficult for probabilistic data streams where massive frequent updates need to be taken into account while respecting data stream constraints. In [28], we propose FEMP (Fast and Exact Mining of Probabilistic data streams), the first solution for exact PFI mining in data streams with sliding windows. FEMP allows updating the frequentness probability of an itemset whenever a transaction is added or removed from the observation window. Using these update operations, we are able to extract PFI in sliding windows with very low response times. Furthermore, our method is exact, meaning that we are able to discover the exact probabilistic frequentness distribution function for any monitored itemset, at any time. We implemented FEMP and conducted an extensive experimental evaluation over synthetic and real-world data sets; the results illustrate its very good performance.

### 6.4.2. Itemset Mining over Tuple-Evolving Data Streams

**Participant:** Florent Masseglia.

In many data streaming applications today, tuples inside the streams may get revised over time. This type of data stream brings new issues and challenges to the data mining tasks. In [42] we present a theoretical analysis for mining frequent itemsets from sliding windows over such data. We define conditions that determine whether an infrequent itemset will become frequent when some existing tuples inside the streams have been updated. We design simple but effective structures for managing both the evolving tuples and the candidate frequent itemsets. Moreover, we provide a novel verification method that efficiently computes the counts of candidate itemsets. Experiments on real-world datasets show the efficiency and effectiveness of our proposed method.

## 6.5. Scalable Data Analysis

### 6.5.1. Scalable Mining of Small Visual Objects

**Participants:** Pierre Letessier, Julien Champ, Alexis Joly.

Automatically linking multimedia documents that contain one or several instances of the same visual object has many applications including: salient events detection, relevant patterns discovery in scientific data or simply web browsing through hyper-visual links. Whereas efficient methods now exist for searching rigid objects in large collections, discovering them from scratch is still challenging in terms of scalability, particularly when the targeted objects are small compared to the whole image. In a previous work, we revisited formally the problem of mining or discovering such objects, and then generalized two kinds of existing methods for probing candidate object seeds: weighted adaptive sampling and hashing based methods. This year, we continued working on the subject by improving our high-dimensional data hashing strategy, that works first at the visual level, and then at the geometric level. We conducted new experiments on a dedicated evaluation dataset <sup>5</sup> and we did show that our the recall or our approach definitely outperforms the reference method [46].

Based on this contribution, we then address the problem of suggesting object-based visual queries in a multimedia search engine [22], [36]. State-of-the-art visual search systems are usually based on the query-by-window paradigm: a user selects any image region containing an object of interest and the system returns a ranked list of images that are likely to contain other instances of the query object. User's perception of these tools is however affected by the fact that many submitted queries actually return nothing or only junk results (complex non-rigid objects, higher-level visual concepts, etc.). In [22], we addressed the problem of suggesting only the object's queries that actually contain relevant matches in the dataset. This requires to first discover accurate object's clusters in the dataset (as an offline process); and then to select the most relevant objects according to user's intent (as an on-line process). We therefore introduce a new object's instances clustering framework based on a bipartite shared-neighbours clustering algorithm that is used to gather object's seeds discovered by our visual mining method. Shared nearest neighbours methods were not studied beforehand in the case of bipartite graphs and never used in the context of object discovery. Experiments show that this new method outperforms state-of-the-art object mining and retrieval results on the Oxford Building dataset. We finally describe two real-word object-based visual query suggestion scenarios using the proposed framework and show examples of suggested object queries. A demo was presented at ACM Multimedia 2013 [36].

This method was finally integrated within a visual-based media event detection system in the scope of a French project called the Transmedia Observatory [33]. It allows the automatic discovery of the most circulated images across the main news media (news websites, press agencies, TV news and newspapers). The main originality of the detection is to rely on the transmedia contextual information to denoise the raw visual detections and consequently focus on the most salient trans-media events.

### 6.5.2. Rare Events Identification for Large-Scale Applications

**Participant:** Florent Masegla.

While significant work in data mining has been dedicated to the detection of single outliers in the data, less research has approached the problem of isolating a group of outliers, i.e. rare events representing micro-clusters of less – or significantly less – than 1% of the whole dataset. This research issue is critical for example in medical applications. The problem is difficult to handle as it lies at the frontier between outlier detection and clustering and distinguishes by a clear challenge to avoid missing true positives. In [41], we address this challenge and propose a novel two-stage framework, based on a backward approach, to isolate abnormal groups of events in large datasets. The key of our backward approach is to first identify the core of the dense regions and then gradually augment them based on a density-driven condition. The framework outputs a small subset of the dataset containing both rare events and outliers. We tested our framework on a biomedical application to find micro-clusters of pathological cells. The comparison against two common clustering (DBSCAN) and outlier detection (LOF) algorithms show that our approach is a very efficient alternative to the detection of rare events – generally a recall of 100% and a higher precision, positively correlated with the size of the rare event – while also providing a  $\mathcal{O}(N)$  solution to the existing algorithms dominated by a  $\mathcal{O}(N^2)$  complexity.

### 6.5.3. Large-scale content-based plants identification from social image data

**Participants:** Hervé Goëau, Alexis Joly, Julien Champ, Saloua Litayem.

<sup>5</sup><http://www-sop.inria.fr/members/Alexis.Joly/BelgaLogos/FlickrBelgaLogos.html>



Speeding up the collection and integration of raw botanical observation data is a crucial step towards a sustainable development of agriculture and the conservation of biodiversity. Initiated in the context of a citizen sciences project in collaboration with the botanists of the AMAP UMR team and Tela Botanica social network, the overall contribution of this work [23] is an innovative collaborative workflow focused on image-based plant identification as a mean to enlist new contributors and facilitate access to botanical data. Since 2010, hundreds of thousands of geo-tagged and dated plant photographs were collected and revised by hundreds of novice, amateur and expert botanists of a specialized social network. An image-based identification tool - available as both a web and a mobile application - is synchronized with that growing data and allows any user to query or enrich the system with new observations. Extensive experiments of the visual search engine as well as system-oriented and user-oriented evaluations of the application did show that it is very helpful to determine a plant among hundreds or thousands of species [23]. As a concrete result, more than 80K people in about 150 countries did download the iPhone end point of the application [32].

From a data management and data analysis perspective, our main contribution concerns the scalability of the system. At the time of writing, the content-based search engine actually works on 120K images covering more than 5000 species (which already makes it the largest identification tool built anytime). The resulting training dataset contains several hundreds of millions feature vectors, each with several hundreds of float attributes (i.e. high-dimensional feature vectors describing the visual content). At query time, thousands of such feature vectors are extracted from the query pictures and have to be searched online in the training set to find the most similar pictures. The underlying search of approximate nearest neighbors is speed-up thanks to a data-dependent high-dimensional hashing framework based on Random Maximum Margin Hashing (RMMH), a new hash function family that we introduced in 2011. RMMH is used for both compressing the original feature vectors into compact binary hash codes and for partitioning the data into a well balanced hash table. Search is then performed through adaptive multi-probe accesses in the hash table and a top-k search refinement step on the full binary hash codes. Last improvements brought in 2013 include a multi-threaded version of the search, the use of a probabilistic asymmetric distance instead of the Hamming distance and the integration of a query optimization training stage in the compressed feature space instead of the original space. A beta version of Pl@ntNet visual search engine based on these new contributions is currently being tested and is about 8 times faster than the one used in production.

Besides scalability and efficiency, we also did work on improving the identification performances of the system [29]. We notably improved the quality of the top-K returned images by weighting each match according to its Hamming distance to the query rather than using a simple vote. We then improved the multi-cue fusion strategy by indexing separately each type of visual features rather than concatenating them in an early phase. We finally did train the optimal selection of features for each of the considered plant organ (flower, leaf, bark, fruit). Beyond the use of the visual content itself, we explored the usefulness of associated metadata and we did prove that some of them like the date can improve the identification performances (contrary to the geo-coordinates that surprisingly degraded the results). Overall, as a result of our participation to ImageCLEF plant identification benchmark [34], we obtained the second best run among 12 international groups and a total of 33 submitted runs.



## ALICE Project-Team

# 5. New Results

## 5.1. Geometry Processing

### 5.1.1. *Fitting Polynomial Volumes to Surface Meshes with Voronoi Squared Distance Minimization*

**Participants:** Gilles-Philippe Paillé, Bruno Lévy.

We propose a method for mapping polynomial volumes. Given a closed surface and an initial template volume grid, our method deforms the template grid by fitting its boundary to the input surface while minimizing a volume distortion criterion. The result is a point-to-point map distorting linear cells into curved ones. Our method is based on several extensions of Voronoi Squared Distance Minimization (VSDM) combined with a higher-order finite element formulation of the deformation energy. This allows us to globally optimize the mapping without prior parameterization. The anisotropic VSDM formulation allows for sharp and semi-sharp features to be implicitly preserved without tagging. We use a hierarchical finite element function basis that selectively adapts to the geometric details. This makes both the method more efficient and the representation more compact. We apply our method to geometric modeling applications in computer-aided design and computer graphics, including mixed-element meshing, mesh optimization, subdivision volume fitting, and shell meshing.

This work was presented at the “ACM Symposium on Geometry Processing” and published in the “Computer Graphics Forum” journal [16].

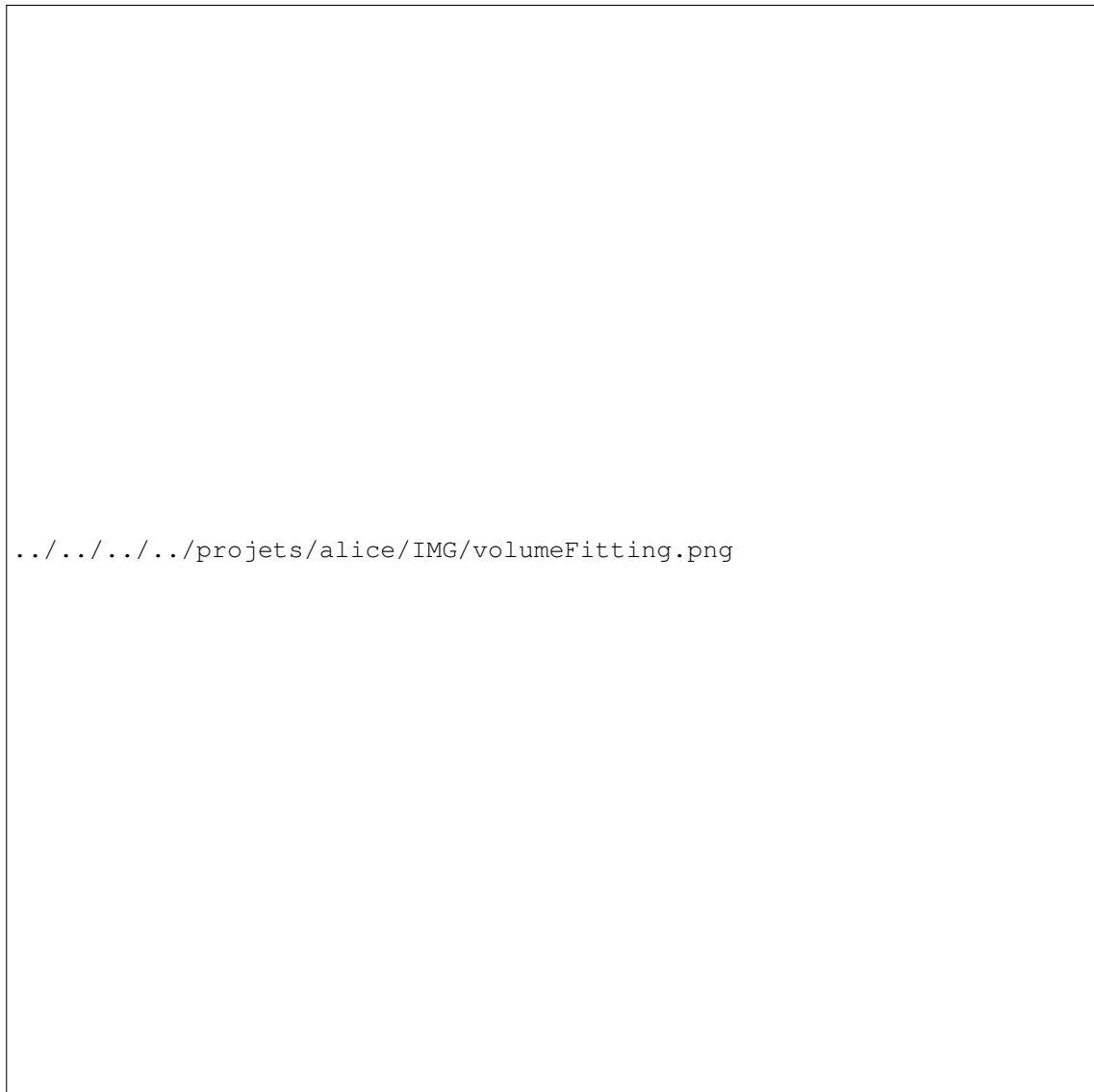
### 5.1.2. *Particle-Based Anisotropic Surface Meshing*

**Participant:** Bruno Lévy.

This paper introduces a particle-based approach for anisotropic surface meshing. Given an input polygonal mesh endowed with a Riemannian metric and a specified number of vertices, the method generates a metric-adapted mesh. The main idea consists of mapping the anisotropic space into a higher dimensional isotropic one, called “embedding space”. The vertices of the mesh are generated by uniformly sampling the surface in this higher dimensional embedding space, and the sampling is further regularized by optimizing an energy function with a quasi-Newton algorithm. All the computations can be re-expressed in terms of the dot product in the embedding space, and the Jacobian matrices of the mappings that connect different spaces. This transform makes it unnecessary to explicitly represent the coordinates in the embedding space, and also provides all necessary expressions of energy and forces for efficient computations. Through energy optimization, it naturally leads to the desired anisotropic particle distributions in the original space. The triangles are then generated by computing the Restricted Anisotropic Voronoi Diagram and its dual Delaunay triangulation. We compare our results qualitatively and quantitatively with the state-of-the-art in anisotropic surface meshing on several examples, using the standard measurement criteria. This work was published in the “ACM Transactions on Graphics” journal (SIGGRAPH conference proceedings) [19].

### 5.1.3. *Approximating Functions on a Mesh with Restricted Voronoi Diagrams*

**Participant:** Bruno Lévy.



*Figure 2. Fitting Polynomial Volumes to Surface Meshes with Voronoi Squared Distance Minimization*

We propose a method that computes a piecewise constant approximation of a function defined on a mesh. The approximation is associated with the cells of a restricted Voronoi diagram. Our method optimizes an objective function measuring the quality of the approximation. This objective function depends on the placement of the samples that define the restricted Voronoi diagram and their associated function values. We study the continuity of the objective function, derive the closed-form expression of its derivatives and use them to design a numerical solution mechanism. The method can be applied to a function that has discontinuities, and the result aligns the boundaries of the Voronoi cells with the discontinuities. Some examples are shown, suggesting potential applications in image vectorization and compact representation of lighting. This work was presented at the “ACM Symposium on Geometry Processing” and published in the “Computer Graphics Forum” journal [15].

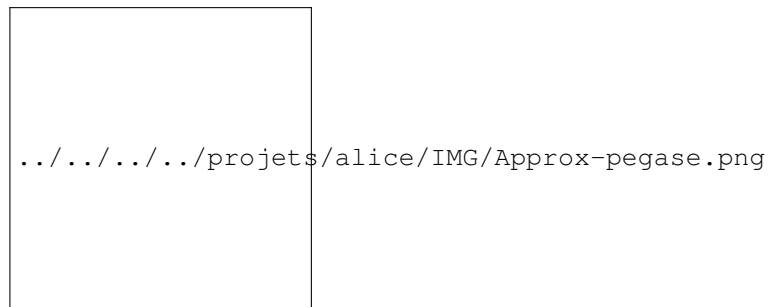


Figure 3. Approximating Functions on a Mesh with Restricted Voronoi Diagrams

#### 5.1.4. Spectral Clustering of Plant Units From 3D Point Clouds

**Participant:** Dobrina Boltecheva.

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

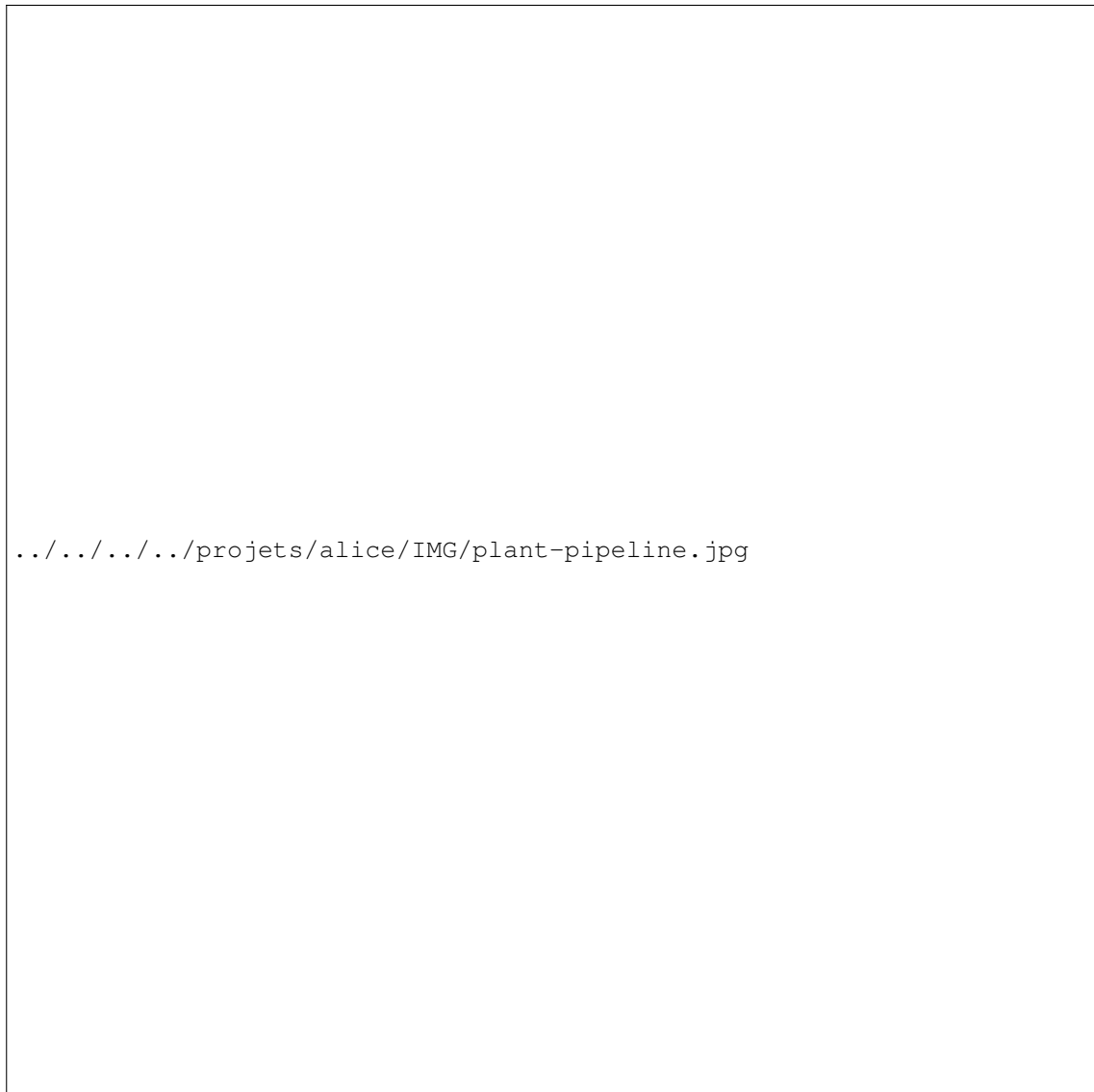
We have developed a new approach based on a delineation algorithm (Fig. 4) that clusters a point cloud into elementary plant units such as internodes, petioles and leaves. The algorithm creates a graph (points + edges) to recover plausible neighbouring relationships between the points and embeds this graph in a spectral space in order to segment the point-cloud into meaningful elementary plant units.

We have presented this work at the 7th International Conference on Functional—Structural Plant Models (FSPM) which took place in Finland this summer [21].

#### 5.1.5. Fixing Normal Constraints for Generation of Polycubes

**Participants:** Nicolas Ray, Dmitry Sokolov.

A polycube is a piecewise linearly defined surface where all faces are squares that are perpendicular to an axis of a global basis. Deforming triangulated surfaces to polycubes provides maps (from the original surface to the polycube) that can be used for a number of applications including hex-meshing. To define such a deformation, it is necessary to determine, for each point of the original surface, what will be its orientation (global axis) in the polycube.



*Figure 4. From a point cloud scan of a plant to a segmentation of its leaves.*

This problem is actually tackled by heuristics that basically affect the closest global axis to the surface normal. Coupled with a mesh deformation as pre-processing and some fixing rules as a post-processing, it is able to provide nice results for a number of surfaces. However, nothing ensures that the surface can be deformed to a polycube having these desired face orientations.

We have worked on a method able to determine if there exists a deformation of the surface that respects a given orientation constraint on each point. We have also designed an automatic solution that can fix constraints that would prevent the existence of a deformation into a polycube (Figure 5).

This study has highlighted that the constraints on desired orientation are global and requires constrained optimization methods to be solved. Our current solution is able to manage many cases where previous works would fail, but we can still produce some complex cases where interactions between dimension may lead to deadlocks.



Figure 5. **Upper row:** the surface is deformed to make its normals closer to major axis, but to reach an equality, we need to have a coherent "wished orientation" of the faces. **Middle row:** we define a valid deformation into a polycube by editing the "wished orientation". **Lower row:** the resolution is performed a dimension at a time.

### 5.1.6. Some Basic Geometric Considerations in Variational Multiview Stereo

**Participant:** Rhaleb Zayer.

We developed a technique for processing correspondences originating from dense variational matching in the context of multiview stereo. Such data tends to be very large and can easily encompass tens or hundreds of millions of points, these figures keep growing as high resolution images are becoming mainstream. Inspired by Lambert's cosine law, we regard the matching as sequences of planar maps across neighboring views, and show how to take advantage of geometric properties of such maps to favor image areas where the cosine angle between the surface normal and the line of sight is maximal. As the approach operates in the planar domain on smaller subsets of neighboring views, it is computationally efficient and has a low memory footprint. A preprint is in preparation.

### 5.1.7. *Multi-frontal Propagation Based Matching*

**Participants:** Rhaleb Zayer, Patricio Galindo.

We consider the propagation-based matching problem, which deals with expanding a limited set of correspondences towards a quasi-dense map across two views. Two issues which have not received much interest in earlier work are raised here. The traversal of weakly textured regions is shown to negatively impact the quality of subsequent correspondences. Analysis of the propagation results using the commonly adopted global best-fit strategy reveals that only a small subset of the input seeds contributes effectively to the propagation, which is probably not optimal since the quality of the matches may deteriorate as the propagation region becomes significantly large as shown in figure 7 -bottom. This research extends existing propagation techniques in two ways: (i) The selection of reliable expansion regions is automatized and adapted to the propagation by categorizing the image into three regions, no-propagation regions, safe-propagation regions and buffer-regions where seeds can propagate but cannot generate new seeds. (ii) A multi-frontal propagation approach is proposed with emphasis on the balance between the greedy nature of the original algorithm and the contribution of the seeds. A preprint is in preparation.

### 5.1.8. *Large Deformations of Slender Objects*

**Participant:** Rhaleb Zayer.

We studied the problem of large spatial deformation in the context of interactive editing of slender curve-like objects. The deformation is analyzed in the local frame of the individual curve segments (beams) and the rigid motion of the local frame is updated using a total Lagrangian approach. Analysis of the virtual work in the light of this decoupling allows formulating the Hessian of the deformation in a simple but principled manner. The resulting representation is sparser than existing derivations and can handle the simultaneous action of torques, and forces, efficiently, so as to reproduce a natural behavior in such path dependent situations. The proposed approach is conceptually simple, easy to implement, and suitable for object editing. The numerical solution is carried out using an efficient iterative scheme which allows stable convergence. A preprint is in preparation.

## 5.2. Computer Graphics

### 5.2.1. *By-example Synthesis of Curvilinear Structured Patterns*

**Participants:** Anass Lasram, Sylvain Lefebvre.

Many algorithms in Computer Graphics require to synthesize a pattern along a curve. This is for instance the case with line stylization, to decorate objects with elaborate patterns (chains, laces, scratches), or to synthesize curvilinear features such as mountain ridges, rivers or roads. We describe a simple yet effective method for this problem. Our method addresses the main challenge of maintaining the continuity of the pattern while following the curve. It allows some freedom to the synthesized pattern: It may locally diverge from the curve so as to allow for a more natural global result. This also lets the pattern escape areas of overlaps or fold-overs. This makes our method particularly well suited to structured, detailed patterns following complex curves. Our synthesizer copies tilted pieces of the exemplar along the curve, following its orientation. The result is optimized through a shortest path search, with dynamic programming. We speed up the process by an efficient parallel implementation. Finally, since discontinuities may always remain we propose an optional post-processing step optimally deforming neighboring pieces to smooth the transitions.



Figure 6. Processing best viewed regions in the Fountain data set (top). Each view represents the central image of a triplet (other two images not shown). The red-colored regions (middle) represent areas best viewed in the triplet. Yellow-marked regions represent regions which are only visible in the triplet and therefore are included even if they do not comply with the best view requirement. The resulting reconstruction (bottom) shows an almost outlier free point cloud.





Figure 7. Typical result of our approach (top), compared to a best-first strategy (bottom). In both experiments, the same initial seeds were used ( $\approx 40$  seeds). The descendants of each initial seed are uniquely colored. Our approach clearly allows all seeds to contribute, whereas the greedy approach marginalize a majority of them.

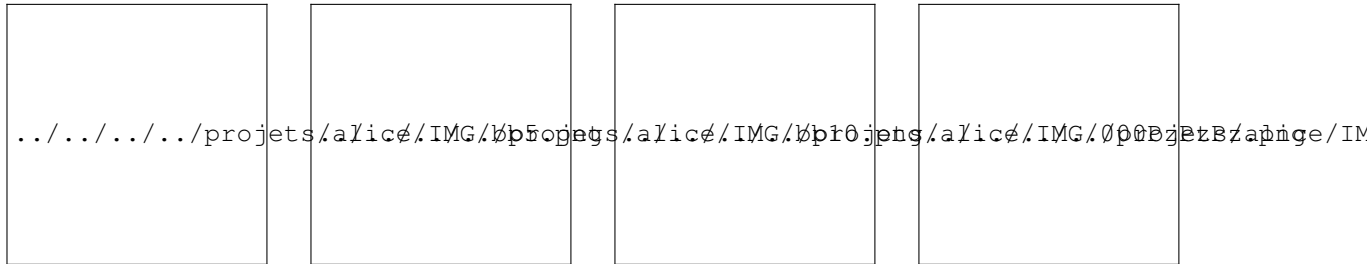


Figure 8. Typical editing examples of slender objects under various constraint, the faded snapshots shows initial or intermediate configurations.

This work was presented at the Eurographics conference and published in the “Computer Graphics Forum” journal [20].

### 5.2.2. Game Level Layout

**Participant:** Sylvain Lefebvre.

This work is a collaboration with the University of British Columbia. We consider a long standing problem in the video game industry: How to automatically generate game levels. Most procedural game levels tend to exhibit a random organization, reducing their interest. Instead, our approach lets a professional game designer describe the global organization of the level through a planar graph, capturing the connectivity and sequencing of different level ‘rooms’. Our approach then automatically generates multiple level geometries that correspond to this high-level description.

The work will be presented at *Eurographics 2014* [13].

### 5.2.3. Dynamic Element Textures

**Participant:** Sylvain Lefebvre.

This work is a collaboration with Microsoft Research Asia. We consider the problem of synthesizing animated details from an example. We first define the notion of a ‘textured’ animation and extract details from the example animation. Intuitively, these are small scale repetitive motions found for instance for leaves in the wind or in swarms. We then propagate these motions to a coarse scale animation. Our techniques work on 1D, 2D and 3D objects.

We published this work in *ACM Transactions on Graphics (SIGGRAPH proceedings)* [14].

### 5.2.4. Make It Stand: Balancing Shapes for 3D Fabrication

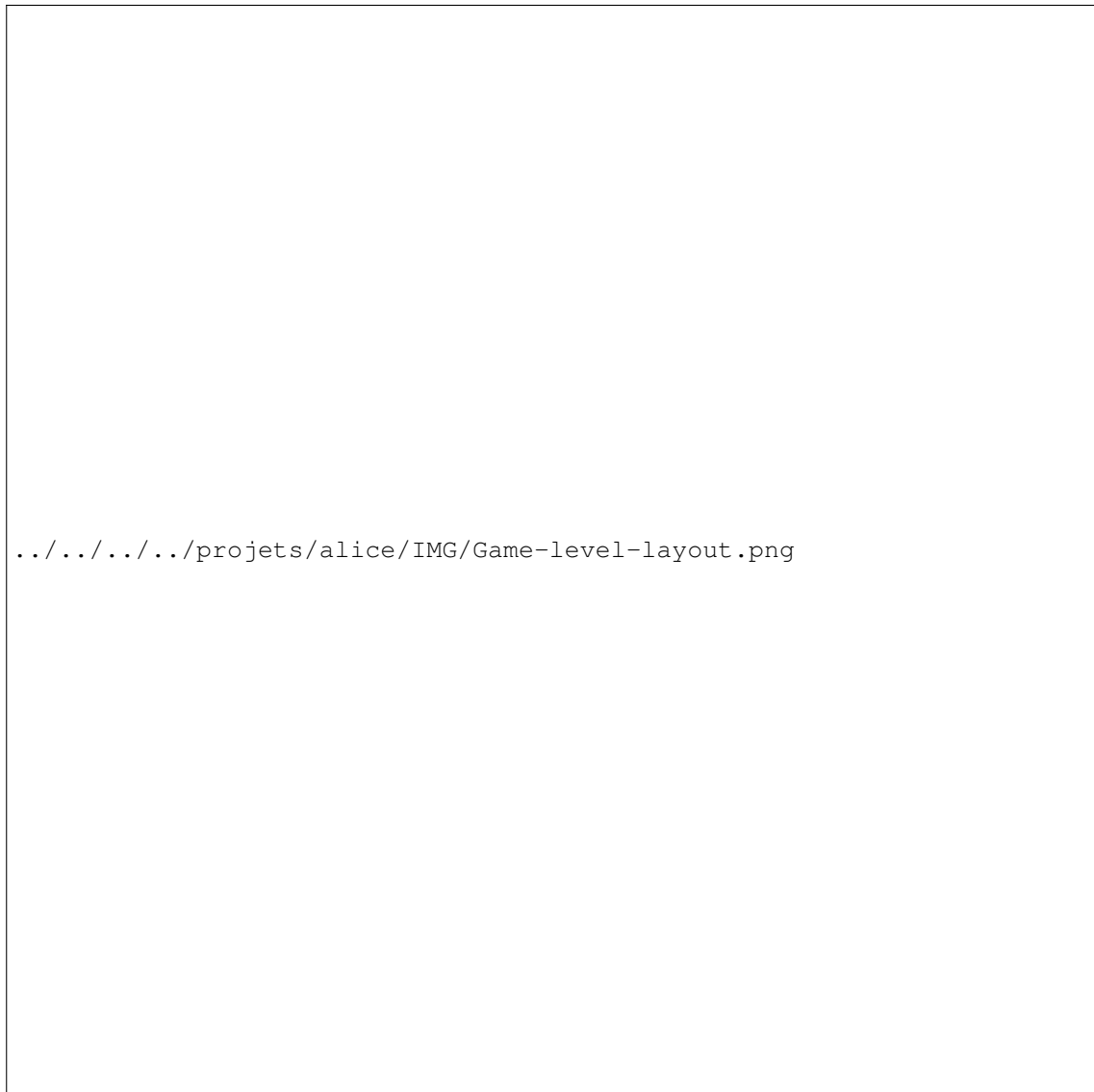
**Participant:** Sylvain Lefebvre.

This work is a collaboration with ETH Zurich. We consider the problem of balancing 3D models so that they stand in static equilibrium on their base of support after printing. We formulate the problem as the joint optimization of a voxel selection inside the model and a continuous detail preserving deformation of the outer surface.

The work has been published in *ACM Transactions on Graphics (SIGGRAPH proceedings)* [18].

### 5.2.5. Clean Colors

**Participants:** Jean Hergel, Sylvain Lefebvre.



*Figure 9. Game Level Layout*



*Figure 10. Dynamic Element Textures*

In this work we consider the problem of tool path planning for low-cost FDM (Fused Deposition Modeling) printers when using multiple filaments. Our method is based on three components which together reduce most of the defects found in such prints. Our algorithm first optimizes the orientation (azimuth angle) of the print so as to minimize defects. It then builds a rampart in close proximity of the model. This captures most of the strings of plastic oozing from idle extruders. Finally, we optimize for navigation paths minimizing the apparition of defects.

The work will be presented at *Eurographics 2014* [22].

### 5.2.6. *Fast Fragment Sorting on the GPU*

**Participants:** Sylvain Lefebvre, Samuel Hornus.

In this work, we build upon our result on “hashing on the GPU” from 2011 [1] to develop new techniques for sorting per-pixel lists of fragments as the latter are rasterized. We can then obtain, for each pixel, the list of surface elements visible through that pixel, sorted according to their distance to the viewpoint. The lists are obtained in a single rasterization pass instead of two for some earlier work; this is a clear win for bandwidth usage and processing time. Two important applications are the possibility to correctly visualize transparent objects and to directly display constructive-solid-geometry models without having to compute their boundary first (the boolean operations are performed on the fly, per pixel).

Our initial work has been published as a research report [25]. It has then been extended into a book chapter [24].

The techniques developed in this work are extensively used in our 3D printing software IceSL (see section 4.2).

### 5.2.7. *Techniques for Shooting Highly Coherent Rays*

**Participant:** Samuel Hornus.

This work explores novel ways to exploit the coherence of some set of rays used in the ray-tracing and other realistic image synthesis techniques. We propose new ways to traverse the usual data-structure for 3D indexes and leverage optimized and exact geometric predicates. Our first results give a faster ray shooting technique for pinhole camera rays and exhibit a remarkable increase in efficiency as the number of rays rises. A manuscript was submitted but not accepted to Eurographics.

## 5.3. Algorithms and analysis

**Participant:** Laurent Alonso.

### 5.3.1. *The Majority Problem*

Given a set of  $n$  elements each of which is either red or blue, Boyer and Moore’s algorithm uses pairwise equal/not equal color comparisons to determine the majority color. We analyze the average behavior of their algorithm, proving that if all  $2^n$  possible inputs are equally likely, the average number of color comparisons used is  $n - \sqrt{2n/\pi} + O(1)$  and has variance  $\frac{\pi-2}{\pi}n - \frac{\sqrt{2n}}{\sqrt{\pi}} + O(1)$ . This joint work with Edward M. Reingold was published in the IPL journal [8].

### 5.3.2. *The $X + Y$ Sorting Problem*

Some combinatorial approaches were taken to try to find bounds on the  $X + Y$  problem: Given two lists:  $X = (x_1, \dots, x_n)$ ,  $Y = (y_1, \dots, y_m)$ , determine the ordering of the values  $x_i + y_j$  for  $i \in [1, n], j \in [1, m]$ .

## 5.4. Fractal Geometry

**Participant:** Dmitry Sokolov.

Fractal geometry is a relatively new branch of mathematics that studies complex objects of non-integer dimensions. It finds applications in many branches of science as objects of such complex structure often exhibit interesting properties.

In 1988 Barnsley presented the Iterative Function System (IFS) model that allows modelling complex fractal shapes with only a limited set of contractive transformations. Later many other models were based on the IFS model such as Language-Restricted IFS, Projective IFS, Controlled IFS and Boundary Controlled IFS. The latter allows modeling complex shapes with control points and specific topology. These models cover classical geometric models such as B-splines and subdivision surfaces as well as fractal shapes.

This year we focused on the analysis of the differential behaviour of the shapes described with Controlled IFS and Boundary Controlled IFS. We derive the necessary and sufficient conditions for differentiability for everywhere dense sets of points. Our study is based on the study of the eigenvalues and eigenvectors of the transformations composing the IFS.

We apply the obtained conditions to modeling curves in surfaces. We describe different examples of differential behaviour presented in shapes modeled with Controlled IFS and Boundary Controlled IFS. We also use the Boundary Controlled IFS to solve the problem of connecting different subdivision schemes. We construct a junction between Doo-Sabin and Catmull-Clark subdivision surfaces and analyse the differential behaviour of the intermediate surface.

An article about this work is in the publication process in LNCS.

## 5.5. Scientific Computing for Linear and Nonlinear Wave Problems

**Participant:** Xavier Antoine.

We consider the Backward Euler SPectral (BESP) scheme that was proposed for computing the stationary states of Bose-Einstein Condensates (BECs) through the Gross-Pitaevskii equation. We show that the fixed point approach introduced earlier fails to converge for fast rotating BECs. A simple alternative approach based on Krylov subspace solvers with a Laplace or Thomas-Fermi preconditioner is given. Numerical simulations (obtained with the associated freely available Matlab toolbox GPELab) for complex configurations show that the method is accurate, fast and robust for 2D/3D problems and multi-components BECs.

This work was published in the journal “Journal of Computational Physics” [9].

## 5.6. Accelerating Structural Biology Software

**Participant:** Xavier Cavin.

This work is a collaboration with Dave Ritchie (team ORPAILLEUR, Nancy). The aim of this project is to leverage parallelism, multi-core computing and GPU in order to speed-up costly computations in cryo-electron microscopy. Several tools have been developed. Two of those “gEM tools” have been the subject of two articles were published in 2013 in “Journal of Structural Biology” [10] and “BMC Structural Biology” [11].

## AVIZ Project-Team

# 6. New Results

## 6.1. Hybrid-Image Visualizations

**Participants:** Petra Isenberg [correspondant], Pierre Dragicevic, Wesley Willett, Anastasia Bezerianos, Jean-Daniel Fekete.

We investigated hybrid-image visualization for data analysis in large-scale viewing environments. Hybrid-image visualizations blend two different visual representations into a single static view, such that each representation can be perceived at a different viewing distance. Our work was motivated by data analysis scenarios that incorporate one or more displays with sufficiently large size and resolution to be comfortably viewed by different people from various distances. Hybrid-image visualizations can be used, in particular, to enhance overview tasks from a distance and detail-in-context tasks when standing close to the display. By taking advantage of humans' perceptual capabilities, hybrid-image visualizations do not require tracking of viewers in front of a display. Moreover, because hybrid-images use a perception-based blending approach, visualizations intended for different distances can each utilize the entire display. In our paper we contributed a design space, discussed the perceptual rationale for our work, provided examples and a set of techniques for hybrid-image visualizations, and described tools for designing hybrid-image visualizations. We will also release software that will help in the construction of hybrid-image visualizations.



*Figure 8. Overview of a treemap showing a subset of the tree of life with a hybrid image visualization. Larger structures are clearly visible from far but do not interfere with reading detail when up close (see in-set).*

## 6.2. Visualization for Interactive Displays

**Participants:** Tobias Isenberg [correspondant], Petra Isenberg.



Because the access to and analysis of information is becoming increasingly important anywhere and at any time, researchers have begun to investigate the role of interactive displays as data analysis platforms. Visualization applications play a crucial role in data analysis and development of dedicated systems and tools for small to large interactive displays to support such application contexts is underway. We contribute a systematic and quantitative assessment of the literature from ten different venues, an open repository of papers, and a code-set that can be used to categorize the research space [22]. We found just over 100 publications at the intersection of interactive surfaces and visualization in our careful examination of 10 different publication venues related to the topic. We found that research has so far largely focused on the development of interaction techniques, for multi-touch tabletop devices, and 2D spatial and abstract visualizations. Together, all publications addressed a wide spectrum of research questions and, given the many possible combinations of interactive surfaces and visualization, the research space is still wide open. While several projects developed applications for data analysis with visualization on interactive surfaces, their availability in practice is still rare. Commercial companies and open-source communities have begun to provide ported versions of their products/tools for tablets and mobile phones (e.g., Tableau Mobile 4 and KiwiViewer 5), showing the need for visualization application on surfaces. Nevertheless, the support for data analysis tasks on these and other interactive surfaces can certainly still be improved—a lot more research with respect to the development and evaluation of the fundamentals of data exploration and analysis is needed for interactive displays. Some example directions of future work in this context were outlined in research agendas published in the journal *IEEE Computer* [24] as well as in *IEEE Computer Graphics and Applications* [21].

In a specific project investigated an interaction design concept for exploratory 3D data visualization that marries direct-touch interaction with stereoscopic vision. The design is inspired by the mental mapping that occurs for mouse interaction where the physical control space is mapped through a mental rotation to the display space. Similarly, we explore touch interaction on a monoscopic tablet, mapped through a mental rotation to the stereoscopic display space. Because this mental mapping can become increasingly skewed we show when and how to re-synchronize the views (see Figure 9).



Figure 9. Tablet-based navigation of a stereoscopically displayed 3D dataset.

### 6.3. Visualization for Soccer Analysis

**Participants:** Charles Perin, Romain Vuillemot, Jean-Daniel Fekete [correspondant].

A new generation of soccer data is now available, as some companies (<http://www.optasports.com/>) collect and provide extensive data covering almost all professional soccer championships, with a wealth of multivariate information related to time, player positions, and types of action, to name a few. Currently, most analysis on such data relate to statistics on individual players or teams. For instance, statistics on “team ball possession” and “number of goal attempts for team A or B” are popular on websites, TV and newspapers and often accompanied by bar charts or plots on a soccer field. However, soccer analysts we collaborated with consider that quantitative analysis alone does not convey the right picture of the game, as context, player positions and phases of player actions are the most relevant aspects.

SoccerStories [] (Figure 10 ) is a visualization interface to support analysts in exploring soccer data and communicating interesting insights that we designed to support the current practice of soccer analysts and to enrich it, both in the analysis and communication stages. Our system provides an overview+detail interface of game phases, and their aggregation into a series of connected visualizations, each visualization being tailored for actions such as a series of passes or a goal attempt. To evaluate our tool, we ran two qualitative user studies on recent games using SoccerStories with data from one of the world’s leading live sports data providers. The first study resulted in a series of four articles on soccer tactics, by a tactics analyst, who said he would not have been able to write these otherwise. The second study consisted in an exploratory follow-up to investigate design alternatives for embedding soccer phases into word-sized graphics. For both experiments, we received a very enthusiastic feedback and participants consider further use of SoccerStories to enhance their current workflow. This article received a Best Paper Honorable Mention in VIS 2013.

We also explored how spectators of a live soccer game can collect detailed data while watching the game [46]. Our motivation arose from the lack of free detailed sport data, contrasting with the large amount of simple statistics collected for every popular games and available on the web. Assuming many spectators carry a smart phone during a game, we implemented a series of input interfaces for collecting data in real time. In a user study, we asked participants to use those interfaces to perform tracking tasks such as locating players in the field, qualifying ball passes, and naming the player with ball while watching a video clip of a real soccer game. Our two main results are 1) the crowd can collect detailed and fairly complex data in real-time with reasonable quality while each participant is assigned a simple task, and 2) a set of design implications for crowd-powered interfaces to collect live sport data. We also discuss the use of such data into SoccerStories, and the design implications coming with the visual communication of missing and uncertain detailed data.

Finally, we presented R2S2 [45] in the SportVis workshop (VIS 2013), a hybrid visualization technique as an intermediate step between Rank Chart and Slope Graph to better understand and analyze team evolutions during soccer championships. Currently used rank tables for soccer are relative (ranked-based) and do not convey the absolute difference between teams. R2S2 provides a way to visualize these differences using the Slope Graph technique (value-based). By interactively setting the parameters of R2S2, we make the distance between teams appear, minimizing the overlaps caused by the Slope Graph technique.

More information about these projects is available at <http://www.aviz.fr/soccer>.

## 6.4. Interaction Model for Visualizations Beyond the Desktop

**Participants:** Yvonne Jansen [correspondant], Pierre Dragicevic.

We introduced an interaction model for beyond-desktop visualizations that combines the visualization reference model with the instrumental interaction paradigm. Beyond-desktop visualizations involve a wide range of emerging technologies such as wall-sized displays, 3D and shape-changing displays, touch and tangible input, and physical information visualizations. While these technologies allow for new forms of interaction, they are often studied in isolation. New conceptual models are needed to build a coherent picture of what has been done and what is possible. We described a modified pipeline model where raw data is processed into a visualization and then rendered into the physical world. Users can explore or change data by directly manipulating visualizations or through the use of instruments. Interactions can also take place in the physical world outside the visualization system, such as when using locomotion to inspect a large scale visualization. Through case



*Figure 10. Using SoccerStories: (a) navigating among soccer phases of a game; (b) mapping a phase on a focus soccer field; (c) exploring the phase by grouping actions into tailored visualizations; and (d) communicating using Sportlines embed into text.*

studies we illustrated how this model can be used to describe both conventional and unconventional interactive visualization systems, and compare different design alternatives.

## 6.5. Network Visualization

**Participants:** Benjamin Bach [correspondant], Basak Alper, Andre Spritzer, Emmanuel Pietriga, Nathalie Henry-Riche, Tobias Isenberg, Jean-Daniel Fekete.

Although much research has been done on finding efficient ways to visualize different kinds of networks (social networks, computer networks, brain networks, etc), many problems are still open. Rather than trying to find optimal layouts, we focus on novel representation and navigation techniques to explore such networks. Our research focusses on three major problems: (i) heterogeneous networks, (ii) comparison of graphs, (iii) dynamic networks, and (iv) generating networks for controlled user evaluations.

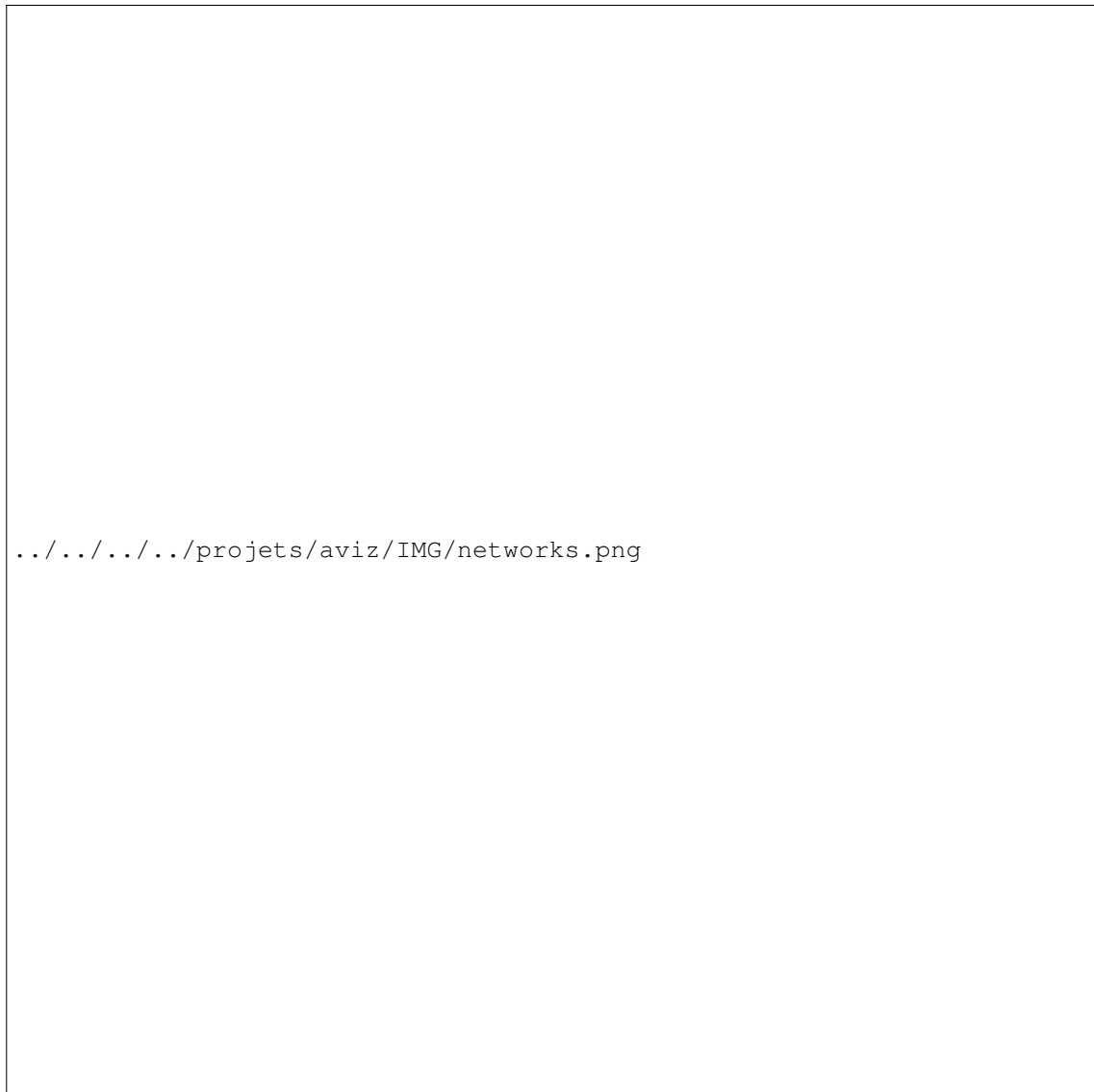
**Heterogeneous Networks:** Heterogeneous networks are networks with multiple node and edge types, such as ontologies in the Semantic Web. Ontologies also provide means to describe type hierarchies on node and edge types as well as other set theoretic relations between these types. Such heterogeneous networks can be explored on two levels, the model (types and possible relations between types), and instances (actual nodes and their connections). In order to allow for interactive exploration of such data, we extended the NodeTrix technique [67] to visualize ontologies. Our prototype is called *OntoTrix* [11] (Figure 11 (a)), which allows for reorganizing matrices by splitting and merging them, traversing node and edge hierarchies and visualize different types of connection between nodes.

**Graph Comparison:** Analyzing brain networks, which can represent anatomical fibers as well as functional correlation between brain regions, is complex in many ways. By analysing brain scientists tasks we concluded that many user tasks can actually performed by comparing two networks. In [28](Figure 11 (b)) we design and discuss several ways to compare two weighted graphs and finally compare the two most promising designs in a controlled user study. We found that our encoding for adjacency matrices outperforms the one for node-link diagrams, even for sparse networks. The implications for brain analysis tools are manifold and our results generalize to other domains that are concerned with comparing (dense and weighted) networks.

**Dynamic Networks:** A very common technique to explore dynamic networks are animations and small multiples, each of which being supporting different tasks, while falling short on others. With *GraphDiaries* [10](Figure 11 (c)), we design an interface based on the combination of both techniques while offering flexible temporal navigation techniques as well as enhanced perceptive feedback to understand changes between time steps. *GraphDiaries* supports further navigation techniques such as temporal aggregation, direct difference views and layout adjustment. While *GraphDiaries* is highly extensible, its techniques are designed to be integrated in existing visualization tools.

While animations and the techniques in *GraphDiaries* are useful for many networks, dense dynamic networks are still an important open problem. We hence generalized the idea of matrices to visualize temporal networks, by describing a visualization and interaction model based on the space time cube metaphor (Figure 11 (d)). In analogy with the physical world, this *Matrix Cube* can be manipulated and decomposed in order to explore the network, while the cube model serves as a consistent visual and mental model of the data and visualization. We implemented an interface called *Cubix* that allows us to perform simple view switches and decomposition operations in the cube. *Cubix* and the *Matrix Cube* was evaluated with two experts, an astronomer and brain scientist, exploring their own real world data. With the *Matrix Cube* and its decomposition operations, we are able to visualize and navigate within very dense dynamic networks such as brain networks, trading flows and technical networks. The design space of possible visualizations that the *Matrix Cube* and its operations offer is both, huge but structured at the same time. It allows us to explore many future designs.

As part of the effort of visualizing publications and work of Jean-Daniel Fekete, we designed a visualization to show his collaborations over the past years, relating his papers and his collaborators [29](Figure 11 (e)). A poster was presented at part of a poster submission to IEEE Vis, 2013 in Atlanta. To the best of our knowledge, no such technique to visualize any sort of dynamic ego networks have been published so far.



*Figure 11. Examples for network visualizations*

**Network Generation:** As any controlled user study, evaluating network visualizations requires control over the data. However, it is hard to find real world data with the desired properties and in reasonable amount for a controlled user study. Synthetic data can help but the output of random graph generators is hard to control and hardly resembles actual real-world data. With GraphCuisine [56] we present an interactive approach to generate graphs. In an iterative process, the computer generates suggestions while the user selects her preferred graphs and graph measures.

## 6.6. GridVis: Visualisation of Island-Based Parallel Genetic Algorithms

**Participants:** Waldo Cancino [correspondant], Hugo Gilbert, Benjamin Bach, Evelyne Lutton, Pierre Collet.

Island Model parallel genetic algorithms rely on various migration models and their associated parameter setting. A fine understanding of how the islands interact and exchange informations is an important issue for the design of efficient algorithms. GridVis, is an interactive tool that has been developed for visualising the exchange of individuals and the propagation of fitness values between islands. GridVis has been developed in Java, to monitor how the islands communicate: when and how much individuals of which fitness they effectively exchange during a run. We model the computer cluster that is running the island model, as dynamic network and use an adjacency matrix to show the relations (exchange between individuals) between nodes (computers) in the cluster (Figure 12 (a)). Several experiments have been performed on a grid and on a cluster to evaluate GridVis' ability to visualise the activity of each machine and the communication flow between machines. Experiments have been made on the optimisation of a Weierstrass function using the EASEA language, with two schemes: a scheme based on uniform islands and another based on specialised islands (Exploitation, Exploration and Storage Islands).



*Figure 12. Visualisation of a grid with 20 machines: Each computer in the cluster appears twice in the matrix, once as row and once as column. Cells inside the matrix indicate information about the interaction of computers during evolution, for example, the amount of individuals exchanged (read from row to column). Similar to heat maps activity (exchange of individuals) is mapped to darkness (dark cells indicate high exchange, bright cells show low exchange).*



## HYBRID Project-Team

### 6. New Results

#### 6.1. 3D interactive techniques

##### 6.1.1. Navigating in virtual environments with omnidirectional rendering

**Participants:** Jérôme Ardouin [contact], Anatole Lécuyer [contact], Maud Marchal.

The “FlyVIZ” enables humans to experience a real-time 360° vision of their surroundings for the first time. The visualization device combines a panoramic image acquisition system (positioned on top of the user’s head) with a Head-Mounted Display (HMD). The omnidirectional images are transformed to fit the characteristics of HMD screens. As a result, the user can see his/her surroundings, in real-time, with 360° images mapped into the HMD field-of-view.

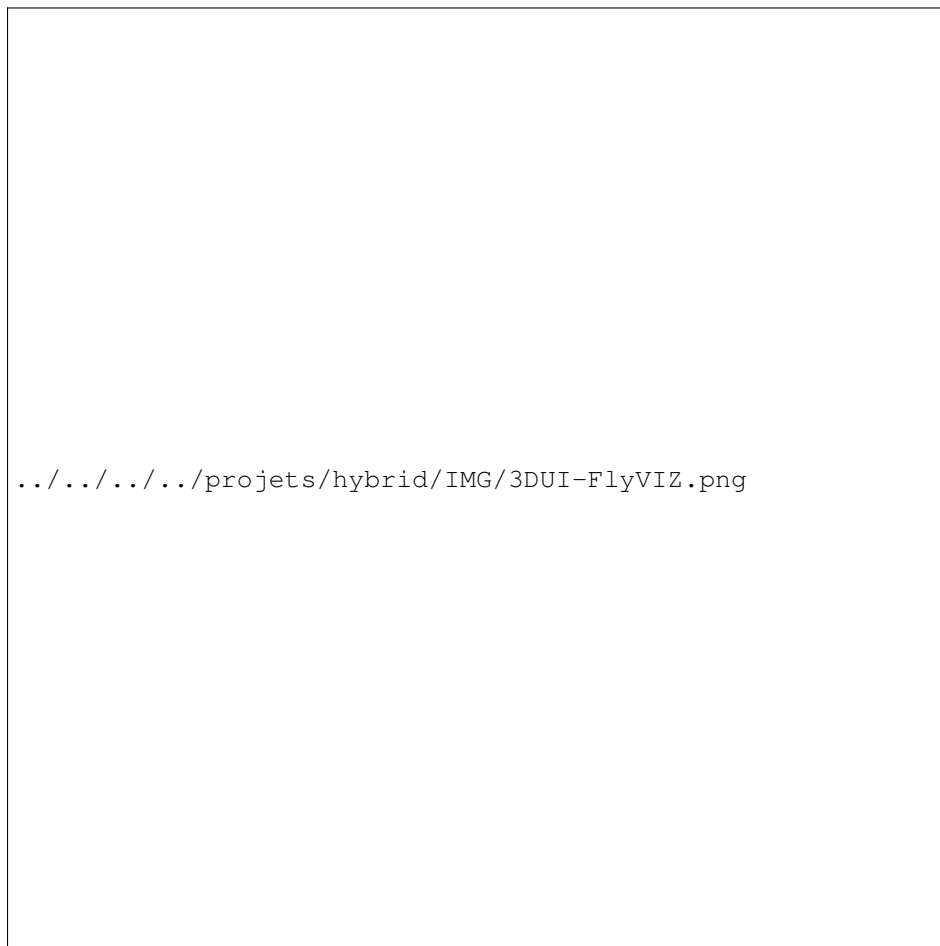


Figure 2. The “FlyVIZ” enables humans to experience in real-time a 360-degree vision of their surroundings.

In order to safely simulate and evaluate our approach, we designed and evaluated [27] several visualization techniques, for navigating in virtual environments (VE). We have conducted an evaluation of different methods compared to a rendering method of reference, i.e. a perspective projection, in a basic navigation task. Our results confirm that using any omnidirectional rendering method could lead to more efficient navigation in terms of average task completion time. Among the different 360° projection methods, the subjective preference was significantly given to a cylindrical projection method (equirectangular). Taken together, our results suggest that omnidirectional rendering could be used in virtual reality applications in which fast navigation or full and rapid visual exploration are important. They pave the way to novel kinds of visual cues and visual rendering methods in virtual reality. This work was a collaboration with the **Lagadic team** (Inria Rennes).

### 6.1.2. *Advances in locomotion interfaces for virtual environments*

**Participants:** Anatole Lécuyer [contact], Maud Marchal [contact], Bruno Arnaldi.

Navigation, a fundamental task in Virtual Reality (VR), is greatly influenced by the locomotion interface being used, by the specificities of input and output devices, and by the way the virtual environment is represented. No matter how virtual walking is controlled, the generation of realistic virtual trajectories is absolutely required for some applications, especially those dedicated to the study of walking behaviors in VR, navigation through virtual places for architecture, rehabilitation and training.

First, we have studied the realism of unconstrained trajectories produced during virtual walking. We proposed a comprehensive evaluation framework consisting on a set of trajecto-graphical criteria and a locomotion model to generate reference trajectories [16]. We considered a simple locomotion task where users walk between two oriented points in space. The travel path was analyzed both geometrically and temporally in comparison to simulated reference trajectories. This work was a collaboration with the **Mimetic team** (Inria Rennes).

Secondly, we have introduced novel “Camera Motions” (CMs) to improve the sensations related to locomotion in virtual environments (VE) [26]. Traditional CMs are artificial oscillating motions applied to the subjective viewpoint when walking in the VE, and they are meant to evoke and reproduce the visual flow generated during a human walk. Our novel CMs are: (1) multistate, (2) personified, and (3) they can take into account the topography of the virtual terrain. In addition, they can then take into account avatar’s fatigue and recuperation, and the topography for updating visual CMs accordingly. Taken together, our results suggest that our new CMs could be introduced in Desktop VR applications involving first-person navigation, in order to enhance sensations of walking, running, and sprinting, with potentially different avatars and over uneven terrains, such as for training, virtual visits or video games.

### 6.1.3. *3D manipulation of virtual objects: 3-Point++*

**Participants:** Thierry Duval [contact], Thi Thuong Huyen Nguyen.

Manipulation in immersive Virtual Environments (VEs) is often difficult and inaccurate because humans have difficulty in performing precise positioning tasks or in keeping the hand motionless in a particular position without any help of external devices or haptic feedback. To address this problem, we proposed a set of four manipulation points attached to objects (called a 3-Point++ tool, including three handle points and their barycenter), by which users can control and adjust the position of objects precisely [39]. By determining the relative position between the 3-Point++ tool and the objects, and by defining different states of each manipulation point (called locked/unlocked or inactive/active), these points can be freely configured to be adaptable and flexible to enable users to manipulate objects of varying sizes in many kinds of positioning scenarios.

### 6.1.4. *A survey of 3D object selection techniques for virtual environments*

**Participant:** Ferran Argelaguet Sanz [contact].

Computer graphics applications controlled through natural gestures are gaining increasing popularity these days due to recent developments in low-cost tracking systems and gesture recognition technologies. Although interaction techniques through natural gestures have already demonstrated their benefits in manipulation, navigation and avatar-control tasks, effective selection with pointing gestures remains an open problem. We surveyed the state-of-the-art in 3D object selection techniques [13]. We reviewed important findings in human control models, analyze major factors influencing selection performance, and classify existing techniques according to a number of criteria. Unlike other components of the application's user interface, pointing techniques need a close coupling with the rendering pipeline, introducing new elements to be drawn, and potentially modifying the object layout and the way the scene is rendered. Conversely, selection performance is affected by rendering issues such as visual feedback, depth perception, and occlusion management. We thus reviewed existing literature paying special attention to those aspects in the boundary between computer graphics and human computer interaction.

### 6.1.5. *Novel pseudo-haptic based interfaces*

**Participants:** Pierre Gaucher, Ferran Argelaguet Sanz, Anatole Lécuyer [contact], Maud Marchal.

Pseudo-haptics is a technique meant to simulate haptic sensations using visual feedback and properties of human visuo-haptic perception. In this course of action, we have extended its usage for gestural interfaces [32] and exploring its usage for the simulation of the local elasticity of images [].

Interacting with virtual objects through free-hand gestures do not allow users to perceive the physical properties of virtual objects. To provide enhanced interaction, we explored how the usage of a pseudo-haptic approach could be introduced while interacting with a 3D Carrousel [32]. In our approach, which is envisioned for showcasing purposes, virtual products are presented using a 3D carousel augmented with physical behavior and a pseudo-haptic effect aiming to attract the user to specific items. The user, through simple gestures, controls the rotation of the carousel, and can select, examine and manipulate the objects presented. Several demos can be tested on-line at [Hybrid website](#).

Secondly, we have introduced the Elastic Images, a novel pseudo-haptic feedback technique which enables the perception of the local elasticity of images without the need of any haptic device []. The proposed approach focuses on whether visual feedback is able to induce a sensation of stiffness when the user interacts with an image using a standard mouse. The user, when clicking on a Elastic Image, is able to deform it locally according to its elastic properties. A psychophysical experiment was conducted to quantify this novel pseudo-haptic perception and determine its perceptual threshold (or its Just Noticeable Difference). The results showed that users were able to recognize up to eight different stiffness values with our method and confirmed that it provides a perceivable and exploitable sensation of elasticity.

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

**Participant:** Valérie Gouranton [contact].

We designed a public experience and exhibition organized during the French National Days of Archaeology. This was the result of an interdisciplinary collaboration between archaeologists and computer scientists, centered on the immersive virtual reality platform Immersia, a node of the European Visionair project. This public exhibition had three main goals: (i) presenting our interdisciplinary collaboration, (ii) communicating on the scientific results of this collaboration, and (iii) offering an immersive experience in the past for visitors. In [33] we could present the scientific context of the event, its organization, and a discussion on feedbacks.

In the frame of the CNPAO project (section 8.1.3 ) we have also worked on the reconstitution of six archaeological sites located in the west of France ranging from prehistory to the Middle Ages: the Cairn of Carn Island, the covered pathway of Roh Coh Coet, the GohMin Ru megalithic site, the gallo-roman mansion of Vanesia, the keep of the Château de Sainte-Suzanne, the Porte des Champs of the Château d'Angers. Other proposals are currently under study [29].

### 6.1.7. *Perception of affordances in virtual reality*

**Participants:** Anatole Lécuyer [contact], Maud Marchal.



Figure 3. "Touching the past" experience during the French National Days of Archaeology.

The perception of affordances could be a potential tool for sensorimotor assessment of physical presence, that is, the feeling of being physically located in a virtual place. We have evaluated the perception of affordances for standing on a virtual slanted surface [25]. Participants were asked to judge whether a virtual slanted surface supported up right stance. The objective was to evaluate if this perception was possible in virtual reality (VR) and comparable to previous works conducted in real environments. We found that the perception of affordances for standing on a slanted surface in virtual reality is possible and comparable (with an underestimation) to previous studies conducted in real environments. We also found that participants were able to extract and to use virtual information about friction in order to judge whether a slanted surface supported an upright stance. Finally, results revealed that the person's position on the slanted surface is involved in the perception of affordances for standing on virtual grounds. Taken together, our results show quantitatively that the perception of affordances can be effective in virtual environments, and influenced by both environmental and person properties. Such a perceptual evaluation of affordances in VR could guide VE designers to improve their designs and to better understand the effect of these designs on VE users.

## 6.2. Haptic Feedback and Physical Simulation

### 6.2.1. Haptic feedback to improve audiovisual experience

**Participants:** Fabien Danieau, Anatole Lécuyer [contact].

Haptics have been employed in a wide set of applications ranging from teleoperation and medical simulation to arts and design, including entertainment, aircraft simulation and virtual reality. As for today, there is also a growing attention from the research community on how haptic feedback can be integrated with profit to audiovisual systems. We have first reviewed [18] the techniques, formalisms and key results on the enhancement of audiovisual experience with haptic feedback. We first reviewed the three main stages in the pipeline which are (i) production of haptic effects, (ii) distribution of haptic effects and (iii) rendering of haptic

effects. We then highlighted the strong necessity for evaluation techniques in this context and discuss the key challenges in the field. By building on technology and results from virtual reality, and tackling the specific challenges in the enhancement of audiovisual experience with haptics, we believe the field presents exciting research perspectives for which financial and societal stakes are significant.

We have then developed a novel approach called HapSeat for simulating motion sensations in a consumer environment. Multiple force-feedbacks are applied to the seated user's body to generate a 6DoF sensation of motion while experiencing passive navigation as illustrated Figure 4. A set of force-feedback devices such as mobile armrests or headrests are arranged around a seat so that they can apply forces to the user. The forces are computed consistently with the visual content (visual acceleration) in order to generate motion sensations. This novel display device has been patented and was demonstrated this year at ACM SIGGRAPH 2013 Emerging-Technologies [55], and ACM CHI 2013 Interactivity [54].



Figure 4. The HapSeat device: force-feedback is applied on the user's body with mobile armrests or headrests in order to generate motion sensations that are consistent with the visual content.

This work was a collaboration with the **Mimetic team** (Inria Rennes).

### 6.2.2. Vibrotactile rendering of splashing fluids

**Participants:** Anatole Lécuyer, Maud Marchal [contact].

Compelling virtual reality scenarios involving physically based virtual materials have been demonstrated using hand- based and foot-based interaction with visual and vibrotactile feedback. However, some materials, such as water and other fluids, have been largely ignored in this context. For VR simulations of real-world environments, the inability to include interaction with fluids is a significant limitation. Potential applications include improved training involving fluids, such as medical and phobia simulators, and enhanced user experience in entertainment, such as when interacting with water in immersive virtual worlds. We introduced the use of vibrotactile feedback as a rendering modality for solid-fluid interaction, based on the physical processes that generate sound during such interactions [15]. This rendering approach enables the perception

of vibrotactile feedback from virtual scenarios that resemble the experience of stepping into a water puddle or plunging a hand into a volume of fluid.

### 6.2.3. *Six-DoF haptic interaction with fluids, solids, and their transitions*

**Participants:** Anatole Lécuyer, Maud Marchal [contact].

Haptic interaction with different types of materials in the same scene is a challenging task, mainly due to the specific coupling mechanisms that are usually required for either fluid, deformable or rigid media. Dynamically-changing materials, such as melting or freezing objects, present additional challenges by adding another layer of complexity in the interaction between the scene and the haptic proxy. We have addressed these issues through a common simulation framework, based on Smoothed-Particle Hydrodynamics, and enable haptic interaction simultaneously with fluid, elastic and rigid bodies, as well as their melting or freezing [30]. We introduced a mechanism to deal with state changes, allowing the perception of haptic feedback during the process, and a set of dynamic mechanisms to enrich the interaction through the proxy. We decouple the haptic and visual loops through a dual GPU implementation. An initial evaluation of the approach was performed through performance and feedback measurements, as well as a small user study assessing the capability of users to recognize the different states of matter they interact with.

### 6.2.4. *Bimanual haptic manipulation*

**Participants:** Anatole Lécuyer [contact], Maud Marchal [contact], Anthony Talvas.

Bimanual haptics is a specific kind of multi-finger interaction that focuses on the use of both hands simultaneously. Several haptic devices enable bimanual haptic interaction, but they are subject to a certain number of limitations for interacting with virtual environments (VEs), such as workspace size issues or manipulation difficulties, notably with single-point interfaces. Interaction techniques exist to overcome these limitations and allow users to perform specific two-handed tasks, such as the bimanual exploration of large VEs and grasping of virtual objects. We have proposed an overview of the current limitations in bimanual haptics and the interaction techniques developed to overcome them. Novel techniques based on the Bubble technique are more specifically presented, with a user evaluation that assesses their efficiency. These include bimanual workspace extension techniques as well as techniques to improve the grasping of virtual objects with dual single-point interfaces. This work was published as a chapter in a book on “Multi-finger Haptic Interaction” [51].

### 6.2.5. *The god-finger method*

**Participants:** Anatole Lécuyer, Maud Marchal [contact], Anthony Talvas.

In physically-based virtual environments, interaction with objects generally happens through contact points that barely represent the area of contact between the user’s hand and the virtual object. This representation of contacts contrasts with real life situations where our finger pads have the ability to deform slightly to match the shape of a touched object. We have proposed a method called god-finger to simulate a contact area from a single contact point determined by collision detection, and usable in a rigid body physics engine [42]. The method uses the geometry of the object and the force applied to it to determine additional contact points that will emulate the presence of a contact area between the user’s proxy and the virtual object. It could improve the manipulation of objects by constraining the rotation of touched objects in a similar manner to actual finger pads. An implementation in a physics engine shows that the method could make for more realistic behaviour when manipulating objects while keeping high simulation rates. This work was presented at IEEE 3DUI Symposium 2013 and has received the best technote award [42].

### 6.2.6. *Collision detection for fracturing rigid bodies*

**Participant:** Maud Marchal [contact].



*Figure 5. 3D interaction techniques for bimanual haptic manipulation.*



In complex scenes with many objects, collision detection plays a key role in the simulation performance. This is particularly true for fracture simulation, where multiple new objects are dynamically created. We have proposed novel algorithms and data structures for collision detection in real-time brittle fracture simulations [21]. 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 proposed 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 reduced 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 brittle fracture simulations involving many dynamically created objects.



Figure 6. Example of brittle fracture with collision detection.

This work was a collaboration with the **Mimetic team** (Inria Rennes).

### 6.2.7. Collision detection with high performance computing on GPU

**Participants:** Bruno Arnaldi, Valérie Gouranton [contact], François Lehericéy.

We have first proposed IRTCD, a novel Iterative Ray-Traced Collision Detection algorithm that exploits spatial and temporal coherency. Our approach uses any existing standard ray-tracing algorithm and we propose an iterative algorithm that updates the previous time step results at a lower cost with some approximations. Applied for rigid bodies, our iterative algorithm accelerates the collision detection by a speedup up to 33 times compared to non-iterative algorithms on GPU [34].

Then, we have presented two methods to efficiently control and reduce the interpenetration without noticeable computation overhead. The first method predicts the next potentially colliding vertices. These predictions are used to make our IRTCD algorithm more robust to the approximations, therefore reducing the errors up to 91%. We also present a ray re-projection algorithm that improves the physical response of ray-traced collision detection algorithm. This algorithm also reduces, up to 52%, the interpenetration between objects in a virtual environment. Our last contribution showed that our algorithm, when implemented on multi-GPUs architectures, is far faster [35].

Finally, we proposed a distributed and anticipative model for collision detection and propose a lead for distributed collision handling, two key components of physically-based simulations of virtual environments. This model is designed to improve the scalability of interactive deterministic simulations on distributed systems such as PC clusters. Our main contribution consists of loosening synchronism constraints in the collision detection and response pipeline to allow the simulation to run in a decentralized, distributed fashion.



*Figure 7. Real-time simulation of iterative ray-traced collision detection algorithm.*

We could show the potential for distributed load balancing strategies based on the exchange of grid cells, and explain how anticipative computing may, in cases of short computational peaks, improve user experience by avoiding frame-rate drop-downs [31].

## 6.3. Brain-Computer Interfaces and Virtual Environments

### 6.3.1. Multi-user BCI video game

**Participant:** Anatole Lécuyer [contact].

How can we connect two brains to a video game by means of a BCI, and what will happen when we do so? How will the two users behave, and how will they perceive this novel common experience? We have created a multi-user videogame called “BrainArena” in which two users can play a simple football game by means of two BCIs [14], as illustrated Figure 8 . They can score goals on the left or right side of the screen by simply imagining left or right hand movements. To add another interesting element , the gamers can play in a collaborative manner (their two mental activities are combined to score in the same goal), or in a competitive manner (the gamers must push the ball in opposite directions). Two experiments were conducted to evaluate the performance and subjective experience of users in the different conditions. Taken together our results suggest that multi-user BCI applications can be operational, effective, and more engaging for participants.

This work was a collaboration with the **Potioc team** (Inria Bordeaux).

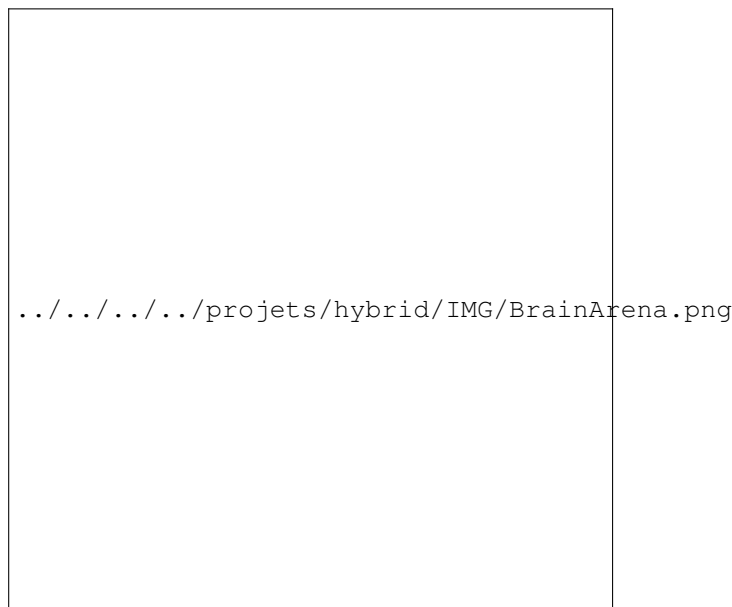


Figure 8. Multi-user football videogame in which two players can score goals to the left or right by imagining left or right hand movements. The users can play together using their brain activities either in a collaboration mode (same goal) or in a competitive mode (one versus the other).

### 6.3.2. Contextual SSVEP-based BCI control

**Participants:** Jozef Legény, Anatole Lécuyer [contact].

One main disadvantage of Brain-Computer Interfaces is that they are not completely reliable. In order to increase BCI performances, some adjustments can be made on low levels, such as signal processing and on high levels by modifying the controller paradigm. We have explored a novel, context-dependent, approach for SSVEP-based BCI controller [22]. This controller uses two kinds of behaviour alternation, commands can be added and removed if their use is irrelevant to the context or the actions resulting from their activation can be weighted depending on the likeliness of the actual intention of the user. This controller has been integrated within a BCI computer game and its influence in performance and mental workload has been addressed through a pilot experiment. Preliminary results have shown a workload reduction and performance improvement with the context-dependent controller while keeping the engagement levels untouched.

This work was a collaboration with the [Universidad de Jaen \(Spain\)](#).

### 6.3.3. *Can we use a BCI and manipulate a mouse at the same time?*

**Participants:** Jonathan Mercier-Ganady, Anatole Lécuyer [contact], Maud Marchal.

In most setups using a BCI, the user is explicitly asked to remain as motionless as possible, since muscular activity is commonly admitted to add noise and artifacts in brain electrical signals. Thus, as for today, people have been rarely let using other classical input devices such as mice or joysticks simultaneously to a BCI-based interaction. We have conducted an experimental study on the influence of manipulating an input device such as a standard computer mouse on the performance of a BCI system [37]. The study uses a simple virtual environment inspired by the well-known Pac-Man videogame and based on BCI and mouse controls. As expected the BCI performance was found to slightly decrease in presence of motor activity. However, we found that the BCI could still be successfully used in all conditions, even in presence of a highly-demanding mouse manipulation. These promising results pave the way to future experimental studies with more complex mental and motor activities, but also to novel 3D interaction paradigms that could mix BCI and other input devices for virtual reality and videogame applications.

### 6.3.4. *Adaptive VR simulators combining visual, haptic, and BCIs*

**Participants:** Anatole Lécuyer [contact], Maud Marchal.

What if the next generation of virtual reality simulators would take into account a novel user's input: his/her mental state, as measured with electrodes and Brain-Computer Interfaces ? This would lead to adaptive simulators that could match the "hidden" expectations of the user optimally? We have initiated and illustrated this promising path with a virtual reality setup in which the force-feedback of a guidance system is adapted in real-time to the "mental workload" of the user [23]. A first application of this approach is a medical simulator in which virtual assistances are automatically adapted to surgeon and trainee's mental activity as illustrated Figure 9 . Such results pave the way to future virtual reality systems which would automatically reconfigure and adapt to cerebral inputs and cognitive processes.

## 6.4. Collaborative Virtual Environments

### 6.4.1. *Collaborative exploration in multi-scale shared virtual environments*

**Participants:** Thierry Duval [contact], Thi Thuong Huyen Nguyen.

Exploration of large-scale 3D Virtual Environments (VEs) is often difficult because of lack of familiarity with complex virtual worlds, lack of spatial information that can be offered to users and lack of sensory details compared to the exploration of real environments. To address this problem, we presented a set of metaphors for assisting users in collaborative navigation to perform common exploration tasks in shared collaborative virtual environments [38], [56]. Our propositions consist in three guiding techniques in the form of navigation aids to enable one or several users to help one main user (exploring user) to explore the VE efficiently. These three techniques consist in drawing directional arrows, lighting up path to follow, and orienting a compass to show a direction to the exploring user. Our experimental results could show that although the directional arrows and compass surpassed the light source in a navigation task, these three techniques are completely appropriate for guiding a user in 3D complex VEs.



Figure 9. Medical simulator adapted to a BCI output. The user manipulates a virtual needle and has to insert it into a virtual liver to reach a tumor. Visual and haptic assistances are activated when a high mental workload is detected which corresponds to a more difficult manipulation of the needle.

#### 6.4.2. Improving the awareness of collaboration in 3D virtual environments

**Participants:** Thierry Duval [contact], Thi Thuong Huyen Nguyen, Valérie Gouranton.

When a user is fully immersed within a Virtual Environments (VE) through a large immersive display system, his feeling of presence can be altered because of disturbing interactions with his 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 and to embed them in the VE. In [19] we could present how we use the Immersive Interactive Virtual Cabin (IIVC) model to obtain such a virtual representation of the physical environment of the user and we illustrated how it can be used to guide efficiently a user for a 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 users in order to understand the actions of the other users even if they are not fully immersed in the shared virtual environment. Last, we explained how we could enhance 3D interaction and collaboration by embedding a symbolic 3D representation of the user that would give 3D information about his posture.

#### 6.4.3. Sharing and bridging information: application to ergonomics

**Participant:** Thierry Duval [contact].

We introduced a collaborative virtual environment usable to conduct ergonomic design sessions, involving the worker, ergonomists and engineers [40]. We focused particularly on the representation of the ergonomic evaluation and the interaction between an ergonomist and the main user (worker). An ergonomic evaluation of the postures was presented. An interaction architecture between the main user and an ergonomist based on the combination of animation modes of two linked manikins was also proposed. Preliminary results and future developments of the CVE (e.g. additional ergonomic evaluation tools, graphical enhancement, interaction enhancement) were then presented.

#### **6.4.4. User embodiment and collaboration in virtual environments for training**

**Participants:** Bruno Arnaldi, Valérie Gouranton [contact], Thomas Lopez, Florian Nouviale, Rozenn Bouville Berthelot.

In Collaborative Virtual Environments for Training (CVET), a group can learn and practice the completion of a task as a team using all the assets provided by Virtual Reality. We presented a novel mechanism that allows real and virtual humans to dynamically exchange the control of their embodiment in virtual environments [41]. Such a mechanism raises two important issues: the possibility of dynamic embodiment exchanges between real humans and virtual humans and the continuity of actions of the team members after an exchange. To address these issues we introduce a new entity, the Perceptive Puppet that abstracts real and virtual humans into one common entity containing its own knowledge.

In addition, in CVET different roles need to be played by actors, i.e. virtual agents or users. In order to abstract an actor from its embodiment in the virtual world, we have introduced a new entity, the Shell [36]. Through the Shell, users and virtual agents are able to collaborate in the same manner during the training. In addition to the embodiment's control, the Shell gathers and carries knowledge and provides interaction inputs. This knowledge and those inputs can be accessed and used homogeneously by both users and virtual agents to help them to perform the procedure.

## IMAGINE Project-Team

# 5. New Results

## 5.1. Introduction

We are developing user-centred, knowledge-based models in three main domains: shape, motion and narrative design, leading us to three research axes. The fourth one is the combination of these models with intuitive interaction tools, in order to set up interactive creative environments dedicated to specific categories of content. The following sections describe our activities in 2013 for each axis.

## 5.2. High level model for shapes

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

### 5.2.1. Implicit modeling

**Participants:** Antoine Bégault, Adrien Bernhardt, Marie-Paule Cani, Mohamed-Galal Koraa, Cédric Zanni.

Implicit surfaces are an appealing representation for free-form, volumetric shapes. In addition to being able to represent shapes of arbitrary topological genus, they have the ability to be constructed by successively blending different components, which eases interactive modeling.

In collaboration with Loic Barthe in Toulouse, we contributed to a new binary blending operator, called Gradient Blending [7], which enables us to blend implicit shapes not only in function of their field values but also of their gradients. This solves a number long standing problems in implicit modeling: we can generate bulge-free blending, ensure that the topological genus of the blended shape remains the one of the union of the input one, and avoid the blur of small details.



Figure 4. Example of surface generated using our SCALIS approach.

Within Cédric Zanni's PhD [2] we introduced closed-form solutions for convolution surfaces along helical skeletons and extended Gabor-noise texturing to enable the creation of repetitive geometric details along implicit surfaces. We also developed a novel extension to convolution surfaces, so-called SCALe-invariant Integral Surfaces (SCALIS) [15], see Figure 4. Thanks to their scale invariant blending properties, these surfaces have three main advantages: the radius of the surface around a skeleton can be explicitly controlled, shapes are self-similar regardless of the scale of the model, and thin components are not smoothed-out when blended into larger ones. This is done while preserving the main benefits of integral surfaces, namely n-ary blending with a simple plus, and shape preservation whatever the way the skeletons is split into smaller primitives. We are currently extending this work to enable the topology of the implicit surface to always reflect the one of the skeleton

### 5.2.2. Analysis of CAD models

**Participants:** Francois Faure, Stefanie Hahmann, Jean-Claude Léon, Olivier Palombi, Flavien Boussuge, Ahmad Shahwan.

CAD models, as part of assemblies defining manufactured products, are often shaped in accordance with their physical counterpart. However, one can observe that the shape of some components, as modeled in CAD, may differ from that of their physical instance. In addition, assemblies representing products are most often reduced to a collection of CAD models representing each component and the designation of each component is neither a reliable information nor a faithful connection with one or more functions of a component. As a result, geometric interfaces between components are unknown and they cannot be reduced to contacts. Interferences may exist that are also relevant for several applications. Determining precisely, the geometric interfaces between components is a first requirement to enrich geometric models with functional information because a subset of functions derive from interfaces between components.

As an example, this is particularly useful for structural mechanics to be able to generate rapidly a Finite Element model of assemblies and it is especially critical when assemblies get very complex. [9] addresses the problem to generate automatically a class of geometric interfaces for very complex assemblies (see fig. 5). GPU-based algorithms have proved suitable to obtain reliable results on CAD models.

Using these geometric interfaces as well as the newly introduced concept of conventional interfaces, [6], [4] and [11], [12], [19] have proposed an approach using qualitative reasoning, ontology reasoning to connect CAD components, their geometric interfaces, to functions and functional designations of components: an intrinsic identifier of a component in an assembly that connects it to its function. As a result, it is shown how geometric models of components need to be restructured, which extends the concept of annotation presently reduced to a elementary link between geometric models and symbolic information.

At the level of assembly components, shape analysis [30] is particularly useful to generate dimensionally reduced models needed for structural mechanics. [3] 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 [18], [32].

### 5.2.3. Knowledge-based shape transfert

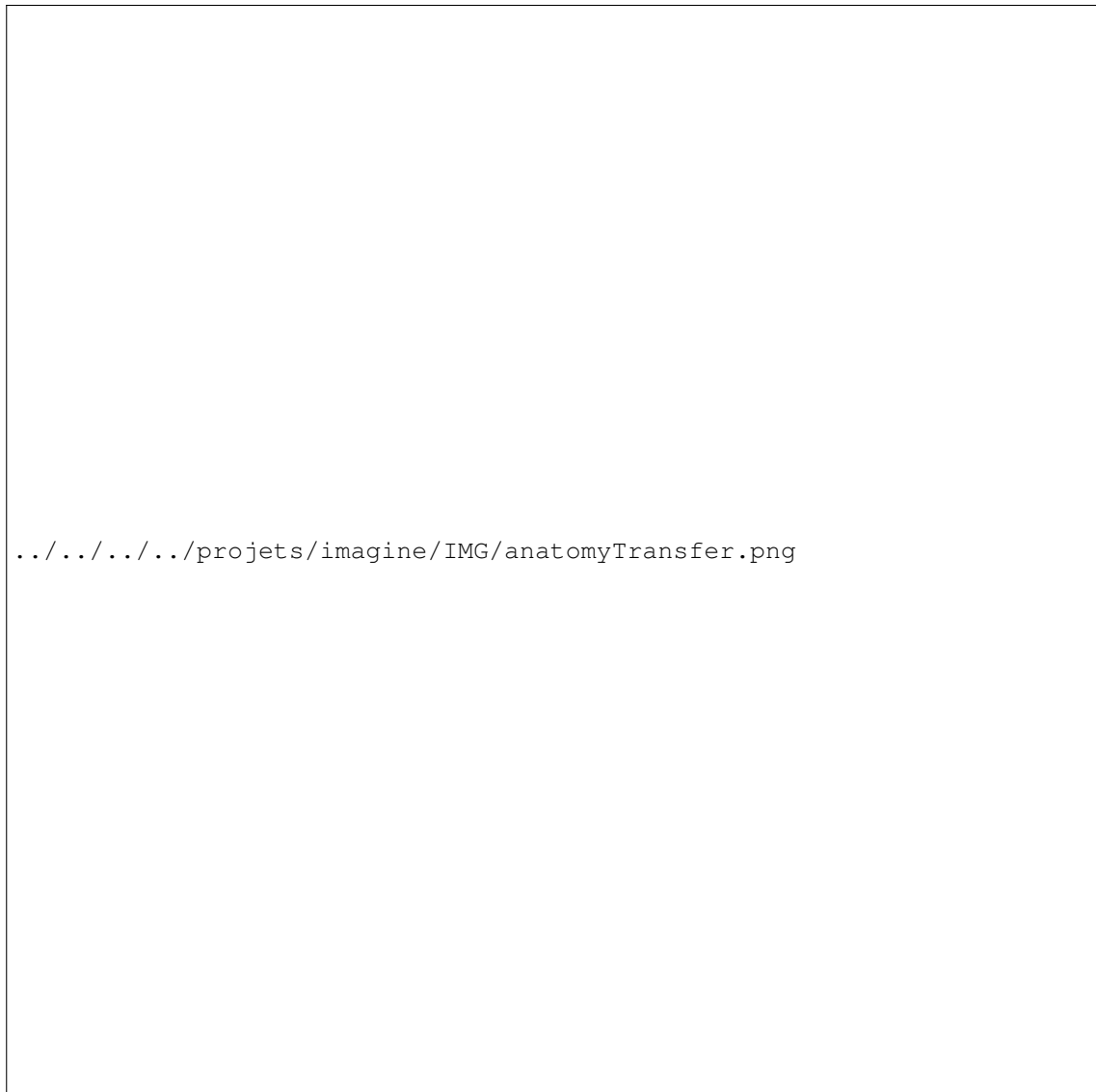
**Participants:** Marie-Paule Cani, Ali Dicko, Francois Faure, Olivier Palombi.

Characters with precise internal anatomy are important in film and visual effects, as well as in medical applications. We have proposed the first semi-automatic method for creating anatomical structures, such as bones, muscles, viscera and fat tissues [5], as illustrated in 6. This is done by transferring a reference anatomical model from an input template to an arbitrary target character, only defined by its boundary representation (skin). The fat distribution of the target character needs to be specified. We can either infer this information from MRI data, or allow the users to express their creative intent through a new editing tool. The rest of our method runs automatically: it first transfers the bones to the target character, while maintaining their structure as much as possible. The bone layer, along with the target skin eroded using the fat thickness information, are then used to define a volume where we map the internal anatomy of the source model using harmonic (Laplacian) deformation. This way, we are able to quickly generate anatomical models for a large range of target characters, while maintaining anatomical constraints.





*Figure 5. Computation of assembly interfaces in NURBS representation.*



*Figure 6. A reference anatomy (left) is automatically transferred to arbitrary humanoid characters. This is achieved by combining interpolated skin correspondences with anatomical rules.*

### 5.3. Models for motion and animation

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

#### 5.3.1. Physical models

**Participants:** Marie-Paule Cani, François Faure, Pierre-Luc Manteaux.

**Frame-based deformable solids** Our frame-based deformable model was published as a book chapter [31]. It combines the realism of physically based continuum mechanics models and the usability of frame-based skinning methods, allowing the interactive simulation of objects with heterogeneous material properties and complex geometries. The degrees of freedom are coordinate frames. In contrast with traditional skinning, frame positions are not scripted but move in reaction to internal body forces. The deformation gradient and its derivatives are computed at each sample point of a deformed object and used in the equations of Lagrangian mechanics to achieve physical realism. We introduce novel material-aware shape functions in place of the traditional radial basis functions used in meshless frameworks, allowing coarse deformation functions to efficiently resolve non-uniform stiffnesses. Complex models can thus be simulated at high frame rates using a small number of control nodes.



Figure 7. A dam break simulation with 5000 particles simulated with WCSPH (on the left) and with our adaptive method (on the right). On the right image, blue corresponds to full-dynamics particles, green to transition particles and red to restrained particles.

**Adaptive particle simulation** In collaboration with the NANO-D Inria Team, we have explored the use of Adaptively Restrained (AR) particles for graphics simulations [25]. Contrary to previous methods, Adaptively Restrained Particle Simulations (ARPS) do not adapt time or space sampling, but rather switch the positional degrees of freedom of particles on and off, while letting their momenta evolve. Therefore, inter-particles forces do not have to be updated at each time step, in contrast with traditional methods that spend a lot of time there. We first adapted ARPS to particle-based fluid simulations, as illustrated in 7 and proposed an efficient incremental algorithm to update forces and scalar fields. We then introduced a new implicit integration scheme enabling to use ARPS for cloth simulation as well. Our experiments showed that this new, simple strategy for adaptive simulations can provide significant speedups more easily than traditional adaptive models.

#### 5.3.2. Skinning virtual characters

**Participants:** Marie-Paule Cani, Damien Rohmer.

Skinning is a widely used technique to deform articulated virtual characters. It can be computed fastly and therefore can deliver real-time feedback at the opposite of physically based simulation. Still standard skinning approaches cannot handle well large deformations and may require manual corrections.

In collaboration with Loic Barthe and Rodolphe Vaillant from University of Toulouse, and collaborators from Victoria University, Inria Bordeaux and University of Bath, we develop a new automatic correction for skinning deformation that has been published in SIGGRAPH [14]. Based on the volumetric implicit representation paradigm, it adjust the mesh vertices and improves the visual appearance of the deformed surface. Moreover, it seamlessly handle skin contact ensuring that no self collision can occurs as seen in fig. 8 . Finally, the method can mimic muscular bulges controlled by the implicit blending operators described in the work [7].



Figure 8. Example of large skinning deformation obtained in [14].

### 5.3.3. Animating crowds

**Participants:** Marie-Paule Cani, Quentin Galvane, Kevin Jordao, Kim Lim.

Crowd animation is an interesting case, since it can be either computed by developing artificial intelligence methods, by using physically-based simulation of some extended particle systems, or by applying a kinematic

texturing methodology, made possible by the repetitive nature of crowd animations. We launched this new topic in the group in 2013, enabling us to explore the two last crowd animation methods:

Firstly, in collaboration with the University SAINTS, Malaysia, we extended particle-based crowd simulation to the case when 4 different populations, with different goals and behaviors, are interacting within the same environment [24]. This as illustrated by a cultural heritage application, with the reconstruction of past life in a harbor in Malaysia in the 19th century: see Figure 9 .



Figure 9. Crowd simulation with 4 different populations from [24].

Secondly, within the ANR project CHROME with Inria Rennes, we adopted the crowd-patches technique, i.e. the idea of combining patches carrying pre-computed crowd trajectories, for quickly populating very large

environments [23]. We are currently developing novel methods for enabling the interactive space-time editing of these animations (a paper will be published at the next Eurographics conference).

## 5.4. Knowledge-based models for narrative design

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

### 5.4.1. Cinematographic virtual camera control

**Participants:** Marie-Paule Cani, Quentin Galvane, Vineet Gandhi, Chen Kim Lim, Rémi Ronfard.

Steering Behaviors for Autonomous Cameras [21] : We proposed a new method for automatically filming crowd simulations with autonomous cameras, using specialized camera steering behaviors and forces. Experimental results show that the method provides a good coverage of events in moderately complex crowds simulations, with consistently correct image composition and event visibility.

The prose storyboard language [26] : We presented a formal language for describing movies shot by shot, where each shot is described with a unique sentence. The language uses a simple syntax and limited vocabulary borrowed from working practices in traditional movie-making, and is intended to be readable both by machines and humans. The language is designed to serve as a high-level user interface for intelligent cinematography and editing systems.

### 5.4.2. Virtual actors

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

Audio-Visual Speaker Conversion using Prosody Features [17] : We presented a new approach towards speaker identity conversion using speech signals and 3D facial expressions. Audio prosodic features are extracted from time alignment information for a better conversion of speaking styles. A subjective evaluation was performed to illustrate that the converted sequences are perceived as belonging to the target speakers. We are working to extend that approach to visual prosody features and to apply it to the situation where a director controls the expressions of a virtual actor, while maintaining its personality traits.

### 5.4.3. Narrative analysis of video

**Participants:** Vineet Gandhi, Rémi Ronfard.

Naming and detecting actors in movies [22]: We proposed a generative model for localizing and naming actors in long video sequences. More specifically, the actor's head and shoulders are each represented as a constellation of optional color regions. Detection can proceed despite changes in view-point and partial occlusions. This work is being extended to the case of theatre actors during performances and rehearsals. It also opens the way to future work in automatic analysis of cinematographic and editing styles in real movie scenes. This was also presented as a poster at the International Conference on Computational Photography (ICCP).

Recording theatre rehearsals [29] : We presented a contribution to the International Federation for Theatre Research describing our ongoing collaboration with the Theatre des Celestins in Lyon, emphasising that high quality video recordings make it possible to study the genetic evolution of a theatre performance, and make it an object of scientific study as well as an object of aesthetic appreciation.

## 5.5. Creating and interacting with virtual prototypes

- **Scientist in charge:** Jean-Claude Léon
- **Other permanent researchers:** Marie-Paule Cani, Olivier Palombi, Damien Rohmer, Rémi Ronfard.

### 5.5.1. Sketch-based modeling

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

The Line of Action [8] : The line of action is a conceptual tool often used by cartoonists and illustrators to help make their figures more consistent and more dramatic. In this paper, we proposed a mathematical definition of the line of action (LOA), which allows us to automatically align a 3D virtual character to a user-specified LOA by solving an optimization problem. This work is now being extended to the more challenging case of creating complete animations from storyboard-like hand-drawn sketches (see fig. 10).



*Figure 10. The line of action defines the pose of the virtual character.*

### **5.5.2. Sculpting methods**

**Participants:** Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer, Lucian Stanculescu.

Sculpting methods is a very powerful approach to design virtual models from an existing model. In the work of Lucian Stanculescu [13] we extend the standard sculpting paradigm of surfaces, to multi-dimensional nested structures. In this method, lower dimensional structures such as points and curves can be defined on the surface to defined a nested structure. Each part can follow a specific deformation behaviors. We therefore categorize the geometrical and topological behavior of the structure (such as rigidity or mutability) to develop a wider range of possible deformation. This method facilitate the persistence of sharp features that automatically split or merge with variable rigidity, even when the shape chages genus. This approach enable to deform a surface exhibiting typical behavior of both CAD model with sharp edges, and CG model with smooth surfaces as seen in fig. 11 .



Figure 11. Deformed surface with nested structure from [13].

We also extends the sculpting approach to handle detailed surfaces. During sculpting deformation such as local stretching, the surface details should not extend as the global shape, but rather duplicates to ensure that the surface keeps his detailed appearance. We studied this question under two different approaches.

The first one, in collaboration with Max-Planck Institute focussed on the deformation of 1D-like parametric structure such as castle walls of centripede characters. The method enable to freely extend, compress, split and merges parts of the structures. The deformed structure is generated by an assembly of basic parts whose behaviors are encoded using a discrete shape grammar. During deformation, the system finds the most suitable



collection of parts to assemble and ensure that the global shapes is coherent with the input rules. This work as been published in Eurographics 2014 [10].

The second approach consists in extending sculpting to continuous freeform deformation of a 2D surface with details. During the deformation gesture of stretching or compression, details on the surfaces should seamlessly appear or disappear continuously. In this work, we studied the simpler case of a planar surface with high field details and presented our result in GTMG [27]. We now work on the more general extension to this work as a collaboration with Max-Plack Institute and University of College London.

### 5.5.3. *Interactive control of procedural models: terrains and waterfalls*

**Participants:** Adrien Bernhardt, Marie-Paule Cani, Arnaud Emilien, Ulysse Vimont.

Procedural models, used for easily modeling large, natural environments, pose a specific challenge in terms of user control: how can these automatic methods, useful for quickly generating a huge number of self-similar details, be adapted to allow the coarse to fine level of control needed by the users?

This topic was first explored within Adrien Bernhardt's PhD thesis [1], where we introduced a real-time terrain modeling tool using a fast GPU-based terrain solver with a lightweight CPU-based data structure. This tool was recently extended in collaboration with Cambridge University, to enable first-person sketch-based editing of terrains models.

Secondly, we have been working on interactive procedural modeling of plausible waterfalls, in collaboration with Montreal University. Offering interactive user control for this application is particularly challenging, since the shape taken by a fall heavily depends on the underlying terrain. Our solution, based on vectorial user-control, on a flow solver, and on procedural adaptation of the underlying terrain, enable users to quickly create plausible flowing water, while controlling which fall segments are in contact with the terrain (vs. in free fall), the topology of the network, and how much the flow should adapt to the current terrain, vs. the terrain to the user-designed trajectories (fig. 12 ). A paper is under review and a presentation has been made at the AFIG conference [16].



Figure 12. Waterfalls modeling using our approach developed in [16]

### 5.5.4. *Interaction methods*

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

We are currently exploring the use of multi-touch tables for the interactive design and editing of 3D scenes, in collaboration with the human-computer interaction group of LIG laboratory. The main challenge here is to find out how to use a 2D interaction media for editing 3D content, hence how to intuitively control the third dimension (depth, non-planar rotations, 3D deformations, etc).

Our first work consisted in an user study where we analyzed all possible hand interactions on table-tops and explored the ways users would intuitively try to manipulate 3D environments, either for changing the camera position or for moving objects around [20]. 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 are to be bent or twisted in 3D using 2D interaction.

## IN-SITU Project-Team

# 6. New Results

## 6.1. Interaction Techniques

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

We explore interaction 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 through standard desktop and tabletops. This year, we investigated how people can use different body parts and limbs to convey information to interactive systems. *BodyScape* provides a framework for analysing and designing interaction techniques that involve the entire human body. Both *WristPointing*, which overcomes the limited range of motion of the wrist, and *HeadPad*, which takes the user's head orientation into account, are whole body techniques that facilitate target acquisition. *Arpege* can interpret a wide range of chord gestures, designed according to the range of motion and limitations of the human hand, and includes a dynamic guide with integrated feedforward/feedback to enhance learning by novices, without slowing down experts. On mobile devices, we designed novel interaction techniques that increase the expressivity of gestures by a single finger, including *ThumbRock*, based on movement dynamics, *SidePress*, which senses pressure on the device, and *Powerup*, which detects proximity. We also continued to develop advanced interactive visualization techniques, including *Gimlenses*, which supports focus+context representations for navigating within 3D scenes.

*BodyScape* – The entire human body plays a central role in interaction. The *BodyScape* design space [34] (honorable mention at CHI 2013) explores the relationship between users and their environment, specifically how different body parts enhance or restrict movement for specific interactions. *BodyScape* can be used to analyze existing techniques or suggest new ones. In particular, we used it to design and compare two free-hand techniques, on-body touch and mid-air pointing, first separately, then combined. We found that touching the torso is faster than touching the lower legs, since it affects the user's balance; and touching targets on the dominant arm is slower than targets on the torso because the user must compensate for the applied force.

*HeadPad* – Rich interaction with high-resolution wall displays is not limited to remotely pointing at targets. Other relevant types of interaction include virtual navigation, text entry, and direct manipulation of control widgets. However, most techniques for remotely acquiring targets with high precision have studied remote pointing in isolation, focusing on pointing efficiency and ignoring the need to support these other types of interaction. We investigated high-precision pointing techniques capable of acquiring targets as small as 4 millimeters on a 5.5 meters wide display while leaving up to 93 of a typical tablet device's screen space available for task-specific widgets [27]. We compared these techniques to state-of-the-art distant pointing techniques and have shown that two of our techniques, a purely relative one and one that uses head orientation, perform as well or better than the best pointing-only input techniques while using a fraction of the interaction resources.

*WristPointing* – Wrist movements are physically constrained and take place within a small range around the hand's rest position. We explored pointing techniques that deal with the physical constraints of the wrist and extend the range of its input without making use of explicit mode-switching mechanisms [33]. Taking into account elastic properties of the human joints, we investigated designs based on rate control. In addition to pure rate control, we examine a hybrid technique that combines position and rate-control and a technique that applies non-uniform position-control mappings. Our experimental results suggest that rate control is particularly effective under low-precision input and long target distances. Hybrid and non-uniform position-control mappings, on the other hand, result in higher precision and become more effective as input precision increases.

*Arpege* – While multi-touch input has become a standard for interacting with devices equipped with a touchscreen with simple techniques like pinch-to-zoom, the number of gestures systems are able to interpret remains rather small. Arpège [23] is a progressive multitouch input technique for learning chords, as well as a robust recognizer and guidelines for building large chord vocabularies. We conducted two experiments to evaluate our approach. Experiment one validated our design guidelines and suggests implications for designing vocabularies, i.e. users prefer relaxed to tense chords, chords with fewer fingers and chords with fewer tense fingers. Experiment two demonstrated that users can learn and remember a large chord vocabulary with both Arpège and cheat sheets, and Arpège encourages the creation of effective mnemonics.

*ThumbRock* – Compared with mouse-based interaction on a desktop interface, touch-based interaction on a mobile device is quite limited: most applications only support tapping and dragging to perform simple gestures. Finger rolling provides an alternative to tapping but uses a recognition process that relies on either per-user calibration, explicit delimiters or extra hardware, making it difficult to integrate into current touch-based mobile devices. We introduce ThumbRock [19], a ready-to-use micro gesture that consists in rolling the thumb back and forth on the touchscreen. Our algorithm recognizes ThumbRocks with more than 96% accuracy without calibration nor explicit delimiter by analyzing the data provided by the touch screen with a low computational cost. The full trace of the gesture is analyzed incrementally to ensure compatibility with other events and to support real-time feedback. This also makes it possible to create a continuous control space as we illustrate with our MicroSlider, a 1D slider manipulated with thumb rolling gestures.

*SidePress* – Virtual navigation on a mobile touchscreen is usually performed using finger gestures: drag and flick to scroll or pan, pinch to zoom. While easy to learn and perform, these gestures cause significant occlusion of the display. They also require users to explicitly switch between navigation mode and edit mode to either change the viewport's position in the document, or manipulate the actual content displayed in that viewport, respectively. SidePress [31] augments mobile devices with two continuous pressure sensors co-located on one of their sides (Figure 9 -(Left)). It provides users with generic bidirectional navigation capabilities at different levels of granularity, all seamlessly integrated to act as an alternative to traditional navigation techniques, including scrollbars, drag-and-flick, or pinch-to-zoom. We built a functional hardware prototype and developed an interaction vocabulary for different applications. We conducted two laboratory studies. The first one showed that users can precisely and efficiently control SidePress; the second, that SidePress can be more efficient than drag-and-flick touch gestures when scrolling large documents.

*Powerup* – Current technology like Arduino (<http://arduino.cc/>) opens a large space for designing new electronic device. We built the Power-up Button [30] by combining both pressure and proximity sensing to enable gestural interaction with one thumb (Figure 9 -(Right)). Combined with a gesture recognizer that takes the hand's anatomy into account, the Power-up Button can recognize six different mid-air gestures performed on the side of a mobile device. This gives it, for instance, enough expressive power to provide full one-handed control of interface widgets displayed on screen. This technology can complement touch input, and can be particularly useful when interacting eyes-free. It also opens up a larger design space for widget organization on screen: the button enables a more compact layout of interface components than what touch input alone would allow. This can be useful when, e.g., filling the numerous fields of a long Web form, or for very small devices.

*Gimlenses* – Complex 3D virtual scenes such as CAD models of airplanes and representations of the human body are notoriously hard to visualize. Those models are made of many parts, pieces and layers of varying size, that partially occlude or even fully surround one another. Gimlenses [28] provides a multi-view, detail-in-context visualization technique that enables users to navigate complex 3D models by interactively drilling holes into their outer layers to reveal objects that are buried, possibly deep, into the scene (see Figure 10 ). These holes are constantly adjusted so as to guarantee the visibility of objects of interest from the parent view. Gimlenses can be cascaded and constrained with respect to one another, providing synchronized, complementary viewpoints on the scene. Gimlenses enable users to quickly identify elements of interest, get detailed views of those elements, relate them, and put them in a broader spatial context.

*Dashboard Exploration* – Visual stories help us communicate knowledge, share and interpret experiences and have become a focus in visualization research in recent years. In this paper we discuss the use of storytelling



Figure 9. (left) The SidePress prototype has two pressure sensors on one of its sides ; (right) The Power-Up button prototype is equipped with a button for sensing both pressure and proximity.

in Business Intelligence (BI) analysis [21] (Best Paper Award). We derive the actual practices in creating and sharing BI stories from in-depth interviews with expert BI analysts (both story “creators” and “readers”). These interviews revealed the need to extend current BI visual analysis applications to enable storytelling, as well as new requirements related to BI visual storytelling. Based on these requirements we designed and implemented a storytelling prototype tool with appropriate interaction techniques, that is integrated in an analysis tool used by our experts, and allows easy transition from analysis to story creation and sharing. We report experts’ recommendations and reactions to the use of the prototype to create stories, as well as novices’ reactions to reading these stories.

*Hybrid-Image Visualizations* – Data analysis scenarios often incorporate one or more displays with sufficiently large size and resolution to be comfortably viewed by different people from various distances. Hybrid-image visualizations [15] blend two different visual representations into a single static view, such that each representation can be perceived at a different viewing distance. They can thus be used to enhance overview tasks from a distance and detail-in-context tasks when standing close to the display. Viewers interact implicitly with these visualizations by walking around the space. By taking advantage of humans’ perceptual capabilities, hybrid-image visualizations show different content to viewers depending on their placement, without requiring tracking of viewers in front of a display. Moreover, because hybrid-images use a perception-based blending approach, visualizations intended for different distances can each utilize the entire display.

*Evolutionary Visual Exploration* – In a high-dimensionality context, the visual exploration of information is challenging, as viewers are often faced with a large space of alternative views on the data. We present [14], a system that combines visual analytics with stochastic optimization to aid the exploration of multidimensional datasets characterized by a large number of possible views or projections. Starting from dimensions whose values are automatically calculated by a PCA, an interactive evolutionary algorithm progressively builds (or evolves) non-trivial viewpoints in the form of linear and non-linear dimension combinations, to help users discover new interesting views and relationships in their data. The system calibrates a fitness function (optimized by the evolutionary algorithm) to take into account the user interactions to calculate and propose new views. Our method leverages automatic tools to detect interesting visual features and human interpretation to derive meaning, validate the findings and guide the exploration without having to grasp advanced statistical concepts. Our prototype was evaluated through an observational study with five domain experts, and helped



*Figure 10. Exploring the CAD drawing of a car engine. The three Gimlenses provide detailed views of different constituent parts of the engine, at different magnification levels and with varying orientation, while revealing their location inside the global 3D model. (a) Context view. (b) Magnified side view of a knot behind, and thus originally hidden by, the cylinder head cover. (c) View fully revealing a poppet valve in-context from a different angle than the main view, with (d) another Gimlens configured so as to provide a low-angled point of view on the valve.*

them quantify qualitative hypotheses, try out different scenarios to dynamically transform their data, and to better formulate their research questions and build new hypotheses for further investigation.

## 6.2. Research Methods

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

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.

However, we not only use research methods, we also investigate and develop them. As organizers of the CHI’13 conference in Paris, which had record-breaking numbers of submissions (over 2000) and participants (3500), we instituted a number of innovations in both the process of creating the program and presenting information to conference participants. In collaboration with researchers at MIT, we introduced an “author-sourcing” process (with an 87% participation rate) for collecting affinity data. We then developed a collaborative, interactive, visualization system on the WILD wall display, combined with the *Cobi* interactive constraint-solving system, that enabled us to resolve all presenter conflicts and successfully place all 500+ papers and events in relevant sessions ([35], [26]). We also replaced the “CHI Madness” series of 25-second presentations with “Video Previews”, in which each research paper, course, panel or other event has a 30-second video preview. These are now available on the CHI’13 website, the ACM/CHI YouTube channel and in the ACM Digital Library, before the paywall. We also developed and field-tested the Interactive Schedule on large, interactive displays, which allowed conference attendees to both view upcoming Video Previews and use their mobile phones to search for particular content and create customized playlists [29]. We also developed two interactive table-top applications that were presented at CHI’13 Interactivity, that allowed attendees to visualize and explore conference events as well as to create customized video playlists.

The RepliCHI workshop at CHI’13, co-organized by Wendy Mackay, examined issues with respect to encouraging replication of controlled experiments, and introduced the RepliCHI award to top research articles that offer strong empirical contributions that include replication. She also organized a session called *Interacting with CHI* in which participants explained the technologies and processes they developed to support the CHI conference design and execution.

In the context of our work with Interactive Paper to support music composition, we developed Paper Tonnetz, a paper-based interface to composing melodies and chords based on musical patterns expressed in Euler’s Tonnetz, and demonstrated it at CHI’13 Interactivity ([22]). We also explored how to create an interactive event for the “Fête de la Science”, called “Design Me a Sound Landscape”, in which participants can create their own ways of expressing a landscape and add natural sounds, such as wind, rain, moving water, that another participant can experience as they move on an interactive floor. Finally, we explored the drawing process, with the *Drawing Assistant* ([25]) in which users receive guidance and feedback as they learn to draw from photographs.

## 6.3. Engineering of interactive systems

**Participants:** Caroline Appert, Michel Beaudouin-Lafon [correspondant], Olivier Chapuis, Stéphane Huot, 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.

Our work has focused on developing middleware for the WILD platform, InSitu's experimental ultra-high-resolution interactive room for studying collaborative interaction and the visualization of very large datasets [2]. WILD features a wall-sized display with 32 monitors, a multitouch table, a motion-tracking system and various mobile devices. Running applications on WILD requires developing advanced distributed systems that coordinate, in real time, the 16 computers of the cluster driving the wall display with a variety of clients and servers running on other computers, including mobile devices.

We investigated the use of Web standards and protocols to develop and deploy such applications. Hydrascope [24] introduces the concept of *meta-application* that combines, adapts and/or repurposes existing web applications for an environment such as WILD. It uses a web browser (or even a web engine, e.g. WebKit) as a rendering and interaction toolkit and Web protocols (HTTP and WebSockets) for communication. We demonstrated how to control a wall-size presentation tool built on Google Present and a wall-size map built on Google Maps without modifying these applications but by taking advantage of the capability of web applications for introspection.

This approach was used to develop CHIWall, an application designed to help us schedule the CHI 2013 conference that InSitu chaired this year in Paris. The resulting tool combines a wall-size display of the full program with a constraint-detection and constraint-solving assistant called Cobi [26], which itself uses crowd-sourced information from the authors. The resulting application supports collaborative work to fine-tune the program (Figure 11) and features a flexible architecture that has been reused for other prototype applications.

In summary, InSitu has continued to make significant progress towards mature toolkits that support post-WIMP and distributed user interfaces. These toolkits, in turn, have enabled us to experiment with novel interaction techniques using rapid prototyping. Conversely, our work on novel interaction techniques has driven the development of software toolkits that embody their underlying principles, facilitating further exploration. This back-and-forth between techniques, methods and tools is a defining feature of InSitu, captured by the *Designing Interaction* [11] framework. As the focus of our research on interaction techniques has shifted from on-the-desktop to off-the-desktop, this approach has proven more valid than ever: improving interaction in such environments requires more complex software architectures and tools; in turn, these tools and architectures are a key step to getting these technologies outside the lab.





*Figure 11. A group of users interacting with the CHIwall application to fine tune the schedule of the CHI 2013 conference (500 events (400 research articles) in 200 sessions, in 16 parallel tracks over four days). The tablet interface provides additional detail about the content of a session, such as the abstract and the details of the affinity data crowd-sourced from the authors.*

## MANAO Team

# 5. New Results

## 5.1. Axis 1: Analysis and Simulation

### 5.1.1. Second Order Analysis of Variance in Multiple Importance Sampling

**Participants:** H. Lu, R. Pacanowski, X. Granier

Monte Carlo Techniques are widely used in Computer Graphics to generate realistic images. Multiple Importance Sampling reduces the impact of choosing a dedicated strategy by balancing the number of samples between different strategies. However, an automatic choice of the optimal balancing remains a difficult problem. Without any scene characteristics knowledge, the default choice is to select the same number of samples from different strategies and to use them with heuristic techniques (e.g., balance, power or maximum). We introduced [16] a second-order approximation of variance for balance heuristic. Based on this approximation, we automatically distribute samples for direct lighting without any prior knowledge of the scene characteristics. For all our test scenes (with different types of materials, light sources and visibility complexity), our method actually reduces variance in average (see Figure 9). This approach will help developing new balancing strategies.



Low to high glossy materials with five diffuse area light sources      Glossy materials with high-frequency environment map lighting

Figure 9. Our per-pixel second-order approximation of the variance leads to a new and automatic approach for balancing the number of samples between two different sampling strategies. Except for light sources, the inset images show the sample distribution for each pixel. The yellow corresponds to the default balance heuristic strategy [101]. Compared to the balance heuristic, the variance is reduced by **(Left)** 26% and **(Right)** 20% in average (14% and 11% for the standard deviation).

### 5.1.2. Rational BRDF

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

Over the last two decades, much effort has been devoted to accurately measuring Bidirectional Reflectance Distribution Functions (BRDFs) of real-world materials and to use efficiently the resulting data for rendering. Because of their large size, it is difficult to use directly measured BRDFs for real-time applications, and fitting the most sophisticated analytical BRDF models is still a complex task.

We have presented Rational BRDF [21], a general-purpose and efficient representation for arbitrary BRDFs, based on Rational Functions (RFs). Using an adapted parametrization, Rational BRDFs offer 1) a more compact and efficient representation using low-degree RFs, 2) an accurate fitting of measured materials with guaranteed control of the residual error, and 3) efficient importance sampling by applying the same fitting process to determine the inverse of the Cumulative Distribution Function (CDF) generated from the BRDF for use in Monte-Carlo rendering.

### 5.1.3. *Decomposing intensity gradients into information about shape and material*

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

Recent work has shown that the perception of 3D shapes, material properties and illumination are inter-dependent, although for practical reasons, each set of experiments has probed these three causal factors independently. Most of these studies share a common observation though: that variations in image intensity (both their magnitude and direction) play a central role in estimating the physical properties of objects and illumination. Our aim is to separate retinal image intensity gradients into contributions of different shape and material properties, through a theoretical analysis of image formation [11].

We find that gradients can be understood as the sum of three terms: variations of surface depth conveyed through surface-varying reflectance and near-field illumination effects (shadows and inter-reflections); variations of surface orientation conveyed through reflections and far-field lighting effects; and variations of surface micro-structures conveyed through anisotropic reflections. We believe our image gradient decomposition constitutes a solid and novel basis for perceptual inquiry. We first illustrate each of these terms with synthetic 3D scenes rendered with global illumination. We then show that it is possible to mimic the visual appearance of shading and reflections directly in the image, by distorting patterns in 2D. Finally, we discuss the consistency of our mathematical relations with observations drawn by recent perceptual experiments, including the perception of shape from specular reflections and texture. In particular, we show that the analysis can correctly predict certain specific illusions of both shape and material.

## 5.2. Axis 2: From Acquisition to Display

### 5.2.1. *Interactive Spatial Augmented Reality*

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

We propose the revealing flashlight [26], a new interaction and visualization technique in spatial augmented reality that helps to reveal the details of cultural heritage artifacts (see Figure 3), since they often contain details that are difficult to distinguish due to aging effects such as erosion. We locally and interactively augment a physical artifact by projecting an expressive 3D visualization that highlights its features, based on an analysis of its previously acquired geometry at multiple scales.

Our novel interaction technique simulates and improves the behavior of a flashlight: according to 6-degree-of-freedom input, we adjust the numerous parameters involved in the expressive visualization - in addition to specifying the location to be augmented. This makes advanced 3D analysis accessible to the greater public with an everyday gesture, by naturally combining the inspection of the real object and the virtual object in a co-located interaction and visualization space. The revealing flashlight can be used by archeologists, for example, to help decipher inscriptions in eroded stones, or by museums to let visitors interactively discover the geometric details and meta-information of cultural artifacts. We confirm its effectiveness, ease-of-use and ease-of-learning in an initial preliminary user study and by the feedbacks of two public exhibitions.

### 5.2.2. *High Dynamic Range, Multispectral, Polarization, and Light-Field Imaging*

**Participants:** A. Manakov, J. Restrepo, R. Hegedüs, I. Ihrke



*Figure 10. (Top) Novel optical converter module that can be placed between a camera and its lens. This module can be configured flexibly to allow for multi-spectral, polarization, high-dynamic range, or light field imaging. It works by splitting the original image into a number of copies that can be optical filtered separately (Bottom, Left). Computational post-processing allows for unprecedented flexibility in image post-processing such as post-capture control of illumination, polarization state, exposure setting or focus. (Bottom,Right) The basis of the operational principle of the aforementioned prototype is imaging in mirror systems.*

In [5] we propose a non-permanent add-on that enables plenoptic imaging with standard cameras (see also Figure 10 top and left). Our design is based on a physical copying mechanism that multiplies a sensor image into a number of identical copies that still carry the plenoptic information of interest. Via different optical filters, we can then recover the desired information. A minor modification of the design also allows for aperture sub-sampling and, hence, light-field imaging. As the filters in our design are exchangeable, a reconfiguration for different imaging purposes is possible. We show in a prototype setup that high dynamic range, multispectral, polarization, and light-field imaging can be achieved with our design.

### 5.2.3. *Structure of a Planar Mirror System from Multiple Observations of a Single Point*

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

We have investigated the problem of identifying the position of a viewer inside a room of planar mirrors with unknown geometry in conjunction with the room's shape parameters [25] (see also Figure 10 bottom right). We consider the observations to consist of angularly resolved depth measurements of a single scene point that is being observed via many multi-bounce interactions with the specular room geometry. Applications of this problem statement include areas such as calibration, acoustic echo cancelation and time-of-flight imaging. We theoretically analyze the problem and derive sufficient conditions for a combination of convex room geometry, observer, and scene point to be reconstructable. The resulting constructive algorithm is exponential in nature and, therefore, not directly applicable to practical scenarios.

To counter the situation, we propose theoretically devised geometric constraints that enable an efficient pruning of the solution space and develop a heuristic randomized search algorithm that uses these constraints to obtain an effective solution. We demonstrate the effectiveness of our algorithm on extensive simulations as well as in a challenging real-world calibration scenario.

### 5.2.4. *Mirrors in Computer Graphics, Computer Vision and Time-of-Flight Imaging*

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

We have investigated the state of the art in dealing with the geometry of mirror systems [28].

Mirroring is one of the fundamental light/surface interactions occurring in the real world. Surfaces often cause specular reflection, making it necessary to design robust geometry recovery algorithms for many practical situations. In these applications the specular nature of the surface is a challenge. On the other side, mirrors, with their unique reflective properties, can be used to improve our sensing modalities, enabling applications such as surround, stereo and light field imaging. In these scenarios the specular interactions are highly desirable. Both of these aspects, the utilization and circumvention of mirrors are present in a significant amount of publications in different scientific areas. These publications are covering a large number of different problem statements as well as many different approaches to solutions. In this survey we focused on a collection and classification of the work in this area.

### 5.2.5. *Computational Fabrication and Display of Material Appearance*

**Participant:** I. Ihrke

We have investigated the state of the art in digital material fabrication and active display technology [22].

After decades of research on digital representations of material and object appearance, computer graphics has more recently turned to the problem of creating physical artifacts with controllable appearance characteristics.

While this work has mostly progressed in two parallel streams – display technologies as well as novel fabrication processes – we believe there is a large overlap and the potential for synergies between these two approaches. In this report, we summarize research efforts from the worlds of fabrication display, and categorize the different approaches into a common taxonomy. We believe that this report can serve as a basis for systematic exploration of the design space in future research.

## 5.3. Axis 3: Rendering, Visualization and Illustration

### 5.3.1. *Real-Time Sampling from Captured Environment Map*

**Participants:** H. Lu, R. Pacanowski, X. Granier



Figure 11. Time-varying light samples distribution for one pixel (cyan dot) on the dragon model when lit with a dynamic environment map [95]. This example runs in average at 145 fps using Multiple Importance Sampling with 50 samples for the Lafortune energy conserving Phong BRDF with a shininess exponent set to 150.

We have introduced [23] a simple and effective technique for light-based importance sampling of dynamic environment maps based on the formalism of Multiple Importance Sampling (MIS). The core idea is to balance per pixel the number of samples selected on each cube map face according to a quick and conservative evaluation of the lighting contribution: this increases the number of effective samples. In order to be suitable for dynamically generated or captured HDR environment maps, everything is computed on-line for each frame without any global preprocessing. Our MIS formalism can be easily extended to other strategies such as BRDF importance sampling.

### 5.3.2. Screen-Space Curvature for Production-Quality Rendering and Compositing

**Participants:** N. Mellado, P. Barla, G. Guennebaud, P. Reuter

Curvature is commonly employed for enhancing details in textured 3D models, or to modulate shading at the rendering or compositing stage. However, existing methods that compute curvature in object space rely on mesh-based surfaces and work at the vertex level. Consequently, they are not well adapted to production-quality models that rely on either subdivision surfaces with displacement and bump maps, or on implicit and procedural representations. In practice they would require a view-dependent scene discretization at each frame, to adapt geometry to visible details and avoid aliasing artifacts. Our approach [24] is independent of both scene complexity and the choice of surface representations since it computes mean curvature from scratch at each frame in screen-space. It works without any pre-process and provides a controllable screen-space scale parameter, which makes it ideal for production requirements, either during rendering or compositing.

### 5.3.3. Smooth Surface Contours with Accurate Topology

**Participant:** P. Bénard

Computing the visible contours of a smooth 3D surface is a surprisingly difficult problem, and previous methods are prone to topological errors, such as gaps in the outline. Our approach [13] 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. We demonstrate results computing contours of complex 3D objects, on which our method eliminates the contour artifacts of other methods.

## 5.4. Axis 4: Editing and Modeling

### 5.4.1. Implicit Skinning and Modeling

**Participant:** G. Guennebaud

Geometric skinning techniques, such as smooth blending or dual-quaternions, are very popular in the industry for their high performances, but fail to mimic realistic deformations. Other methods make use of physical simulation or control volume to better capture the skin behavior, yet they cannot deliver real-time feedback. In collaboration with IRIT (Toulouse) and the Imagine team (Grenoble), we developed the first purely geometric method handling skin contact effects and muscular bulges in real-time. Our insight is to exploit the advanced composition mechanism of volumetric, implicit representations for correcting the results of geometric skinning techniques (cf. Figure 12 -a). The mesh is first approximated by a set of implicit surfaces. At each animation step, these surfaces are combined in real-time and used to adjust the position of mesh vertices, starting from their smooth skinning position. This deformation step is done without any loss of detail and seamlessly handles contacts between skin parts. As it acts as a post-process, our method fits well into the standard animation pipeline. Moreover, it requires no intensive computation step such as collision detection, and therefore provides real-time performances. This work has been published at Siggraph this year [20] and featured by the 3DVF website <http://www.3dvv.com/actualite-6678-siggraph-2013-methode-skinning-implicite.html>.

Still in collaboration with IRIT, we addressed the challenging problem of finding adequate bounds for implicit modeling with compact field functions. Recent advances in implicit surface modeling now provide highly controllable blending effects. These effects rely on the field functions of  $\mathbb{R}^3 \rightarrow \mathbb{R}$  in which the implicit surfaces are defined. In these fields, there is an outside part in which blending is defined and an inside part. The implicit surface is the interface between these two parts. As recent operators often focus on blending, most efforts have been made on the outer part of field functions and little attention has been paid on the inner part. Yet, the inner fields are important as soon as difference and intersection operators are used. This makes its quality as crucial as the quality of the outside.

In this work we analyzed these shortcomings, and deduced new constraints on field functions such that differences and intersections can be seamlessly applied without introducing discontinuities or field distortions. In particular, we showed how to adapt state of the art gradient-based union and blending operators to our new constraints. Our approach enables a precise control of the shape of both the inner or outer field boundaries. We also developed a new set of asymmetric operators tailored for the modeling of fine details while preserving the integrity of the resulting fields. This work has been published at Shape Modeling International 2013 [14].

### 5.4.2. Surface reconstruction

**Participants:** J. Chen, G. Guennebaud, P. Barla, X. Granier

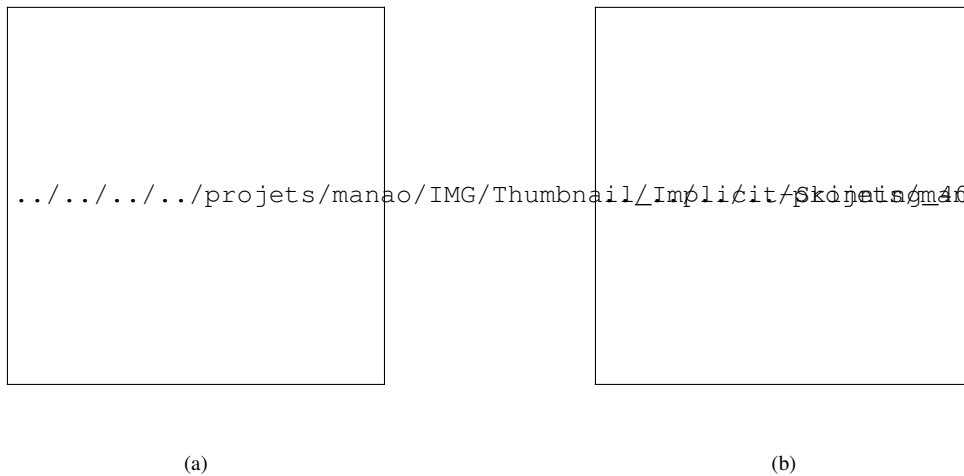


Figure 12. (a) Illustration of the implicit skinning technique. (b) Surface reconstruction from non-oriented normals

Reconstructing a smooth surface from a set of points is still a challenging problem. Most of the popular techniques assume correctly oriented points as inputs. However, in many situations, computing a consistent orientation of the normal field is as difficult as the reconstruction itself. In a recent work, we extended the Algebraic Point Set Surface method to support non oriented normals (cf. Figure 12 -b). By fitting algebraic spheres, our approach outperforms simple local methods based on non-oriented planar fit while still being fast since it involves only local computations. The core of this new technique also proved to be useful for image processing. This work as been published at Computer Graphics Forum [3].

#### 5.4.3. Manipulation of Anisotropic Highlights

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

We have developed [19] 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

# 6. New Results

## 6.1. Visual perception

### 6.1.1. *Decomposing intensity gradients into information about shape and material*

**Participants:** Pascal Barla, Romain Vergne, Roland W. Fleming.

Recent work has shown that the perception of 3D shapes, material properties and illumination are inter-dependent, although for practical reasons, each set of experiments has probed these three causal factors independently. Most of these studies share a common observation though: that variations in image intensity (both their magnitude and direction) play a central role in estimating the physical properties of objects and illumination. Our aim is to separate retinal image intensity gradients into contributions of different shape and material properties, through a theoretical analysis of image formation. We find that gradients can be understood as the sum of three terms: variations of surface depth conveyed through surface-varying reflectance and near-field illumination effects (shadows and inter-reflections); variations of surface orientation conveyed through reflections and far-field lighting effects; and variations of surface micro-structures conveyed through anisotropic reflections. We believe our image gradient decomposition constitutes a solid and novel basis for perceptual inquiry. We first illustrate each of these terms with synthetic 3D scenes rendered with global illumination. We then show that it is possible to mimic the visual appearance of shading and reflections directly in the image, by distorting patterns in 2D. Finally, we discuss the consistency of our mathematical relations with observations drawn by recent perceptual experiments, including the perception of shape from specular reflections and texture. In particular, we show that the analysis can correctly predict certain specific illusions of both shape and material.

### 6.1.2. *Predicting the effects of illumination in shape from shading*

**Participants:** Roland W. Fleming, Romain Vergne, Steven Zucker.

Shading depends on different interactions between surface geometry and lighting. Under collimated illumination, shading is dominated by the 'direct' term, in which image intensities vary with the angle between surface normals and light sources. Diffuse illumination, by contrast, is dominated by 'vignetting effects' in which image intensities vary with the degree of self-occlusion (the proportion of incoming direction that each surface point 'sees'). These two types of shading thus lead to very different intensity patterns, which raises the question of whether shading inferences are based directly on image intensities. We show here that the visual system uses 2D orientation signals ('orientation fields') to estimate shape, rather than raw image intensities and an estimate of the illuminant. We rendered objects under varying illumination directions designed to maximize the effects of illumination on the image. We then passed these images through monotonic, non-linear intensity transfer functions to decouple luminance information from orientation information, thereby placing the two signals in conflict (Figure 6). In Task 1 subjects adjusted the 3D shape of match objects to report the illusory effects of changes of illumination direction on perceived shape. In Task 2 subjects reported which of a pair of points on the surface appeared nearer in depth. They also reported perceived illumination directions for all stimuli. We find that the substantial misperceptions of shape are well predicted by orientation fields, and poorly predicted by luminance-based shape from shading. For the untransformed images illumination could be estimated accurately, but not for the transformed images. Thus shape perception was, for these examples, independent of the ability to estimate the lighting. Together these findings support neurophysiological estimates of shape from the responses of orientation selective cell populations, irrespective of the illumination conditions.

### 6.1.3. *Evaluation of Depth of Field for Depth Perception in DVR*

**Participants:** Pascal Grosset, Charles Hansen, Georges-Pierre Bonneau.



*Figure 6. Top: one single shape is shaded using multiple light source directions. Bottom: one single light source is used to shade multiple shapes. Each of them has approximatively the same percept as for the corresponding light direction. Similarity between orientation fields can be seen in the insets.*

We study the use of Depth of Field for depth perception in Direct Volume Rendering (Figure 7 ). Direct Volume Rendering with Phong shading and perspective projection is used as the baseline. Depth of Field is then added to see its impact on the correct perception of ordinal depth. Accuracy and response time are used as the metrics to evaluate the usefulness of Depth of Field. The on site user study has two parts: static and dynamic. Eye tracking is used to monitor the gaze of the subjects. From our results we see that though Depth of Field does not act as a proper depth cue in all conditions, it can be used to reinforce the perception of which feature is in front of the other. The best results (high accuracy & fast response time) for correct perception of ordinal depth is when the front feature (out of the users were to choose from) is in focus and perspective projection is used. Our work has been published in the proceedings of the Pacific Graphics conference in 2013 [16].



Figure 7. A user study is conducted to evaluate the use of Depth of Field for depth perception in Direct Volume Rendering. The user has to say which of the two highlighted features is in the front. Statistical analysis of the results is performed to assess the effect of Depth of Field in the visualization. Results are detailed in the Pacific Graphics paper [16].

## 6.2. Visualization

### 6.2.1. Morse-Smale complexes

**Participants:** Léo Allemand-Giorgis, Georges-Pierre Bonneau, Stefanie Hahmann.

Preserving meaningful local extrema of scalar data in a visualization while removing nearby extrema with similar values is a powerful way for enhancing the appearance of significant features. For the special case of monotonic data, e.g. data with no local extrema in the interior of the domain, the visualization should not introduce spurious local extrema. We study a new piecewise polynomial interpolant that preserves the monotonicity of scalar data defined on a 2D uniform grid. Based on this interpolant, we also plan to

introduce a new method for visualizing data that has been simplified according to its Morse-Smale complex, a combinatorial structure connecting the critical points and partitioning the domain into a set of monotonic regions. In contrast with previous analogous works, our approach uses piecewise polynomial functions defined in each monotonic region instead of optimizing values on the original mesh vertices. We have presented our first results in a workshop and have submitted a paper for a book chapter about our new monotonic interpolant.

### 6.2.2. *Computation of components' interfaces in highly complex assemblies*

**Participants:** François Jourdes, Georges-Pierre Bonneau, Stefanie Hahmann, François Faure.

The preparation of CAD models from complex assemblies for simulation purposes is a very time-consuming and tedious process, since many tasks such as meshing and idealization are still completed manually. Herein, the detection and extraction of geometric interfaces between components of the assembly is of central importance not only for the simulation objectives but also for all necessary shape transformations such as idealizations or detail removals. It is a repetitive task in particular when complex assemblies have to be dealt with. This paper proposes a method to rapidly and fully automatically generate a precise geometric description of interfaces in generic B-Rep CAD models. The approach combines an efficient GPU ray-casting technique commonly used in computer graphics with a graph-based curve extraction algorithm. Not only is it able to detect a large number of interfaces efficiently, but it also provides an accurate Nurbs geometry of the interfaces, that can be stored in a plain STEP file for further downstream treatment. We demonstrate our approach on examples from aeronautics and automotive industry, see Figure 8. Our results have been funded in by the ANR Project ROMMA. They have been published as a journal paper in [13], and presented at the Solid and Physical Modeling conference in 2013.

## 6.3. Image creation and editing

### 6.3.1. *Accurate Binary Image Selection from Inaccurate User Input*

**Participants:** Subr Kartic, Paris Sylvain, Soler Cyril, Kautz Jan.

Selections are central to image editing, since they are the starting point of common operations such as copy-pasting and local edits. Creating them by hand is particularly tedious and scribble-based techniques have been introduced to assist the process. By interpolating a few strokes specified by users, these methods generate precise selections. However, most of the algorithms assume a 100% accurate input, and even small inaccuracies in the scribbles often degrade the selection quality, which imposes an additional burden on users. In this work, we propose a selection technique tolerant to input inaccuracies (See example on Figure 9). We use a dense conditional random field (CRF) to robustly infer a selection from possibly inaccurate input. Further, we show that patch-based pixel similarity functions yield more precise selection than simple point-wise metrics. However, efficiently solving a dense CRF is only possible in low-dimensional Euclidean spaces, and the metrics that we use are high-dimensional and often non-Euclidean. We address this challenge by embedding pixels in a low-dimensional Euclidean space with a metric that approximates the desired similarity function. The results show that our approach performs better than previous techniques and that two options are sufficient to cover a variety of images depending on whether the objects are textured. This work has been published to the Eurographics Conference [15].

### 6.3.2. *Discrete Texture Design Using a Programmable Approach*

**Participants:** Hugo Loi, Thomas Hurtut, Romain Vergne, Thollot Joëlle.

Many rendering methods use discrete textures (planar arrangements of vector elements) instead of classic bitmaps. Discrete textures are resolution-insensitive and easily allow to modify the elements' geometry or spatial distribution. However, manually drawing such textures is a time-consuming task. Automating this production is a long-time studied subject. The methods designed for this purpose deal with a difficult tradeoff between the reachable variety of textures and the usability for a community of users. In this work, we show that considering discrete textures as programs allow for a larger variety of textures than relying on a given model. This work has been as a Siggraph 2013 talk [18].

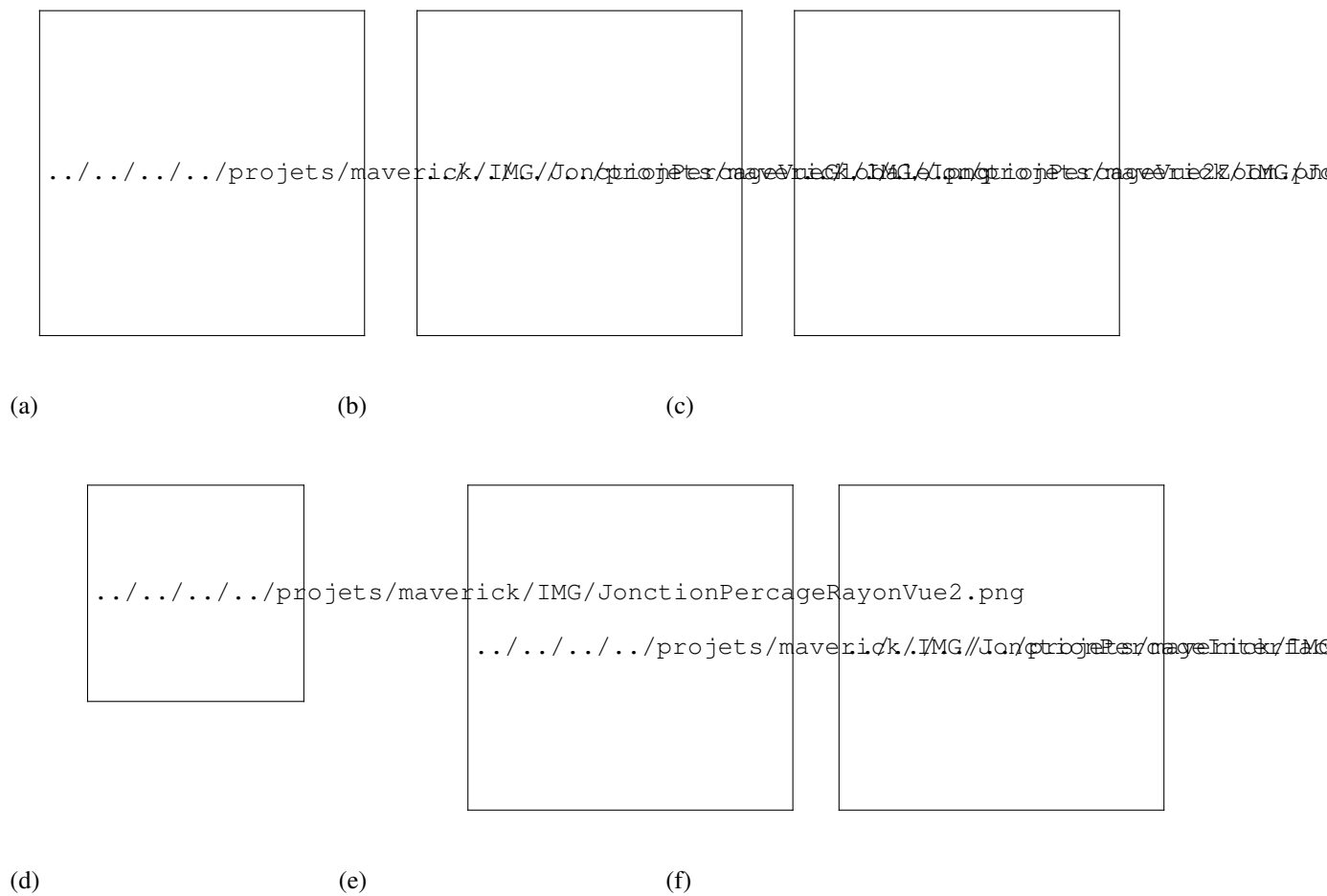


Figure 8. Computing Interfaces of an Aircraft part assembling the wings with the body of an aircraft (model courtesy of EADS). (a,b) two views of the components, (c) exploded view, (d) ray casting for detecting proximities between parts, (e) boundary reconstruction, (f) final interfaces



*Figure 9. Accurate selection from inaccurate input. Although our algorithm is fed with inaccurate selection, it succeeds in separating the background from the foreground components in the image.*



Figure 10. Discrete element texture generated with our programmable approach.

## 6.4. Complex scenes

### 6.4.1. Appearance pre-filtering

**Participants:** Heitz Eric, Neyret Fabrice.

Last year work and HPG'12 paper "Representing Appearance and Pre-filtering Subpixel Data in Sparse Voxel Octrees" was dealing with the light and view dependant aspect of complex surfaces due to sub-pixels details. This was done by replacing sub-voxel height fields by gaussian slope distribution and height-correlated colors by its gradient, feeding a Cook-Torrance-like microfacet brdf.

In continuation of this and in the same spirit of replacing sub-pixel values by gaussian distributions to be shaded using the frame of microfacets brdf, this year we addressed the filtering of color maps (on surfaces and per se), displacement map, and reflectance maps, thus obtaining a complete model of the local rendering integral (see Figure 11).

Note that Eric did his work partly during his 6-monthes stay of University of Montreal in the scope of Exploradoc regional founding. He also colaborated with nVIDIA for an on-going work related to animation of GigaVoxels, and we were invited for a stay of several weeks at Weta Digital, NZ to help them applying our techniques.

### 6.4.2. Filtering Color Mapped Textures and Surfaces

**Participants:** Heitz Eric, Neyret Fabrice, Nowrouzezahrai Derek, Poulin Pierre.

Indeed, several ubiquitous CG operations like filtering non-linear functions of the data are still mostly unsolved despite their important flaws. Typically, density, noise data, normals or height are filtered before feeding a color look-up texture, despite the strong non-linearity of the transform forbids factoring it out of the integral. This result on very visible flaws such as thin blue bones+air foams appearing as red muscle at distance in volume visualisation, silhouettes and horizon getting the middle tint instead of the integral of tints, procedural noise bump maps and height fields appearing as smooth instead of rough.



Figure 11. The resulting pixel color is the integral of the local rendering on the surface, which combines 4 fields which are: the color texture parameter, the visibility from eye, the visibility from light, and the shading. As these are varying and non-linear, the ubiquitous simplification of averaging each term separately is not valid.

Assuming Gaussian distribution of colors within a pixel or voxel, the filtered colors values can be represented as color lobes (i.e. histogramms) instead of scalars. In all the cases where the subpixel/voxel raw data can also be represented as gaussian distribution (e.g. Perlin noise), the filtering is just the inner product of the two lobes. It can easily be tabulated as a 1D LUT MIP-map which LOD corresponds to the standard deviation and thus the scale. Since microfacet brdf models allow to estimate the *visible* slope statistics accounting for light and view visibility, this allows for emerging light-view dependant color effect both accurately and very efficiently. Note that the same scheme applies for colors correlated to orientation rather than heights (see Figure 12). This provides a multiscale representation where subpixel/subvoxel data is represented through lobes which can be precalculated or calculated on demand from the thinner level.

This work was published at ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D) 2013 [17]. An extended version "Filtering Non-Linear Transfer Functions on Surfaces" was published at IEEE Transactions on Visualization and Computer Graphics 2013 [11].

#### 6.4.3. Linear Efficient Antialiased Displacement and Reflectance Mapping

**Participants:** Dupuy Jonathan, Heitz Eric, Neyret Fabrice.

Here, the last term of the local rendering integration is addressed: the filtering of subpixel/subvoxel geometry and brdf as an apparent brdf applied on a macro-geometry. By re-derivating accurately the brdf of a displacement map assumed to have sub-pixel gaussian distribution (with an exact masking term, more accurate cross-correlated light-view, and offsetted apparent lobe), and by noting that the reflectance of the environment can be pre-filtered like the textures of the previous paper, we finally obtain a complete model of pre-filtered appearance of surfaces (see Figure 13). This work, co-first-authored with Jonathan Dupuy, was published at ACM Transactions on Graphics and presented at Siggraph Asia [8].

## 6.5. Realistic rendering

### 6.5.1. Interactive Rendering of Acquired Materials on Dynamic Geometry Using Frequency Analysis

**Participants:** Bagher M. Mahdi, Soler Cyril, Subr Kartic, Belcour Laurent, Holzschuch Nicolas.





Figure 12. (Left:) Correct pre-filtering of non-linear functions (e.g., LUT) of a noise texture. (Right:) Correct pre-filtering of the appearance of details, with light and view macroscopic dependency and color correlated with depth.

Shading acquired materials with high-frequency illumination is computationally expensive. Estimating the shading integral requires multiple samples of the incident illumination. The number of samples required may vary across the image, and the image itself may have high- and low-frequency variations, depending on a combination of several factors. Adaptively distributing computational budget across the pixels for shading is a challenging problem. In this work, we depict complex materials such as acquired reflectances, interactively, without any precomputation based on geometry. In each frame, we first estimate the frequencies in the local light field arriving at each pixel, as well as the variance of the shading integrand. Our frequency analysis [1] accounts for combinations of a variety of factors: the reflectance of the object projecting to the pixel, the nature of the illumination, the local geometry and the camera position relative to the geometry and lighting. We then exploit this frequency information (bandwidth and variance) to adaptively sample for reconstruction and integration. For example, fewer pixels per unit area are shaded for pixels projecting onto diffuse objects, and fewer samples are used for integrating illumination incident on specular objects (See Figure 14 ). This work has been published in IEEE Transactions on Visualization and Computer Graphics [3], as a follow up of a previous paper published at the I3D conference.

### 6.5.2. 5D Covariance Tracing for Efficient Defocus and Motion Blur

**Participants:** Belcour Laurent, Soler Cyril, Subr Kartic, Holzschuch Nicolas, Durand Frédo.

The rendering of effects such as motion blur and depth-of-field requires costly 5D integrals. We dramatically accelerate their computation through adaptive sampling and reconstruction based on the prediction of the anisotropy and bandwidth of the integrand. For this, we develop a new frequency analysis of the 5D temporal light-field, and show that first-order motion can be handled through simple changes of coordinates in 5D. We further introduce a compact representation of the spectrum using the co-variance matrix and Gaussian approximations. We derive update equations for the  $5 \times 5$  covariance matrices for each atomic light transport event, such as transport, occlusion, BRDF, texture, lens, and motion. The focus on atomic operations makes our work general, and removes the need for special-case formulas. We present a new rendering algorithm that computes 5D covariance matrices on the image plane by tracing paths through the scene, focusing on the single-bounce case. This allows us to reduce sampling rates when appropriate and perform reconstruction of images with complex depth-of-field and motion blur effects (See Figure 15 ). This work was published at ACM Transactions on Graphics [5] and presented at Siggraph'2013.



*Figure 13. (Left:) Correct pre-filtering of non-linear functions (e.g., LUT) of a noise texture. (Right:) Correct pre-filtering of the appearance of details, with light and view macroscopic dependency and color correlated with depth.*



*Figure 14. Our simplified bandwidth prediction technique is suitable for real-time rendering. It allows us to only compute a subset of the image pixels while concentrating integration cost to the pixels that have the largest variance of the integrand.*



*Figure 15. Our covariance analysis of the power spectrum of local illumination allows us to accurately predict the sampling rates an reconstruction filters to significantly increase the convergence of path tracing.*

### 6.5.3. Accurate and Efficient Filtering using Anisotropic Filter Decomposition

**Participants:** Soler Cyril, Bagher Mahdi, Nowrouzezahrai Derek.

Efficient filtering remains an important challenge in computer graphics, particularly when filters are spatially-varying, have large extent, and/or exhibit complex anisotropic profiles. We explored an efficient filtering approach for these difficult cases based on anisotropic filter decomposition (IFD). By decomposing complex filters into linear combinations of simpler, displaced isotropic kernels, and precomputing a compact prefiltered dataset, we are able to interactively apply any number of—potentially transformed—filters to a signal (See Figure 16 ). Our performance scales linearly with the size of the decomposition, not the size nor the dimensionality of the filter, and our prefiltered data requires reasonable storage, comparing favorably to the state-of-the-art. We apply IFD to interesting problems in image processing and realistic rendering. This work is currently under submission and a technical report is already available [21].



Figure 16. Isotropic filter decomposition allows us to efficiently perform anisotropic filtering at the cost of minimal pre-computation.

#### 6.5.4. Double- and Multiple-scattering Effects in Translucent Materials

**Participants:** Holzschuch Nicolas, Gascuel Jean-Dominique.

Some materials, such as coffee, milk or marble, have a soft translucent aspect because of sub-surface scattering: light enters them, is scattered several times inside before leaving in a different place. A full representation of sub-surface scattering effects in illumination simulation is computationally expensive. The main difficulty comes from multiple scattering events: the high number of events increases the uncertainty on the result, forcing us to allocate more time for the computations. Recently, we showed that there is a strong correlation between the surface effects of multiple scattering inside the material and the effects after just two scatter events. This knowledge will help in accelerating multiple scattering effects (see figure 17). We exploited this knowledge to provide a model and implementation for fast computation of double-scattering events, using a precomputed density function stored in a compact way. This work has been published in IEEE Computer Graphics and Applications [12].

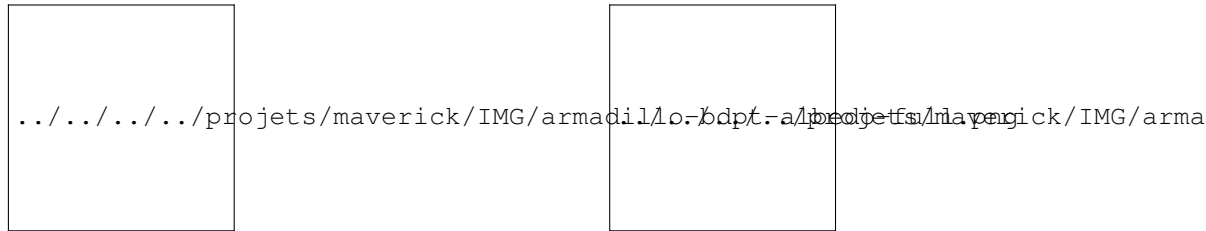


Figure 17. Our approximation achieves results very close to the reference solution: (left) the BDPT reference, in 32min; (right) our solution, in 1.7 min. .

## 6.6. Inverse problems

### 6.6.1. Sparse BRDF Approximation using Compressive Sensing

**Participants:** Zupancic Benoit, Soler Cyril.

BRDF acquisition is a tedious operation, since it requires measuring 4D data. On one side of the spectrum lie explicit methods, which perform many measurements to potentially produce very accurate reflectance data after interpolation. These methods are generic but practically difficult to setup and produce high volume data. On the other side, acquisition methods based on parametric models implicitly reduce the infinite dimensionality of the BRDF space to the number of parameters, allowing acquisition with few samples. However, parametric methods require non linear optimization. They become unstable when the number of parameters is large, with no guaranty that a given parametric model can ever fit particular measurements.

We experiment a new acquisition method where the measurement of the BRDF is performed from a single image, knowing the normals and illumination. To tackle such a severely underconstrained problem, we express the BRDF in a high dimensional basis, and perform the reconstruction using *compressive sensing*, looking for the most sparse solution to the linear problem of fitting the measurement image. Doing so, we leverage the coherency between the measured pixels, while keeping the high dimension of the space the BRDF is searched into.

This work is a very first attempt at reconstructing BRDFs using compressive sensing. In Fig 18 we used a synthetic input image, for the sake of checking the feasibility of the recovery algorithm, in the particular case of an isotropic spatially constant BRDFs. The possibility to extend our theory to non spatially varying and anisotropic BRDFs is currently under investigation. We would like to orient our work toward BRDF acquisition with consumer hardware. In particular, our preliminary results indicate that compressive sensing

could achieve an very accurate acquisition with additional input, such as a video of a static object under probed lighting.

This work has been published as a poster to the Siggraph Asia'2013 conference [24].



*Figure 18. Reconstruction of a BRDF from a single image. This work stands up as a preliminary experiment to prove that BRDF acquisition can be performed with a much lower amount of data than what the Nyquist rate predicts for data reconstruction, if using compressive sensing methods.*

### **6.6.2. Floating tangents for approximating spatial curves with G1 piecewise helices**

**Participants:** Derouet-Jourdan Alexandre, Bertails-Descoubes Florence, Thollot Joëlle.

Curves are widely used in computer science to describe real-life objects such as slender deformable structures. Using only 3 parameters per element, piecewise helices offer an interesting and compact way of representing digital curves. In our work [7], we present a robust and fast algorithm to approximate Bezier curves with G1 piecewise helices. Our approximation algorithm takes a Bezier spline as input along with an integer N

and returns a piecewise helix with  $N$  elements that closely approximates the input curve. The key idea of our method is to take  $N+1$  evenly distributed points along the curve, together with their tangents, and interpolate these tangents with helices by slightly relaxing the points. Building on previous work, we generalize the proof for Ghosh's co-helicity condition, which serves us to guarantee the correctness of our algorithm in the general case. Finally, we demonstrate both the efficiency and robustness of our method by successfully applying it to various datasets of increasing complexity, ranging from synthetic curves created by an artist to automatic image-based reconstructions of real data such as hair, heart muscular fibers or magnetic field lines of a star.

### **6.6.3. Inverse Dynamic Hair Modeling with Frictional Contact**

**Participants:** Derouet-Jourdan Alexandre, Daviet Gilles, Bertails-Descoubes Florence, Thollot Joëlle.

In the latest years, considerable progress has been achieved for accurately acquiring the geometry of human hair, thus largely improving the realism of virtual characters. In parallel, rich physics-based simulators have been successfully designed to capture the intricate dynamics of hair due to contact and friction. However, at the moment there exists no consistent pipeline for converting a given hair geometry into a realistic physics-based hair model. Current approaches simply initialize the hair simulator with the input geometry in the absence of external forces. This results in an undesired sagging effect when the dynamic simulation is started, which basically ruins all the efforts put into the accurate design and/or capture of the input hairstyle. In this work [6] we propose the first method which consistently and robustly accounts for surrounding forces — gravity and frictional contacts, including hair self-contacts — when converting a geometric hairstyle into a physics-based hair model. Taking an arbitrary hair geometry as input together with a corresponding body mesh, we interpret the hair shape as a static equilibrium configuration of a hair simulator, in the presence of gravity as well as hair-body and hair-hair frictional contacts. Assuming that hair parameters are homogeneous and lie in a plausible range of physical values, we show that this large underdetermined inverse problem can be formulated as a well-posed constrained optimization problem, which can be solved robustly and efficiently by leveraging the frictional contact solver of the direct hair simulator. Our method was successfully applied to the animation of various hair geometries, ranging from synthetic hairstyles manually designed by an artist to the most recent human hair data automatically reconstructed from capture.



## MIMETIC Project-Team

# 6. New Results

## 6.1. Biomechanics and Motion Analysis

### 6.1.1. Modeling gesture in sports: tennis serve

**Participants:** Nicolas Bideau, Guillaume Nicolas, Benoit Bideau, Richard Kulpa.

In the midst of the INSEP project and the PhD of Caroline Martin, the tennis serve has been studied with biomechanical analyses. To this end, we have done kinematic and dynamic analyses based on motion capture, force plate and electromyographic systems. They provided information on how the gesture is performed and how it is related to injuries. Moreover, these analyses have been done on several level of players including top-level ones. A comparison of the kinematic and dynamic data can then be done. Our objective is to use these data in virtual reality to study the interaction between a tennis server and a receiver. We are creating a tool that displays a virtual server in front of a real receiver. The control of the virtual server is then done based on these biomechanical data. The objective is to analyze the reaction of the receiver depending on the movement of the server and its level of expertise.

### 6.1.2. Motion modeling in clinical applications

**Participant:** Armel Crétual.

We have developed a new index of gait quantification based on muscular activity called KeR-EGI. After having proved that this index is consistent and complementary with kinematics-based indices, we have shown that it is reproducible in patients even when their impairment level is high. This index is now used in clinical routine in adults. It will be also used in pediatrics in the next few months.

In orthopedics, we have proposed a novel method to quantify shoulder's global mobility called SCSV. It is based on the reachable volume in the whole configuration space of the shoulder, i.e. a 3-dimensional angular space. Clinical evaluations of shoulder's range of motion are quite always based on the analysis of only one axis, and the most usual refers to maximal external rotation from rest posture (ER1). Considering several mono-axial amplitudes, we have shown that ER1 is actually the worst choice to estimate global mobility. Instead of the ER1 procedure, we proposed to use the sum of 3 mono-axial amplitudes: external/internal amplitude at 90° lateral elevation, abduction and flexion/extension.

As shoulder is actually a complex of three articulations (gleno-humeral, scapulo-thoracic and sternocalvicular), we have evaluated the contribution of each of them on global mobility. This has been done through a cadaveric study where we measured SCSV in any possible blocking conditions of these three articulations (from 0 to 3).

## 6.2. VR and Ergonomics

**Participants:** Charles Pontonnier [contact], Georges Dumont, Franck Multon, Pierre Plantard.

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.

First, the development of motion analysis tools is mandatory in order to provide additional information to the ergonomists and help them to analyse the work environment. Particularly, an analysis of the muscle forces involved in the motion generation is a very important information with regard to the ergonomics of a task. Several methods can lead to an estimation of these muscle forces. In a study we developed, we tried to assess the level of confidence for results obtained with an inverse dynamics method from real captured work tasks. The chosen tasks were meat cutting tasks, well known to be highly correlated to musculoskeletal troubles appearance in the slaughter industry.

The experimental protocol consisted in recording three main data during meat cutting tasks, and analysing their variation when some of the workstation design parameters were changing.

Then the motion was replayed in the AnyBody modeling system (AnyBody, Aalborg, Denmark) in order to obtain muscle forces generated during the motion. A trend comparison has been done, comparing recorded and computed muscle activations. Results showed that most of the computed activations were qualitatively close from the recorded ones (similar shapes and peaks), but quantitative comparison led to major differences between recorded and computed activations (the trend followed by the recorded activations in regard of a workstation design parameter, such as the table height, is not obtained with the computed activations) [15]. We currently explore those results to see if the fact that co-contraction of single joints muscles is badly estimated by classical inverse dynamics method can be a reason of this issue. We also work on the co-contraction simulation in order to improve the results.

This work has been done in collaboration with the Center for Sensory-motor Interaction (SMI, Aalborg University, Aalborg, Denmark), particularly Mark de Zee (Associate Professor) and Pascal Madeleine (Professor).

Furthermore, the fidelity of the VR simulator has to be evaluated. For example, a simulator for assembly task has been evaluated in comparing different types of interaction : real, virtual and virtual + force feedback. Objective and subjective metrics of discomfort led to highlight the influence of the environment on motor control and sensory feedback, changing more or less deeply the way the task is performed. The results particularly showed a distortion between the user's subjective rating of discomfort and the objective value associated to the postures they reached during the task execution. Nevertheless, scores obtained in real and virtual environments for objective and subjective indicators of discomfort were highly correlated [17], [16]. It indicates that despite the differences, the gap between real and virtual environments can be fulfilled. This work has been done within the frame of the european project FP7 VISIONAIR.

At last we proposed in collaboration with Thierry Duval (Hybrid team, Rennes) a new architecture for information sharing and bridging in collaborative virtual environments in application to ergonomics studies. This work has been awarded with a best paper award at The 4th IEEE conference on Cognitive Infocommunications (CogInfoCom 2013) [28].

### 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).

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. The extremes of the distance between the two knees along the longitudinal axis provides us with very accurate gait events detection compared to previous works.

A second contribution consisted in defining a global gait asymmetry index according to depth images provided by a Kinect. In previous works this index relied on computing ratio between joint angles. With a Kinect, joint angles may be very noisy that could affect the asymmetry index. We have introduced a new index which is directly deduced from depth images without any joint angle estimation nor skeleton fitting. The method consists in building a model of the gait cycle of the patient by averaging depth images recorded along several cycles. As a consequence the noise within the instantaneous depth images is filtered leading to accurate surfaces of the patient gait (leading to a 3D+time data structure). The main vertical axis of the surface is used to define a symmetry plane. Consequently surfaces of the right part of the body can be symmetrized to be compared to the left part at compatible times in the gait cycle (such as a right foot strike is symmetrized to be compared to a left foot strike). The comparison between the two surfaces leads to a promising asymmetry index. The results (see Figure 4) demonstrate that this method is able to significantly distinguish asymmetrical gaits obtained by adding a 5cm sole under one of the feet of healthy subjects. Ongoing works consist in comparing this index to previously published ones which were based on accurate motion capture data. It will also be applied to unimpaired gaits of pathological subjects.



*Figure 4. Longitudinal (DAI) and lateral (LAI) Asymmetry indexes computed thanks to depth images for normal gait and two artificial modification (adding a 5cm sole below the left or the right foot). The asymmetry index is computed all along the gait cycle and was able to statistically distinguish asymmetrical gaits.*

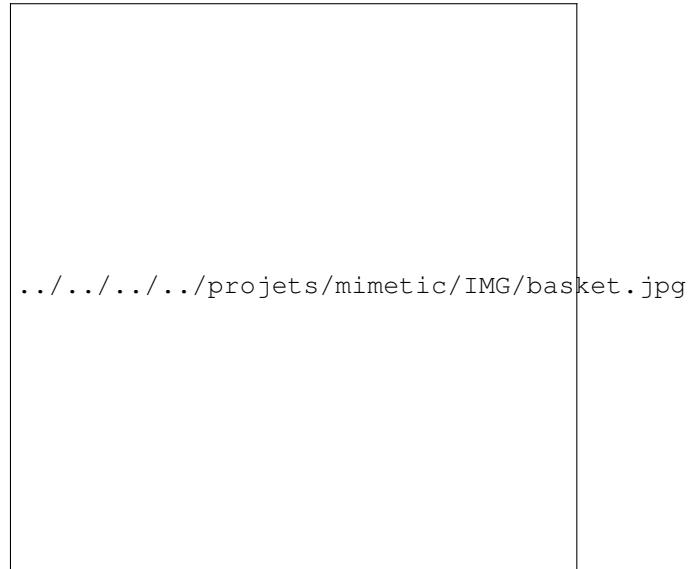
## 6.4. VR and Sports

**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 Room, Rennes): 1) first-person view (see Figure 5), 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 difference exists in ball speed between first-person view condition compared to real condition whereas no difference 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 Brassov in Romania.

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). Pseudohaptics 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 6. Thus, to simulate the lifting of a "heavy" dumbbell, the 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. 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. First-person view condition in the basket free-throw performed in a cave (Immersia Room, France).*



*Figure 6. 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 conditions from the lightest to the heaviest mass.*

## 6.5. Autonomous Virtual Humans

### 6.5.1. Space and Time Constrained Task Scheduling for Crowd Simulation

**Participants:** Carl-Johan Jorgensen, Fabrice Lamarche [contact].

Crowd distribution in cities highly depends on how people schedule their daily activities. When performing an intended activity, people decisions and behavior mainly consist in scheduling tasks that compose this activity, planning paths between locations where these tasks should be performed, navigating along the planned paths and performing the scheduled tasks.

We proposed a task scheduling model aims at selecting where, when and in which order several tasks, representing an intended activity, should be performed. The proposed model handles spatial and temporal constraints relating to the environment and to the agent itself. Personal preferences, characterizing the agent, are also taken into account. Produced task schedules are optimized on the long term and exhibit adequate choices of locations and times with respect to the agent intended activity and its environment. Once computed, these task schedules are relaxed and used to drive a microscopic crowd simulation in which observable flows of pedestrians emerge from the scheduled individual activities. Such simulations are easy to produce and do not require the use of a complex decisional model. In terms of validation, we conducted an experiment that shows that our algorithm produces task schedules which are representative of humans' ones.

This work is part of the iSpace&Time project in which virtual cities are populated with virtual pedestrians and vehicles.

### 6.5.2. Long term planning and opportunism

**Participants:** Philippe Rannou, Fabrice Lamarche [contact].

Autonomous virtual characters evolve in dynamic virtual environments in which changes may be unpredictable. One main problem when dealing with long term action planning in dynamic environment is that an agent should be able to behave properly and adapt its behavior to perceived changes while still fulfilling its goals.

We propose a system that combines long term action planning with failure anticipation and opportunism. The system is based on a modified version of an HTN planning algorithm. It generates plans enriched with information that enable a monitor to detect relevant changes of the environment. Once such changes are detected, a plan adaptation is triggered. Such adaptations include modifying the plan to react to a predicted failure and more importantly to exploit opportunities offered by the environment.

This system has been extended to better take into account the relationship between action planning and the environment. It is now combined with our space and time constrained tasks scheduling system (Cf. 6.5 ) to optimize the choice of locations where actions should be performed.

## 6.6. Interactive Virtual Cinematography

**Participants:** Marc Christie [contact], Christophe Lino, Cunka Sanokho.

The domain of Virtual Cinematography explores the operationalization of rules and conventions pertaining to camera placement, light placement and staging in virtual environments. Within the context of the ANR CHROME project, we have tackled the problem of portraying events in complex crowd simulations using steering behaviors. The system we proposed relies on Reynolds' model of steering behaviors to control and locally coordinate a collection of camera agents similar to a group of reporters. In our approach, cameras are either in a scouting mode, searching for relevant events to convey, or in a tracking mode following one or more unfolding events. The key benefit, in addition to the simplicity of the steering rules, holds in the capacity of the system to adapt to the evolving complexity of crowd simulations by self-organizing the cameras to track most of the events. The results have been presented at the Motion in Games conference [21].

We have also created a table-top interactive application to offer collaborative and high-level control on multi-dimensional and temporal data. This has been applied to the collaborative control of cinematographic parameters in a virtual movie, using our cinematographic engine [26].

In the ANR project Cinécitta, we have proposed means to evaluate the sense of balance in synthetic shots. Balance represents the equilibrium of visual weights in the screen, *i.e.* equilibrium of the visual interests one perceives. Balance is a key criteria in the aesthetics of a shot, and only a few approaches have seriously tackled this issue. In our approach, we rely on a dataset of well-balanced shot extracted from real movies to construct a balance feature space. The balance feature space is then used to estimate the sense of balance in new synthetic shot. We have furthermore extended the approach by automatically recomputing viewpoints to improve balance. A journal paper is under submission in Computer Graphics Forum.

## 6.7. Interactive Storytelling

**Participants:** Marc Christie [contact], Hui-Yin Wu.

In 2013, within the Inria Associate Team FORMOSA (see 8.3.1.1 ), we have proposed a framework for the creation of parametrable and personalized stories in interactive storytelling. In any kind of storytelling, the success of the story relies both on the intricate plot design and control of the author as well as the emotional feedback of the user. With the assistance of computing algorithms combined with the maturing understanding of narrative structures, it is possible for interactive stories to create a more personalized, engaging, and well-controlled narrative content to users than traditional linear narrative. And with the emergence of new storytelling technologies, critical issues concern the creation of such complex narratives in virtual 3D environments, and the coherent simulation of these interactive narratives.

In the framework we proposed, the author can specify characteristics on the story structure and fragments (pieces of story) in order to generate variations of interactive stories. The characteristics we consider are genre, story complexity, and Chatman's modes of plot (eg a good hero fails). The story generation model we devised combines a branching story structure with a three-step graph traversal algorithm that filters and recombines story fragments from the characteristics, generating a high-level interactive script that satisfies all authorial constraints, and provides sufficient abstraction from the technical implementation. The script is then simulated in a real-time storytelling system, featuring autonomous characters and automatic camera control. The work has been presented as a short paper in the CASA conference [30].

We then extended this approach to handle temporal aspects of discourse in stories (*i.e.* how to temporally rearrange fragments of a story while maintaining consistency and logic whatever the user's choices). By rewriting our graph traversal algorithm (which filters inconsistent branches, and propagates constraints along the branches), and performing the graph traversal on each choice selected by the user, we enable the simulation of consistent temporal variations in stories. This typically allows the creation of flashbacks, flashforwards, parallel and embedded stories. Early results have been presented as a poster at Motion in Games 2013 [21].

## 6.8. Haptic Cinematography

**Participant:** Marc Christie [contact].

In 2013, we have demonstrated an approach to Haptic Cinematography in very selective events (2013 CHI conference [40], 2013 Siggraph Emerging Technology [41], 2013 UIST conference [39]). This is joint work with members of the Hybrid team (Anatole Lécuyer, Fabien Danieau) and members of the Technicolor Company (Philippe Guillotel, Nicolas Mollet, Julien Fleureau). Haptic cinematography consists in enhancing our audio-visual experience of movies by adding haptic effects related to the semantics of camera motions. Camera motions in movies, which are typically non-diegetic elements in a narrative, tend to enhance user experience both visually and emotionally. The questions we address here are (i) whether the coupling between camera motions and haptic motions improve this experiences and (ii) what are the rules and recommendations for coupling these motions. Results, that we published in the IEEE Multimedia journal [8], demonstrate that (i) the coupling is effective when precisely synchronized, (ii) the direction of motions between the camera and the haptic motions do not need to be correlated, and (iii) haptic metaphors can easily be perceived by the spectators. This opens great perspectives as to how haptic devices can enhance audio-visual contents in more subtle ways than straightforward mappings between diegetic elements and haptic motions.



## 6.9. Biomechanics for avatar animation

**Participants:** Julien Pettré [contact], Charles Pontonnier, Georges Dumont, Franck Multon, Ana Lucia Cruz Ruiz, Steve Tonneau.

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 8.1.5 .

Ana-Lucia Cruz-Ruiz has been recruited as a PhD student since november 2013 to begin to work on musculoskeletal-based methods for avatar animation. More precisely, the goal of this thesis is to define and evaluate a modular and multiscale whole-body musculoskeletal model usable to analyze and human movement and synthesize realistic avatar animations. The specificity of the subject is hidden in the words “modular” and “multiscale”. “Modular” says that the model has to be easily tunable to be modified in accordance with the investigated motor control theories (uncontrolled manifold, motor synergies,...). “Multiscale” means that the model has to exhibit multiple levels of details cohabiting at the same time, depending on the region of interest investigated. At last, the model has to be easily scalable, in order to be applied to different morphologies. Moreover, she currently explores musculoskeletal-based simplified joint behaviors to improve torque-based dynamics applications.

We also address the problem of planning human motion in constrained environment. In previous approach, planning human motion is performed based on robotics planning algorithms the objective of which is to avoid obstacles. In our approach, we suggest that creating contacts with the obstacles of the environment is actually a mean to perform a motion tasks. We thus model human motion as a sequence of contacts between humans and obstacles. A contact planner is being developed, and results being prepared for publication.

## 6.10. Crowds

**Participants:** Julien Pettré [contact], Anne-Hélène Olivier, Julien Bruneau, Jonathan Perrinet, Kevin Jordao, David Wolinski.

### 6.10.1. Analysis of Locomotion Trajectories during Collision Avoidance

The experimental observation of physical interactions between real walkers is for us a great source of inspiration for the design of realistic microscopic models of crowd simulation. This year, we have continued analysing locomotion trajectories of real walkers during collision avoidance tasks. Analysis focused on individual strategies and role set to solve such a reciprocal interaction. Our analysis revealed that walkers combine re-orientation and speed adaptations to avoid collisions, but more importantly, that the strategies, as well as the global amount of adaptations is dependent on the role each one has in the avoidance (e.g., passing first, giving way). Our results are reported in [13]. In addition, we inspected the role of psychologic factors on the metrics of interactions [27].

### 6.10.2. Evaluation of Locomotion Trajectories performed in Virtual Reality

Virtual Reality rooms are physically limited in space, and prevent users virtually walking by really walking in larger virtual spaces: a locomotion interface is employed to overcome this issue. The interface is composed of a peripheral device, such as joystick, as well as of a software component which transform users' actions on the peripheral device into a virtual locomotion. In this work, we wondered if users were performing similar trajectories in virtual than in vivo: such question is important when aiming at using VR for motion analysis purpose. We evaluated the bias introduced by several couples of devices and software components during the execution of goal directed locomotion tasks. As reported in [7], impressive similarities on the formed trajectories even when the device control motions are radically different in comparison with walking motions.

### 6.10.3. Virtual Populations for large-scale digital environments and Cultural Heritage Applications

We are developing techniques dedicated to the animation of large virtual populations at very low computational cost based on the crowd patches techniques. Crowd patches can be described as 3D animated textures that small



groups animations. They are composed in space to form large population. This year, we coupled the crowd patches approach with mutable shape models: such association enable users cdesigning patches composition in an interactive manner, as introduced in [22]. We applied those techniques to design populations of some old Malaysian trading ports [25].

#### ***6.10.4. Macroscopic derivations of microscopic simulation models***

Crowd phenomenon exhibit macroscopic structures which derive from the combination of local interactions between individuals. Together with the IMT in Toulouse in the frame of the ANR-Pedigree project (term. 2012), the microscopic models developed in our team has been derived into macroscopic models to demonstrate their ability to provoke the mergence of some typical macroscopic structures [35], [36].

## MINT Project-Team

# 6. New Results

## 6.1. Human limits in small unidirectional mouse movements

**Participants:** Jonathan Aceituno [correspondant], Géry Casiez, Nicolas Roussel.

Computer mouse sensors keep increasing in resolution. The smallest displacement they can detect gets smaller, but little is known on our ability to control such small movements. Small target acquisition has been previously tackled, but the findings do not apply to the problem of finding the useful resolution of a user with a mouse, which corresponds to the smallest displacement (s)he can reliably produce with that device. In [16], we detail this definition and provide an associated experimental protocol to measure the useful resolution. We then report on the results of a study suggesting that high-end mice are not likely to be used to their full potential. We further comment on the different strategies used by participants to achieve best performance, and derive implications for user interfaces.

## 6.2. Small, Medium, or Large? Estimating the User-Perceived Scale of Stroke Gestures

In [27], we show that a large consensus exists among users in the way they articulate stroke gestures at various scales (i.e., small, medium, and large) and formulate a simple rule that estimates the user-intended scale of input gestures with 87% accuracy. Our estimator can enhance current gestural interfaces by leveraging scale as a natural parameter for gesture input, reflective of user perception (i.e., no training required). Gesture scale can simplify gesture set design, improve gesture- to-function mappings, and reduce the need for users to learn and for recognizers to discriminate unnecessary symbols.

## 6.3. Métamorphe : a shape changing keyboard

Métamorphe is a keyboard with mobile keys [21]. Whether keys are pressed or released, they can be at their usual height, or raised. This mechanism allows both to provide haptic feedback to ease eyes-free interaction, and to access the side of the keys. The sides of the keys can be pushed, like the top of the keys. Therefore each key can be mapped to several actions. For instance this could be useful for command selection.

## 6.4. Designing Intuitive Multi-touch 3D Navigation Techniques

**Participants:** Géry Casiez, Damien Marchal [correspondant], Nicolas Roussel, Clement Moerman.

Multi-touch displays have become commonplace over recent years. Numerous applications take advantage of this to support interactions that build on users' knowledge and correspond to daily practices within the real world. 3D applications are also becoming more common on these platforms, but the multi-touch techniques for 3D operations often lag behind 2D ones in terms of intuitiveness and ease of use. Intuitive navigation techniques are particularly needed to make multi-touch 3D applications more useful, and systematic approaches are direly needed to inform their design: existing techniques are still too often designed in ad-hoc ways. In [25], we propose a methodology based on cognitive principles to address this problem. The methodology combines standard user-centered design practices with optical flow analysis to determine the mappings between navigation controls and multi-touch input. It was used to design the navigation technique of a specific application for our industrial partner Idées3Com. The resulting technique proved to be more efficient and preferred by users when compared to existing ones, which provides a first validation of the approach.



*Figure 1. a) Métamorphe concept: the user presses the control key, keys corresponding to hotkeys rise b) key mounted on a solenoid, with force sensors on the sides c) press on the top of the key d) press on the right of the key e) press on the left of the key.*

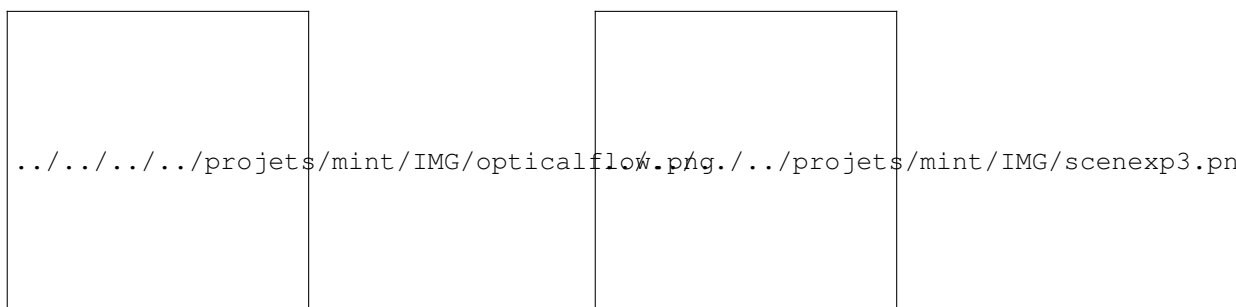


Figure 2. (Left) The optical flow for camera movements are used to design the shape of the interaction gestures. (Right) The evaluation scenario used to compare several state of the art navigation techniques.

## 6.5. Mockup Builder: 3D modeling on and above the surface

Mockup Builder [11] is a semi-immersive environment for conceptual design which allows virtual mockups to be created using gestures. Our goal is to provide familiar ways for people to conceive, create and manipulate three-dimensional shapes. To this end, we developed on-and-above-the-surface interaction techniques based on asymmetric bimanual interaction for creating and editing 3D models in a stereoscopic environment. Our approach combines both hand and finger tracking in the space on and above a multi-touch surface. This combination brings forth an alternative design environment where users can seamlessly switch between interacting on the surface or above it to leverage the benefit of both interaction spaces. A formal user evaluation conducted with experienced users shows very promising avenues for further work towards providing an alternative to current user interfaces for modeling.

## 6.6. Towards Many Gestures to One Command: A User Study for Tabletops

**Participants:** Yosra Rekik, Laurent Grisoni [correspondant], Nicolas Roussel.

This work has been accepted as a long paper at Interact 2013. Multi-touch gestures are often thought by application designers for a one-to-one mapping between gestures and commands, which does not take into account the high variability of user gestures for actions in the physical world; it can also be a limitation that leads to very simplistic interaction choices. Our motivation is to make a step toward many-to-one mappings between user gestures and commands, by understanding user gestures variability for multi-touch systems; for doing so, we set up a user study in which we target symbolic gestures on tabletops. From a first phase study we provide qualitative analysis of user gesture variability; we derive this analysis into a taxonomy of user gestures, that is discussed and compared to other existing taxonomies. We introduce the notion of atomic movement; such elementary atomic movements may be combined throughout time (either sequentially or in parallel), to structure user gesture. A second phase study is then performed with specific class of gesture-drawn symbols; from this phase, and according to the provided taxonomy, we evaluate user gesture variability with a fine grain quantitative analysis. Our findings indicate that users equally use one or two hands, also that more than half of gestures are achieved using parallel or sequential combination of atomic movements. We also show how user gestures distribute over different movement categories, and correlate to the number of fingers and hands engaged in interaction. Finally, we discuss implications of this work to interaction design, practical consequences on gesture recognition, and potential applications.

## 6.7. Sub-space gestures: elements of design for mid-air interaction with distant displays

**Participants:** Hanae Rateau, Laurent Grisoni [correspondant], Bruno de Araujo.

(Research report, accepted to publication in a modified version to IUI 2014). Multi-touch gestures are often thought by application designers for a one-to-one mapping between gestures and commands, which does not take into account the high variability of user gestures for actions in the physical world; it can also be a limitation that leads to very simplistic interaction choices. Our motivation is to make a step toward many-to-one mappings between user gestures and commands, by understanding user gestures variability for multi-touch systems; for doing so, we set up a user study in which we target symbolic gestures on tabletops. From a first phase study we provide qualitative analysis of user gesture variability; we derive this analysis into a taxonomy of user gestures, that is discussed and compared to other existing taxonomies. We introduce the notion of atomic movement; such elementary atomic movements may be combined throughout time (either sequentially or in parallel), to structure user gesture. A second phase study is then performed with specific class of gesture-drawn symbols; from this phase, and according to the provided taxonomy, we evaluate user gesture variability with a fine grain quantitative analysis. Our findings indicate that users equally use one or two hands, also that more than half of gestures are achieved using parallel or sequential combination of atomic movements. We also show how user gestures distribute over different movement categories, and correlate to the number of fingers and hands engaged in interaction. Finally, we discuss implications of this work to interaction design, practical consequences on gesture recognition, and potential applications.

## 6.8. Merging two tactile stimulation principles: Electrovibration and Squeeze film effect

**Participants:** Michel Amberg, Frédéric Giraud, Clément Nadal, Betty Semail [correspondant].

Electrovibration and squeeze film effect can modify the perception a user has of a flat surface, with opposite action. In fact, electrovibration increases the friction of the finger on the surface, while the squeeze film reduces it. These two stimulation principles are compatible, and in this work [23], we wanted to merge them in a tactile stimulator, in order to enhance the control of the lateral force. Our approach was to identify the effect of each tactile stimulation, and we proposed its modelling: the dynamic of the mechanical response of the fingerpulp has to be taken into account between the programmed stimulus and the resulting lateral force. We have shown also that the two techniques may be used simultaneously accounting to a few precautions. From the first experimental trials, the conclusion here is that the squeeze film effect is able to reduce tangential forces generated by the electrostatic forces, by going on acting on the friction coefficient.

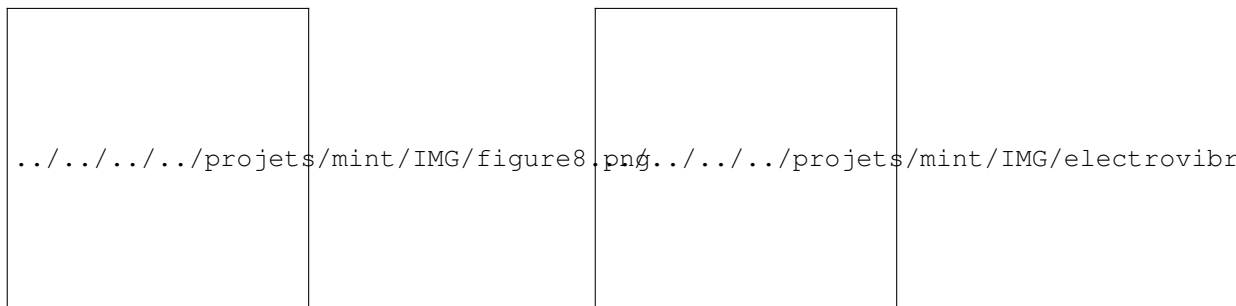


Figure 3. (Left) The experimental test bench to measure the forces produced during the stimulation. (Right) The tactile stimulator merging the two stimulation principles.

## POTIOC Team

# 6. New Results

## 6.1. Navigation techniques in 3D digital cities on mobile touch devices

**Participants:** Jacek Jankowski, Thomas Hulin, Martin Hachet.

This study 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 3 ). 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 accepted for publication at the 3DUI 2014 conference [15].



*Figure 3. Four techniques for navigating in a 3D city on a mobile touch device.*

As part of this project on Navigation in 3D digital cities, the Potioc group also built a tutorial about interaction techniques for 3D environments. It was presented at Eurographics 2013 [13] and Web3D 2013 [14]. The goal of this work is to provide an up-to-date state-of-the-art of this topic to the community.

## 6.2. Interaction with spatial augmented reality for physical drawing

**Participants:** Jérémy Laviolle, Martin Hachet.

We developed tools that enable precise interactive projection on pieces of paper. The sheets of paper are tracked by a camera while the user's inputs (e.g., touch and hovering events) are detected by a Kinect. The paper acts as a screen, its image coming from an overhead projector. The focus of this work is to use such tools to assist the creation of physical drawings and painting. In this context we propose Digital Construction Lines (DCL), in opposition with physical construction lines. Traditionally, the structure of a physical drawing can be created with construction lines which are light pencil strokes. These strokes are then erased during the drawing process. With DCL, it is not required to erase the construction lines anymore. Furthermore, it is possible to create construction lines on fragile material like a canvas for waterpainting or on fresh paint. It also enables construction lines on a dark canvas. In addition to these projection advantages, it is possible to create these DCL interactively and directly onto the support. Consequently, the DCL complement the physical ones during the creation process.

We investigated in a user study if the DCL could effectively replace the physical construction lines, and compared the performance (speed, cleanliness) between the two kinds of construction lines. In this user study we also evaluated the quality and usability of projection of thin lines in a fully controlled environment with a low-cost setup. The study showed that DCL could effectively replace physical construction lines, even though it might not be desirable. The study also showed that the drawing experience was as pleasant with projection, and with the usual tools. The feedback about the quality of tracking and projection was also positive. The only negative evaluation concerned the size of the projection area, which was limited by the resolution of the projector. This work was published as part of Jérémy Laviolle's PhD thesis [4].

PapART was also used as part of a museum exhibition on the Lascaux caves, together with other 3D UI from Potioc. This exhibition has provided us with the opportunity to experiment with touch-based interfaces for manipulating 3D virtual objects. We targeted three tasks: observing rare objects with Cubtile, reassembling object fragments with Toucheo, and reproducing artwork with PapART [7] (see Figure 5). These exhibitions allowed us to experiment our systems in real conditions. It led to a Living Lab, where the visitors can test our devices.

## 6.3. Rouages: Revealing the Mechanisms of Digital Musical Instruments to the Audience

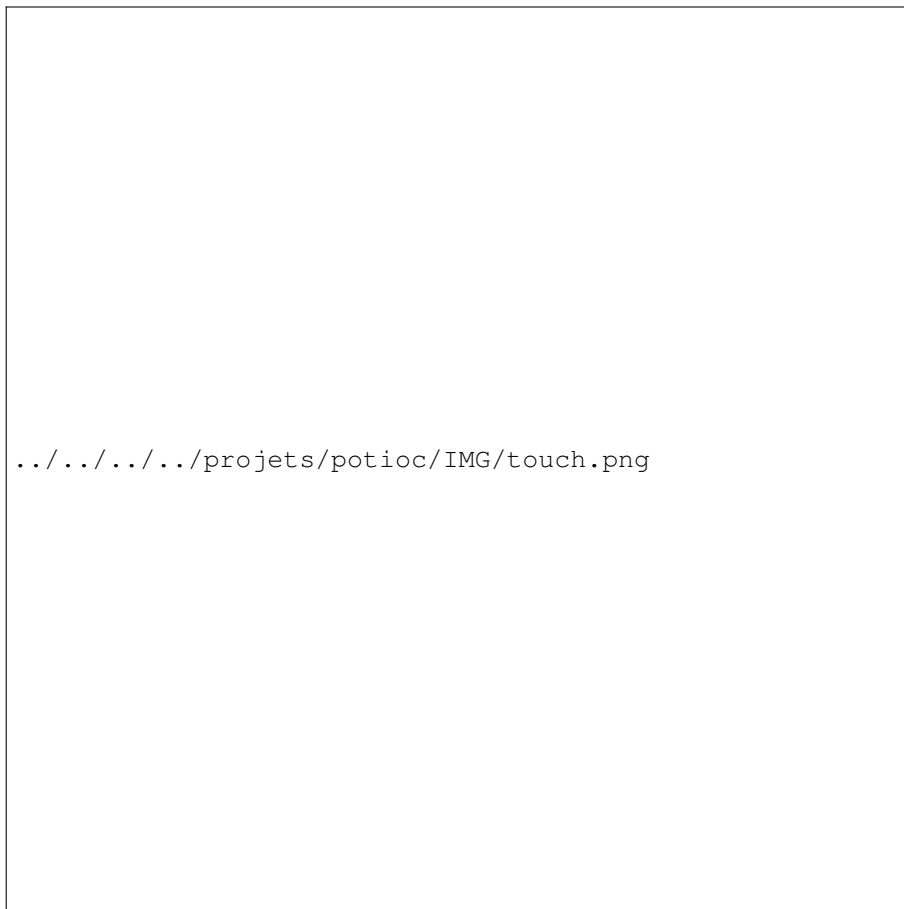
**Participants:** Florent Berthaut, Martin Hachet, Pierre-Marie Plans.

We have developed Rouages [10], a mixed-reality display system associated with a 3D visualization application. Rouages reveals the mechanisms of digital musical instruments in two ways. First, by amplifying musicians' gestures with virtual extensions of the sensors. Second, by representing the sound components with 3D shapes and specific behaviors and by showing the impact of musicians' gestures on these components. In addition, we have explored new setups to enhance collaboration between musicians using our VR-based instruments. This is illustrated in Figure 6.

## 6.4. Gateway driving simulator

**Participants:** Florian Larrue, Pauline Davignon, Martin Hachet.

As part of the SIMCA FUI project, the POTIOC team focuses on the design and evaluation of a gateway driving simulator (see Figure 7), to teach drivers how to drive an airport gateway in virtual reality, i.e., in a safe and cost-effective environment. We conducted a comprehensive user study to assess the impact of various parameters on user performances. This study allowed us to provide a set of recommendations for the design of an actual simulator.



*Figure 4. Using Digital Construction Lines for spatial augmented reality-based physical drawing.*





*Figure 5. Manipulation of a 3D model and lighting conditions for drawing on a prehistoric object in a museum.*



*Figure 6. Collaborative setup for enhancing interaction between immersed musicians.*



Figure 7. Gateway simulator.

## 6.5. Training Approaches for Brain-Computer Interfaces

**Participants:** Fabien Lotte, Florian Larrue, Christian Mühl.

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 human in the loop. However, how well the user masters the BCI skill is also a key element explaining BCI robustness. Unfortunately, despite the importance of BCI training protocols, they have been scarcely studied so far, and used mostly unchanged for years. In our work, we advocate that current human training approaches for spontaneous BCI are most likely inappropriate. We notably studied instructional design literature in order to identify the key requirements and guidelines for a successful training procedure that promotes a good and efficient skill learning. This literature study highlighted that current spontaneous BCI user training procedures satisfy very few of these requirements and hence are likely to be suboptimal. We therefore identified the flaws in BCI training protocols according to instructional design principles. We also proposed new research directions that are theoretically expected to address some of these flaws and to help users learn the BCI skill more efficiently. This work has been published in the *Frontiers in Neuroscience* journal [9].

On a related topic, together with colleagues from Inria Rennes (A. Lécuyer and L. Bonnet, Hybrid team) we explored the design and evaluation of multiuser BCI applications, notably to see their impact on user training and performance. We created a multiuser videogame called BrainArena in which two users can play a simple football game by means of two BCIs. They can score goals on the left or right side of the screen by simply imagining left or right hand movements (see Figure 8). To add another interesting element, the gamers can play in a collaborative manner (their two mental activities are combined to score in the same goal),



Figure 8. Multiplayer BCI-based gaming and feedback [5].

or in a competitive manner (the gamers must push the ball in opposite directions). Interestingly enough, our results showed that compared to a single player version of the same BCI-game, collaborative multiplayer BCI-gaming increased the motivation and performance of the most skilled of each player pairs, while leaving the performance of the other players unchanged, hence proving a useful tool to improve BCI training. This work has been published in the IEEE Transactions on Computational Intelligence and AI in Games journal [5].

## 6.6. Inducing, measuring and estimating mental and psychosocial stress from physiological signals

**Participants:** Christian Mühl, Camille Jeunet, Fabien Lotte.

Stress is a major societal issue with negative impacts on health and economy. Physiological computing offers a continuous, direct, and unobtrusive method for stress level assessment and computer-assisted stress management. However, stress is a complex construct and its physiology can vary depending on its source, for example cognitive load or social evaluation. To study the feasibility of physiology-based load-invariant psychosocial stress-detection, we designed a stress-induction protocol able to independently vary the relevant types of psychophysiological activity: mental and psychosocial stress. In [27], [17], we validate the efficacy of our protocol to induce psychosocial and mental stress. Our participants (N=24) had to perform a cognitive task associated with two workload conditions (low/high mental stress), in two contexts (low/high psychosocial stress), during which we recorded subject's self-reports, behavior, physiology and neurophysiology. Questionnaires showed that the subjectively perceived level of anxiety varied with the psychosocial stress induction, while perceived arousal and mental effort levels vary with mental stress induction. Behavior and physiology corroborated the validity of our protocol further. Heart rate and skin conductance globally increased after psychosocial stress induction relative to the non-stressful condition. Moreover, we demonstrated that higher workload tasks (mental stress) led to decrease in performance and a marked increase of heart rate.



Figure 9. The setup of the experiment to induce and measure mental and psychosocial stress.

Based on this protocol, we also explored the effect of stress on workload estimation. Workload estimation from electroencephalographic signals (EEG) offers a highly sensitive tool to adapt the human-computer interaction to the user state. To create systems that reliably work in the complexity of the real world, a robustness against contextual changes (e.g., mood), has to be achieved. To study the resilience of state-of-the-art EEG-based workload classification against stress, we test the capability of the workload classifier to generalize across affective contexts (stress/non-stress). We show that the classifier is able to transfer between affective contexts, though performance suffers. However, cross-context training is a simple and powerful remedy allowing the extraction of features more resilient to task-unrelated variations in signal characteristics, leading to a performance comparable to within-context training and testing.

## 6.7. Exploring electroencephalography as an evaluation method for human-computer interaction

**Participants:** Jérémy Frey, Léonard Pommereau, Fabien Lotte, Christian Mühl, Martin Hachet.

Evaluating human-computer interaction is essential as a broadening population uses machines, sometimes in sensitive contexts. However, traditional evaluation methods may fail to combine real-time measures, an objective approach and data contextualization. We presented a review seeking how neuroimaging techniques can respond to such needs. We focused on electroencephalography (EEG), as it could be handled effectively during a dedicated evaluation phase. We identified workload, attention, vigilance, fatigue, error recognition, emotions, engagement, flow and immersion as being recognizable by EEG. We find that workload, attention and emotions assessments would benefit the most from EEG. Moreover, we advocate to study further error recognition through neuroimaging to enhance usability and increase user experience. This review paper was published in the proceeding of the Physiological Computing Systems (PhyCS) conference [12].



*Figure 10. Stereoscopic display studied through EEG by varying apparent depth.*

Along this line of research, we also explored whether it was possible to assess the zone of comfort in stereoscopic displays using electroencephalography. Indeed, the conflict between vergence (eye movement) and accommodation (crystalline lens deformation) occurs with every stereoscopic display. It could cause important stress outside the "zone of comfort", when stereoscopic effect is too strong. This conflict has already been studied using questionnaires, during viewing sessions of several minutes. We built an experimental protocol (see Figure 10 ) which compares two different comfort conditions using electroencephalography over short viewing sequences. Analyses showed significant differences both in event-related potentials and in frequency bands power. By extending our protocol it should be possible to study at the same time comfort and depth perception, having a better understanding of stereoscopy.

## REVES Project-Team

### 5. New Results

#### 5.1. Plausible and Realistic Image Rendering

##### 5.1.1. Depth Synthesis and Local Warps for Interactive Image-based Navigation

**Participants:** Gaurav Chaurasia, Sylvain Duchene, George Drettakis.



Figure 3. Novel views generated by the image-based rendering approach of [12] along with a visualization of novel camera position relative to the 3D scene and input cameras. This approach is among the first to handle very complex urban scenes such as those shown here and provide a stable solution for viewpoints that are far from the input cameras.

Modern camera calibration and multi-view stereo techniques enable users to smoothly navigate between different views of a scene captured using standard cameras. The underlying automatic 3D reconstruction methods work well for buildings and regular structures but often fail on vegetation, vehicles and other complex geometry present in everyday urban scenes. Consequently, missing depth information makes image-based rendering for such scenes very challenging. This paper introduces a new image-based rendering algorithm that is robust to missing or unreliable geometry, providing plausible novel views even in regions quite far

from the input camera positions. The approach first oversegments the input images, creating superpixels of homogeneous color content which preserve depth discontinuities. It then introduces a *depth synthesis* step for poorly reconstructed regions. It defines a graph on the superpixels and uses *shortest walk* traversals to fill unreconstructed regions with approximate depth from regions that are well-reconstructed and similar in visual content. The superpixels augmented with synthesized depth allow a local shape-preserving warp which warps each superpixel of the input image to the novel view without incurring distortions and preserving the local visual content within the superpixel. This allows the approach to effectively compensate for missing photoconsistent depth, the lack of which is known to cause rendering artifacts. The final rendering algorithm blends the warped images, using heuristics to avoid ghosting artifacts. The results demonstrate novel view synthesis in real time for multiple challenging scenes with significant depth complexity (see Figure 3), providing a convincing immersive navigation experience. The paper presents comparisons with three of the state of the art image-based rendering techniques and demonstrate clear advantages.

This work was in collaboration with Olga Sorkine-Hornung at ETH Zurich. It has been published in ACM Transactions on Graphics 2013 [12] and presented at SIGGRAPH.

### 5.1.2. *Megastereo: Constructing High-Resolution Stereo Panoramas*

**Participant:** Christian Richardt.

There is currently a strong consumer interest in a more immersive experience of content, such as 3D photographs, television and cinema. A great way of capturing environmental content are panoramas (see Figure 4). We present a solution for generating high-quality stereo panoramas at megapixel resolutions. While previous approaches introduced the basic principles, we show that those techniques do not generalise well to today's high image resolutions and lead to disturbing visual artefacts. We describe the necessary correction steps and a compact representation for the input images in order to achieve a highly accurate approximation to the required ray space. In addition, we introduce a flow-based upsampling of the available input rays which effectively resolves known aliasing issues like stitching artefacts. The required rays are generated on the fly to perfectly match the desired output resolution, even for small numbers of input images. This upsampling is real-time and enables direct interactive control over the desired stereoscopic depth effect. In combination, our contributions allow the generation of stereoscopic panoramas at high output resolutions that are virtually free of artefacts such as seams, stereo discontinuities, vertical parallax and other mono-/stereoscopic shape distortions.

This work was carried out in collaboration with Yael Pritch, Henning Zimmer and Alexander Sorkine-Hornung at Disney Research Zurich. The paper has been published as an oral presentation at CVPR 2013 [20].

### 5.1.3. *Probabilistic Connection Path Tracing*

**Participants:** Stefan Popov, George Drettakis.

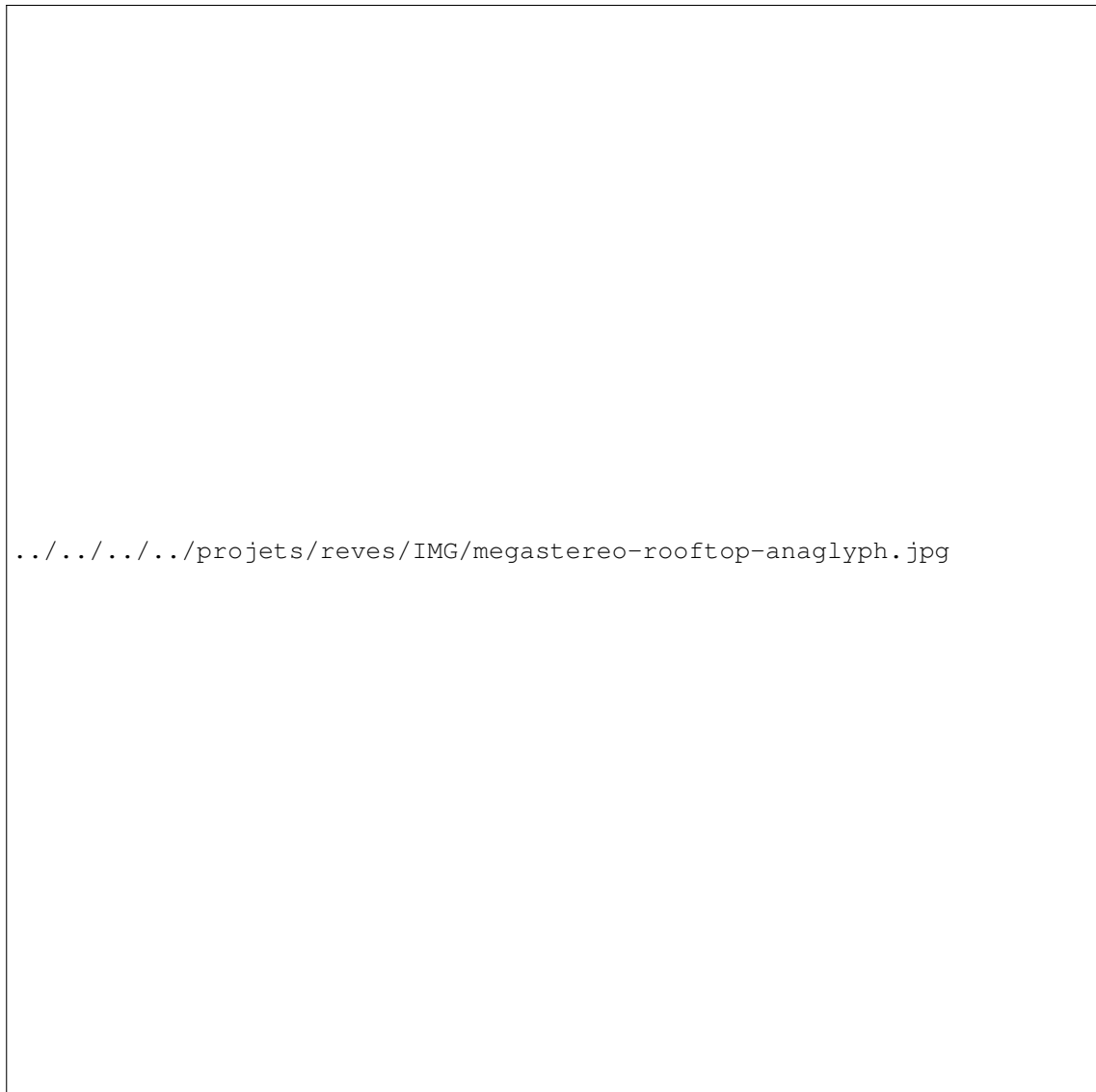
We propose an unbiased generalization of bi-directional path tracing (BPT) that significantly improves its rendering efficiency. Our main insight is that the set of paths traced by BPT contains a significant amount of statistical information, that is not exploited.

BPT repeatedly builds an eye and a light sub-paths, connects them, estimates the contribution to the corresponding pixel and then throws the path away. Instead, we propose to first trace all eye and light sub-paths, and then probabilistically connect each eye sub-path to one or more light sub-paths. From a Monte-Carlo perspective, this will connect each light to each eye sub-path, substantially increasing the number of paths used to estimate the solution. As a result, the convergence will be significantly increased as well.

This work is a collaboration with Frédo Durand from the Massachusetts Institute of Technology, Cambridge and Ravi Ramamoorthi from University of California, Berkeley in the context of the CRISP Associated Team.

### 5.1.4. *Parallelization Strategies for Associative Image Processing Operators*

**Participants:** Gaurav Chaurasia, George Drettakis.



*Figure 4. A stereoscopic panorama corrected and stitched using our techniques, shown as red-cyan anaglyph image.*



Basic image processing operations have been optimized on a case-by-case basis such as prefix sums and recursive filters. Moreover, these optimized algorithms are very complicated to program because parallelization involves non-trivial splitting of the input domain of the operator. The target of this is to generalize the optimization heuristics of a generic class of associative image processing operators by developing an algebraic understanding of the operator and parallelization options. The algebra can transform associative operations such as box filters, summed area table, recursive filters etc. by splitting their domain to smaller subsets of the input image that can be executed in parallel and recombine the intermediate result later. The ultimate target is to develop a compiler front-end based on the Halide language that implements this algebra and is capable of parallelizing associative operators of arbitrary footprints by a few lines of code, thereby relieving the programmer of the tedious task for programming the parallelized algorithms. Such a compiler would allow programmers to easily experiment with a plethora of parallelization strategies in a systematic manner.

This work is in collaboration with Jonathan Ragan-Kelley and Fredo Durand of MIT and Sylvain Paris (Adobe Research).

### 5.1.5. *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. In this work we evaluate different user interface designs for lightfield editing. Our interfaces differ mainly in the way depth is presented to the user and build upon different depth perception cues.

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.

## 5.2. Perception for Plausible Rendering

### 5.2.1. *Perception of Perspective Distortions in Image-Based Rendering*

**Participants:** Peter Vangorp, Christian Richardt, Gaurav Chaurasia, George Drettakis.

Image-based rendering (IBR) creates realistic images by enriching simple geometries with photographs, for example by mapping the photograph of a building façade onto a plane. However, as soon as the viewer moves away from the correct viewpoint, the image in the retina becomes distorted, sometimes leading to gross misperceptions of the original geometry. Two hypotheses from vision science state how viewers perceive such image distortions, one claiming that they can compensate for them (and therefore perceive scene geometry reasonably correctly), and one claiming that they cannot compensate (and therefore can perceive rather significant distortions). We modified the latter hypothesis so that it extends to street-level IBR. We then conducted a rigorous experiment that measured the magnitude of perceptual distortions that occur with IBR for façade viewing. We also conducted a rating experiment that assessed the acceptability of the distortions. The results of the two experiments were consistent with one another. They showed that viewers' percepts are indeed distorted, but not as severely as predicted by the modified vision science hypothesis. From our experimental results, we develop a predictive model of distortion for street-level IBR, which we use to provide guidelines for acceptability of virtual views and for capture camera density. We perform a confirmatory study to validate our predictions, and illustrate their use with an application that guides users in IBR navigation to stay in regions where virtual views yield acceptable perceptual distortions (see Figure 5).

This work is a collaboration with Emily Cooper and Marty Banks at UC Berkeley, within the associate team CRISP. The paper was accepted as a SIGGRAPH 2013 paper and published in the ACM Transactions on Graphics journal [18].

### 5.2.2. *Gloss Perception in Painterly and Cartoon Rendering*

**Participant:** Adrien Bousseau.



*Figure 5. This interactive navigation tool shows an inset (in the lower left) that predicts comfort ratings for all possible camera orientations as seen from the blue camera's viewpoint. The application also restricts the user's motion to regions with acceptable predicted quality (in blue and yellow).*

Depictions with traditional media such as painting and drawing represent scene content in a stylized manner. It is unclear however how well stylized images depict scene properties like shape, material and lighting. In this project, we use non photorealistic rendering algorithms to evaluate how stylization alters the perception of gloss (see Figure 6). Our study reveals a compression of the range of representable gloss in stylized images so that shiny materials appear more diffuse in painterly rendering, while diffuse materials appear shinier in cartoon images.

From our measurements we estimate the function that maps realistic gloss parameters to their perception in a stylized rendering. This mapping allows users of NPR algorithms to predict the perception of gloss in their images. The inverse of this function exaggerates gloss properties to make the contrast between materials in a stylized image more faithful. We have conducted our experiment both in a lab and on a crowdsourcing website. While crowdsourcing allows us to quickly design our pilot study, a lab experiment provides more control on how subjects perform the task. We provide a detailed comparison of the results obtained with the two approaches and discuss their advantages and drawbacks for studies similar to ours.



*Figure 6. The experimental task used for studying gloss perception in stylized images.*

This work is a collaboration with James O'Shea, Ravi Ramamoorthi and Maneesh Agrawala from UC Berkeley in the context of the Associate Team CRISP (see also Section 1 ) and Frédo Durand from MIT. It has been published in ACM Transactions on Graphics 2013 [11] and presented at SIGGRAPH.

### 5.2.3. A High-Level Visual Attention Model

**Participant:** George Drettakis.

The goal of this project is to develop a high-level attention model based on memory schemas and singleton theory in visual perception. We have developed an approach extending a Bayesian approach to attention, which incorporates these high level features and can be directly used in a game engine to improve scene design.

This project is in collaboration with the Tech. University of Crete in the context of the Ph.D. of George Koulieris, supervised by Prof. Katerina Mania and BTU Cottburg (D. Cunningham).

## 5.3. Interaction and Design for Virtual Environments

### 5.3.1. Diffusion Curves: A Vector Representation for Smooth-Shaded Images

**Participant:** Adrien Bousseau.

This paper was selected for presentation in the Communications of the ACM, as an important graphics research result of interest to the entire Computer Science community. We describe a new vector-based primitive for creating smooth-shaded images, called the diffusion curve. A diffusion curve partitions the space through which it is drawn, defining different colors on either side. These colors may vary smoothly along the curve. In addition, the sharpness of the color transition from one side of the curve to the other can be controlled. Given a set of diffusion curves, the final image is constructed by solving a Poisson equation whose constraints are specified by the set of gradients across all diffusion curves (Figure 7 ). Like all vector-based primitives, diffusion curves conveniently support a variety of operations, including geometry-based editing, keyframe animation, and ready stylization. Moreover, their representation is compact and inherently resolution independent. We describe a GPU-based implementation for rendering images defined by a set of diffusion curves in real time. We then demonstrate an interactive drawing system allowing artists to create artwork using diffusion curves, either by drawing the curves in a freehand style, or by tracing existing imagery. Furthermore, we describe a completely automatic conversion process for taking an image and turning it into a set of diffusion curves that closely approximate the original image content.

This work is a collaboration with Alexandrina Orzan, Pascal Barla (Inria / Manao), Holger Winnemöller (Adobe Systems), Joëlle Thollot (Inria / Maverick) and David Salesin (Adobe Systems). This work was originally published in ACM Transactions on Graphics (Proceeding of SIGGRAPH 2008) and was selected for publication in Communications of the ACM July 2013 [15].

### 5.3.2. Natural Gesture-based Interaction for Complex Tasks in an Immersive Cube

**Participants:** Emmanuelle Chapoulie, George Drettakis.

We present a solution for natural gesture interaction in an immersive cube in which users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks. Our solution uses finger tracking coupled with a real-time physics engine, combined with a comprehensive approach for hand gestures, which is robust to tracker noise and simulation instabilities. To determine if our natural gestures are a feasible interface in an immersive cube, we perform an exploratory study for tasks involving the user walking in the cube while performing complex manipulations such as balancing objects. We compare gestures to a traditional 6-DOF Wand, and we also compare both gestures and Wand with the same task, faithfully reproduced in the real world. Users are also asked to perform a free task, allowing us to observe their perceived level of presence in the scene. Our results show that our robust approach provides a feasible natural gesture interface for immersive cube-like environments and is perceived by users as being closer to the real experience compared to the Wand.



*Figure 7. A diffusion curve consists of a Bézier curve (a) enriched with color (b) and blur (c) control points. The final image (d) is obtained by diffusing the colors in the image domain.*

This work is a collaboration with Jean-Christophe Lombardo of SED, with Evanthia Dimara and Maria Roussou from the University of Athens and with Maud Marchal from IRISA-INSA/Inria Rennes - Bretagne Atlantique. The work is under review in the journal Virtual Reality.

### 5.3.3. *Evaluation of Direct Manipulation in an Immersive Cube: a Controlled Study*

**Participants:** Emmanuelle Chapoulie, George Drettakis.

We are pursuing a study for interaction using finger tracking and traditional 6 degrees of freedom (DOF) flysticks in a virtual reality immersive cube. Our study aims at identifying which factors make one interface better than the other and which are the tradeoffs for the design of experiments, thus decomposing the movements into restricted DOF.

### 5.3.4. *The Drawing Assistant: Automated Drawing Guidance and Feedback from Photographs*

**Participants:** Emmanuel Iarussi, Adrien Bousseau.

Drawing is the earliest form of visual depiction and continues to enjoy great popularity with paint systems. However, drawing requires artistic skills that many people feel out of reach. We developed an interactive drawing tool that provides automated guidance over model photographs to help people practice traditional drawing-by-observation techniques. The drawing literature describes a number of techniques to help people gain consciousness of the shapes in a scene and their relationships. We compile these techniques and derive a set of construction lines that we automatically extract from a model photograph (see Figure 8). We then display these lines over the model to guide its manual reproduction by the user on the drawing canvas. Our pen-based interface also allows users to navigate between the techniques they wish to practice and to draw construction lines in dedicated layers. We use shape-matching to register the user's sketch with the model guides. We use this registration to provide corrective feedback to the user. We conducted two user studies to inform the design of our tool and evaluate our approach with a total of 20 users. Participants produced better drawings using the drawing assistant, with more accurate proportions and alignments. They also perceived that guidance and corrective feedback helped them better understand how to draw. Finally, some participants spontaneously applied the techniques when asked to draw without our tool after using it for about 30 minutes.

This work is a collaboration with Theophanis Tsandilas from the InSitu project team - Inria Saclay, in the context of the ANR DRAO project. It has been published at proceedings of UIST 2013 the 26th annual ACM symposium on User interface software and technology [19].

### 5.3.5. *Shape-Aware Sketch Editing with Covariant-Minimizing Cross Fields*

**Participants:** Emmanuel Iarussi, Adrien Bousseau.

Free-hand sketches are extensively used in product design for their ability to convey 3D surfaces with a handful of pen strokes. Skillful artists capture all surface information by strategically positioning strokes so that they depict the feature lines and curvature directions of surface patches. Viewers envision the intended 3D surface by mentally interpolating these lines to form a dense network representative of the curvature of the shape. Our goal is to mimic this interpolation process to estimate at each pixel of a sketch the projection of the two principal directions of the surface, or their extrapolation over umbilic regions. While the information we recover is purely 2D, it provides a vivid sense of the intended 3D surface and allows various shape-aware sketch editing applications, including normal estimation for shading, cross-hatching rendering and surface parameterization for texture mapping.

This work is a collaboration with David Bommès from the Titane project team, Sophia-Antipolis.

### 5.3.6. *Depicting Stylized Materials with Vector Shade Trees*

**Participants:** Jorge Lopez-Moreno, Stefan Popov, Adrien Bousseau, George Drettakis.



*Figure 8. Our drawing assistant provides guidance and feedback over a model photograph that the user reproduces on a virtual canvas (a). We use computer vision algorithms to extract visual guides that enhance the geometric structures in the image (b). In this example, the user first sketched the block-in construction lines (c, blue) before drawing the regions and adding details. This guidance helps users produce more accurate drawings.*



*Figure 9. Depicting Stylized Materials with Vector Shade Trees.*



Vector graphics represent images with compact, editable and scalable primitives. Skillful vector artists employ these primitives to produce vivid depictions of material appearance and lighting. However, such stylized imagery often requires building complex multi-layered combinations of colored fills and gradient meshes. We facilitate this task by introducing vector shade trees that bring to vector graphics the flexibility of modular shading representations as known in the 3D rendering community. In contrast to traditional shade trees that combine pixel and vertex shaders, our shade nodes encapsulate the creation and blending of vector primitives that vector artists routinely use. We propose a set of basic shade nodes that we design to respect the traditional guidelines on material depiction described in drawing books and tutorials. We integrate our representation as an Adobe Illustrator plug-in that allows even inexperienced users to take a line drawing, apply a few clicks and obtain a fully colored illustration. More experienced artists can easily refine the illustration, adding more details and visual features, while using all the vector drawing tools they are already familiar with. We demonstrate the power of our representation by quickly generating illustrations of complex objects and materials.

Figure 9 illustrates how our algorithm works. We use a combination of basic shade nodes composed of vector graphics primitives to describe Vector Shade Trees that represent stylized materials (a). Combining these nodes allows the depiction of a variety of materials while preserving traditional vector drawing style and practice. We integrate our vector shade trees in a vector drawing tool that allows users to apply stylized shading effects on vector line drawings (b,c).

This work is a collaboration with Maneesh Agrawala from University of California, Berkeley in the context of the CRISP Associated Team. The work was accepted as a SIGGRAPH 2013 paper and published in ACM Transactions on Graphics, volume 32, issue 4 [14].

### 5.3.7. Auditory-Visual Aversive Stimuli Modulate the Conscious Experience of Fear

**Participants:** Rachid Guerchouche, George Drettakis.

In a natural environment, affective information is perceived via multiple senses, mostly audition and vision. However, the impact of multisensory information on affect remains relatively undiscovered. In this study, we investigated whether the auditory-visual presentation of aversive stimuli influences the experience of fear. We used the advantages of virtual reality to manipulate multisensory presentation and to display potentially fearful dog stimuli embedded in a natural context. We manipulated the affective reactions evoked by the dog stimuli by recruiting two groups of participants: dog-fearful and non-fearful participants. The sensitivity to dog fear was assessed psychometrically by a questionnaire and also at behavioral and subjective levels using a Behavioral Avoidance Test (BAT). Participants navigated in virtual environments, in which they encountered virtual dog stimuli presented through the auditory channel, the visual channel or both. They were asked to report their fear using Subjective Units of Distress. We compared the fear for unimodal (visual or auditory) and bimodal (auditory-visual) dog stimuli. Dog-fearful participants as well as non-fearful participants reported more fear in response to bimodal audiovisual compared to unimodal presentation of dog stimuli. These results suggest that fear is more intense when the affective information is processed via multiple sensory pathways, which might be due to a cross-modal potentiation. Our findings have implications for the field of virtual reality-based therapy of phobias. Therapies could be refined and improved by implicating and manipulating the multisensory presentation of the feared situations.

This work is a collaboration with Marine Taffou and Isabelle Viaud-Delmon from CNRS-IRCAM, in the context of the European project VERVE. The work was published in the Multisensory Research journal 2013 [17].

### 5.3.8. Memory Motivation Virtual Experience

**Participants:** Emmanuelle Chapoulie, Rachid Guerchouche, George Drettakis.

Memory complaints are known to be one the first stages of Alzheimer's disease, for which -up to now, there is no known chemical treatment. In the context of the European project VERVE, and in collaboration with the Resources and Research Memory Centre of Nice Hospital (CM2R), we performed a study on the feasibility of treating memory complaints using realistic immersive virtual environments. Such environments are created



Figure 10. Pictures of the auditory-visual VEs used to measure the participants' fear when encountering virtual dogs. On the left, the outdoor garden scene and on the right, the indoor hangar scene.

using Image-Based Rendering technique developed by REVES. It is possible to easily provide, realistic 3D environments of places familiar to the participants using only a few photograph, and investigate whether IBR virtual environments can convey familiarity.

This work is a collaboration with Pierre-David Petit and Pr. Philippe Robert from CM2R. The work will be presented in IEEE Virtual Reality conference 2014 and will be published in the conference proceedings.

### 5.3.9. Layered Image Vectorization

**Participants:** Christian Richardt, Adrien Bousseau, George Drettakis.

Vector graphics enjoy great popularity among graphic designers for their compactness, scalability and editability. The goal of *vectorization* algorithms is to facilitate the creation of vector graphics by converting bitmap images into vector primitives. However, while a vectorization algorithm should faithfully reproduce the appearance of a bitmap image, it should also generate vector primitives that are easily editable – a goal that existing methods have largely overlooked. We investigate layered representations which are more compact and editable, and hence better preserve the strengths of vector graphics. This work is in collaboration with Maneesh Agrawala in the context of the CRISP Associated Team and Jorge Lopez-Moreno, now a postdoc at the University of Madrid.

### 5.3.10. True2Form: Automatic 3D Concept Modeling from Design Sketches

**Participants:** Adrien Bousseau.

We developed a method to estimate smooth 3D shapes from design sketches. We do this by hypothesizing and perceptually validating a set of local geometric relationships between the curves in sketches. We then algorithmically reconstruct 3D curves from a single sketch by detecting their local geometric relationships and reconciling them globally across the 3D curve network.

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.

## TITANE Team

# 6. New Results

## 6.1. Analysis

### 6.1.1. Detecting Parametric Objects in Large Scenes by Monte Carlo Sampling

**Participants:** Yannick Verdie, Florent Lafarge.

Point processes constitute 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 performance of 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 [11]. 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 as a function of spatial information extracted from the input data. The performance of the sampler is analyzed through a set of experiments on various object detection problems from large scenes, including comparisons to the existing algorithms. The sampler is also evaluated as an optimization algorithm for MRF-based labeling problems (Figure 1).



Figure 1. Detection and modeling of trees from large-scale Lidar scans by our parallel Monte Carlo sampler.

### 6.1.2. Recovering Line-networks in Images by Junction-Point Processes

**Participant:** Florent Lafarge.

*In collaboration with Dengfeng Chai (Zhejiang University, China) and Wolfgang Forstner (University of Bonn, Germany).*

We tackle the automatic extraction of line-networks from images. Appearance and shape considerations have been deeply explored in the literature to improve accuracy in presence of occlusions, shadows, and a wide variety of irrelevant objects. However most existing work has ignored the structural aspect of the problem. We present an original method which provides structurally-coherent solutions [13]. Contrary to the pixel-based and object-based methods, our result is a graph in which each node represents either a connection or an ending in the line-network. Based on stochastic geometry, we develop a new family of point processes consisting in sampling junction-points in the input image by using a Monte Carlo mechanism. The quality of a configuration is measured by a probability density which takes into account both image consistency and shape priors. Our experiments on a variety of problems illustrate the potential of our approach in terms of accuracy, flexibility and efficiency (Figure 2).



Figure 2. Line-network extraction from images using a junction-point process.

## 6.2. Approximation

### 6.2.1. Integer-Grid Maps for Reliable Quad Meshing

**Participants:** David Bommès, Pierre Alliez.

*In collaboration with Leif Kobbelt from RWTH Aachen.*

Quadrilateral remeshing approaches based on global parametrization enable many desirable mesh properties. Two of those are (1) high regularity due to explicit control over irregular vertices and (2) smooth distribution of distortion achieved by convex variational formulations. In this work [2] we propose a novel convex Mixed-Integer Quadratic Programming (MIQP) formulation which ensures by construction that the resulting map is within the class of so called Integer-Grid Maps that are guaranteed to imply a quad mesh. In order to overcome the NP-hardness of MIQP we propose two additional optimizations: a complexity reduction algorithm and singularity separating conditions. While the former decouples the dimension of the MIQP search space from the input complexity of the triangle mesh, the latter improves the continuous relaxation, which is crucial for the success of modern MIQP optimizers. Our algorithm also enables the global search for high-quality coarse

quad layouts as illustrated in Figure 3, a difficult task solely tackled by insufficient greedy methodologies before.



Figure 3. Integer-grid maps for reliable quad meshing.

### 6.2.2. QEx: Robust Quad Mesh Extraction

**Participant:** David Bommes.

*In collaboration with Leif Kobbelt from RWTH Aachen.*

Among the class of quad remeshing techniques, the ones based on parameterization strive to generate an integer-grid map, i.e., a parametrization of the input surface in 2D such that the canonical grid of integer iso-lines forms a quad mesh when mapped back onto the surface in 3D. An essential, albeit broadly neglected aspect of these methods is the quad extraction step. This step is not a trivial matter: ambiguities induced by numerical inaccuracies and limited solver precision, as well as imperfections in the maps produced by most methods (unless costly countermeasures are taken) pose significant hurdles to the quad extractor. In this work [6] we present a method to sanitize a provided parametrization such that it becomes numerically consistent even with limited precision floating point arithmetic. We also devise a novel strategy to cope with common local fold-overs in the parametrization. We can generate all-quadrilateral meshes where otherwise holes, non-quad polygons or no output at all would have been produced like for the example in Figure 4.

### 6.2.3. Advanced Automatic Hexahedral Mesh Generation from Surface Quad Meshes

**Participant:** David Bommes.

*In collaboration with Leif Kobbelt (RWTH Aachen).*

A purely topological approach for the generation of hexahedral meshes from quadrilateral surface meshes of genus zero has been proposed by M. Müller-Hannemann: in a first stage, the input surface mesh is reduced to a single hexahedron by successively eliminating loops from the dual graph of the quad mesh; in the second stage, the hexahedral mesh is constructed by extruding a layer of hexahedra for each dual loop from the first stage in reverse elimination order. We introduce several techniques to extend the scope of target shapes of the



Figure 4. QEx: Robust Quad Mesh Extraction.

approach and significantly improve the quality of the generated hexahedral meshes [14]. While the original method can only handle almost-convex objects and requires mesh surgery and remeshing in case of concave geometry, we propose a method to overcome this issue by introducing the notion of concave dual loops in order to handle non-convex objects like the one displayed in Figure 5. Furthermore, we analyze and improve the heuristic to determine the elimination order for the dual loops such that the inordinate introduction of interior singular edges, i.e., edges of degree other than four in the hexahedral mesh, can be avoided in many cases.

#### 6.2.4. On the Equilibrium of Simplicial Masonry Structures

**Participant:** Pierre Alliez.

*In collaboration with Mathieu Desbrun, Fernando de Goes and Houman Owhadi from Caltech.*

We contributed a novel approach for the analysis and design of self-supporting simplicial masonry structures [4]. A finite-dimensional formulation of their compressive stress field is derived, offering a new interpretation of thrust networks through numerical homogenization theory. We further leverage geometric properties of the resulting force diagram to identify a set of reduced coordinates characterizing the equilibrium of simplicial masonry. We finally derive computational form-finding tools that improve over previous work in efficiency, accuracy, and scalability.

### 6.3. Reconstruction

#### 6.3.1. Noise-Adaptive Shape Reconstruction from Raw Point Sets

**Participants:** Simon Giraudot, Pierre Alliez.

*In collaboration with David Cohen-Steiner (GEOMETRICA project-team)*



Figure 5. Advanced Automatic Hexahedral Mesh Generation from Surface Quad Meshes.

We devised a noise-adaptive shape reconstruction method specialized to smooth, closed shapes [7]. Our algorithm takes as input a defect-laden point set with variable noise and outliers, and comprises three main steps. First, we compute a novel noise-adaptive distance function to the inferred shape, which relies on the assumption that the inferred shape is a smooth submanifold of known dimension. Second, we estimate the sign and confidence of the function at a set of seed points, through minimizing a quadratic energy expressed on the edges of a uniform random graph. Third, we compute a signed implicit function through a random walker approach with soft constraints chosen as the most confident seed points computed in the previous step.

### 6.3.2. Surface Reconstruction through Point Set Structuring

**Participants:** Florent Lafarge, Pierre Alliez.

We present a method for reconstructing surfaces from point sets [8]. The main novelty lies in a structure-preserving approach where the input point set is first consolidated by structuring and resampling the planar components, before reconstructing the surface from both the consolidated components and the unstructured points. Structuring facilitates the surface reconstruction as the point set is substantially reduced and the points are enriched with structural meaning related to adjacency between primitives. Our approach departs from the common dichotomy between smooth/piecewise-smooth and primitive-based representations by gracefully combining canonical parts from detected primitives and free-form parts of the inferred shape (Figure 6).

### 6.3.3. Hybrid Multi-view Stereo for Modeling Urban Scenes

**Participant:** Florent Lafarge.

*In collaboration with Renaud Keriven (Acute3D), Mathieu Bredif (IGN), and Hiep Vu (Ecole des Ponts ParisTech).*

We present an original multi-view stereo reconstruction algorithm which allows the 3D-modeling of urban scenes as a combination of meshes and geometric primitives [9]. The method provides a compact model while preserving details: irregular elements are described by meshes whereas regular structures are described by



Figure 6. Structuring of a church.

canonical geometric primitives. We adopt a two-step strategy consisting first in segmenting the initial mesh-based surface using a multi-label Markov Random Field based model and second, in sampling primitive and mesh components simultaneously on the obtained partition by a Jump-Diffusion process. The quality of a reconstruction is measured by a multi-object energy model which takes into account both photo-consistency and semantic considerations (i.e. geometry and shape layout). The segmentation and sampling steps are embedded into an iterative refinement procedure which provides an increasingly accurate hybrid representation (Figure 7).

#### 6.3.4. Indoor Scene Reconstruction using Primitive-driven Space Partitioning and Graph-cut

**Participants:** Sven Oesau, Florent Lafarge, Pierre Alliez.

*In collaboration with EADS ASTRIUM*

We present 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 [15]. 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 experiments produces watertight reconstructed models of multi-level buildings and complex scenes (Figure 8).

#### 6.3.5. Watertight Scenes from Urban LiDAR and Planar Surfaces

**Participant:** Thijs Van Lankveld.

*In collaboration with Marc Van Kreveld and Remco Veltkamp*





*Figure 7. Reconstruction of a facade from Multi-View Stereo images using a multi-shape hybrid sampler.*



*Figure 8. Reconstruction of a multi-floor indoor scene from an input point cloud.*

The demand for large geometric models is increasing, especially of urban environments. This has resulted in production of massive point cloud data from images or LiDAR. Visualization and further processing generally require a detailed, yet concise representation of the scene's surfaces. Related work generally either approximates the data with the risk of over-smoothing, or interpolates the data with excessive detail. Many surfaces in urban scenes can be modeled more concisely by planar approximations. We present a method that combines these polygons into a watertight model [10]. The polygon-based shape is closed with free-form meshes based on visibility information. To achieve this, we divide 3-space into inside and outside volumes by combining a constrained Delaunay tetrahedralization with a graph-cut. We compare our method with related work on several large urban LiDAR data sets. We construct similar shapes with a third fewer triangles to model the scenes. Additionally, our results are more visually pleasing and closer to a human modeler's description of urban scenes using simple boxes (Figure 10).



Figure 9. Reconstruction of watertight model from LIDAR data sets.

### 6.3.6. Feature-Preserving Surface Reconstruction and Simplification from Defect-Laden Point Sets

**Participant:** Pierre Alliez.

*In collaboration with David Cohen-Steiner, Julie Digne, Mathieu Desbrun and Fernando de Goes*

We introduce a robust and feature-capturing surface reconstruction and simplification method that turns an input point set into a low triangle-count simplicial complex [5]. Our approach starts with a (possibly non-manifold) simplicial complex filtered from a 3D Delaunay triangulation of the input points. This initial approximation is iteratively simplified based on an error metric that measures, through optimal transport, the distance between the input points and the current simplicial complex, both seen as mass distributions. Our approach exhibits both robustness to noise and outliers, as well as preservation of sharp features and boundaries (Figure 10). Our new feature-sensitive metric between point sets and triangle meshes can also be used as a post-processing tool that, from the smooth output of a reconstruction method, recovers sharp features and boundaries present in the initial point set.



Figure 10. Robust reconstruction through optimal transportation.

### 6.3.7. Splat-based Surface Reconstruction from Defect-Laden Point Sets.

**Participant:** Pierre Alliez.

*In collaboration with Mariette Yvinec (EPI GEOMETRICA), Ricard Campos (University of Girona), Raphael Garcia (University of Girona)*

We introduce a method for surface reconstruction from point sets that is able to cope with noise and outliers. First, a splat-based representation is computed from the point set. A robust local 3D RANSAC-based procedure is used to filter the point set for outliers, then a local jet surface – a low-degree surface approximation – is fitted to the inliers. Second, we extract the reconstructed surface in the form of a surface triangle mesh through Delaunay refinement (Figure 11 ). The Delaunay refinement meshing approach requires computing intersections between line segment queries and the surface to be meshed. In the present case, intersection queries are solved from the set of splats through a 1D RANSAC procedure. [3].



*Figure 11. Splat-based reconstruction.*

## ALPAGE Project-Team

# 6. New Results

## 6.1. Unsupervised segmentation of Mandarin Chinese

**Participants:** Pierre Magistry, Benoît Sagot.

In Chinese script, very few symbols can be considered as word boundary markers. The only easily identifiable boundaries are sentence beginnings and endings, as well as positions before and after punctuation marks. Although the script doesn't rely on typography to define (orthographic) "words", a word-level segmentation is often required for further natural language processing, which is a highly non-trivial task.

A great variety of methods have been proposed in the literature, mostly in supervised machine learning settings. Our work addresses the question of unsupervised segmentation, i.e., without any manually segmented training data. Although supervised learning typically performs better than unsupervised learning, we believe that unsupervised systems are worth investigating as they require less human labour and are likely to be more easily adaptable to various genres, domains and time periods. They can also provide more valuable insight for linguistic studies.

Amongst the unsupervised segmentation systems described in the literature, two paradigms are often used: Branching Entropy (BE) and Minimum Description Length (MDL). The system we have developed relies on both. We have introduced a new algorithm [22] which searches in a larger hypothesis space using the MDL criterion, thus leading to lower Description Lengths than other previously published systems. Still, this improvement concerning the Description Length does not come with better results on the Chinese word segmentation task, which raises interesting issues. However, it turns out that it is possible to add very simple constraints to our algorithm in order to adapt it to the specificities of Mandarin Chinese in a way that leads to results better than the state-of-the-art on the Chinese word segmentation task.

Moreover, an important part of discrepancies between the various segmentation guidelines concerns the so-called "factoids." This term covers a variety of language phenomena that include: numbers, dates, addresses, email addresses, proper names, and others. We have shown that specific treatment of a subset of such expressions is both sound (as factoids to not resort to general language, which we try and capture with our segmentation model, both rather to conventions that are easy to encode as rules). By augmenting the local grammars of SxPipe to deal with the aforementioned expressions in Chinese, and use them as a pre-processing for our task, we can discard the matched expressions from the training data and segment them accordingly to the guidelines as a post-processing step. Our results show a significant improvement over previous results.

## 6.2. Dynamic extension of a French lexical resources based a text stream

**Participants:** Damien Nouvel, Benoît Sagot, Rosa Stern, Virginie Mouilleron, Marion Baranes.

Lexical incompleteness is a recurring problem when dealing with natural language and its variability. It seems indeed necessary today to regularly validate and extend lexica used by tools processing large amounts of textual data. This is even more true when processing real-time text flows. In this context, we have introduced two series of techniques for addressing words unknown to lexical resources, and applied them to French within the context of the EDyLex ANR project:

- **Extending a morphological lexicon** We have studied neology (from a theoretic and corpus-based point of view) and developed modules for detecting neologisms in AFP news wires in real time and inferring information about them (lemma, category, inflectional class) [24]. We have shown that we are able, using among others modules for analyzing derived and compound neologisms, to generate lexical entries candidates in real time and with a good precision, to be added in the *Lefff* lexicon.
- **Extending an entity database** We have also extended our previous work on named entities detection and linking in order to be able to extract new named entities from AFP news wires and create candidate entries for the *Aleda* entity database.

### 6.3. Transferring lexical knowledge from a resourced language to a closely-related resource-free language

**Participants:** Yves Scherrer, Benoît Sagot.

We have developed a generic approach for the transfer of part-of-speech (POS) annotations from a resourced language (RL) towards an etymologically closely related non-resourced language (NRL), without using any bilingual (i.e., parallel) data. We rely on two hypotheses. First, on the lexical level, the two languages share a lot of cognates, i.e., word pairs that are formally similar and that are translations of each other. Second, on the structural level, we admit that the word order of both languages is similar, and that the set of POS tags is identical. Thus, we suppose that the POS tag of one word can be transferred to its translational equivalent in the other language.

The proposed approach consists of two main steps. In the first step, we induce a translation lexicon from monolingual corpora. This step relies on several methods, including a character-based statistical machine translation model to infer cognate pairs, and 3-gram and 4-gram contexts to infer additional word pairs on the basis of their contextual similarity. This step yields a list of  $\langle w_{\text{NRL}}, w_{\text{RL}} \rangle$  pairs. In the second step, the RL lexicon entries are annotated with POS tags with the help of an existing resource, and these annotations are transferred onto the corresponding NRL lexicon entries. We complete the resulting tag dictionary with heuristics based on suffix analogy. This results in a list of  $\langle w_{\text{NRL}}, t \rangle$  pairs, covering the whole NRL corpus.

We have evaluated our methods on several language pairs. We have worked among others on five language pairs of the Iberic peninsula, where Spanish and Portuguese play the role of RLs: Aragonese–Spanish, Asturian–Spanish, Catalan–Spanish, Galician–Spanish and Galician–Portuguese [27]. We have also conducted experiments on Germanic [28] and Slavic languages. We have also applied it in a slightly different context, in collaboration with Tomaž Erjavec (IJS, Slovenia), namely that of inducing resources for historical Slovene based on existing resources for contemporary Slovene [26]. Although no direct comparison can be performed, because of the novelty of the task, our results are very satisfying in so far that they are almost as high as published result on a related but simpler task, that of unsupervised part-of-speech tagging — which, contrarily to our work, relies on an existing morphological lexicon for the language at hand.

### 6.4. Building a large-scale translation graph

**Participants:** Valérie Hanoka, Benoît Sagot.

Large-scale general-purpose multilingual translation databases are useful in a wide range of Natural Languages Processing (NLP) tasks. This is especially true concerning researches tackling problems specific to under-resourced languages, as translation databases can be used for adapting existing resources in other languages. This has been applied for example for the development of wordnets in languages other than English. There is thus a real need in NLP for *open-source* multilingual lexical databases that compile as many translations as can be found on any freely available resource in any language.

We have developed, and are about to release, a new open-source heavily multilingual (over 590 languages) translation database built using several sources, namely various wiktionaries and the OPUS parallel corpora.

Our graph was built in several steps. We first extracted a preliminary set of translation and synonym pairs, which we stored in a large translation and synonym graph. We then applied filtering techniques for increasing the accuracy of this graph. We have evaluated the accuracy of our graph as being as high as 98% for translations extracted from wiktionaries.

### 6.5. Computational morphology

**Participant:** Benoît Sagot.

In 2013, following previous collaborative work [92], [105], we have designed and developed Alexina<sub>PARSLI</sub> in collaboration with Géraldine Walther (LLF and DDL), a formalism for encoding inflectional descriptions (lexicon and grammar) that aims at filling the gap between morphologically and typologically motivated approaches on the one hand and implemented approaches on the other hand, as will be discussed in the remainder of this section. Indeed, Alexina<sub>PARSLI</sub> is both:

- an **implementation formalism for PARSLI**, a formal model of inflectional morphology [106] that accounts for concepts underlying the canonical approach of morphological typology;
- an **extension of the Alexina lexical framework** developed at Alpage for modeling lexical information and developing lexical resources. The Alexina framework now supports both morphological grammars that use the original Alexina morphological formalism as well as new grammars developed in Alexina<sub>PARSLI</sub>.

The Alexina<sub>PARSLI</sub> formalism and tools have been proven greatly beneficial to works both in descriptive and formal morphology, in particular in studies about Latin passivisation and Maltese verbal inflection [106] and in studies comparing the compacity of morphological descriptions [106], [92], [105], as well as in NLP, for the efficient development of a large-scale and linguistically sound morphological lexicon for German (a paper describing this new lexicon is to be presented at the LREC 2014 conference).

In collaboration with Géraldine Walther and Guillaume Jacques (CRLAO, CNRS), within operation LR4.11 from strand 6 of the LabEx EFL, we have also developed two Alexina<sub>PARSLI</sub> descriptions of (part of) the Khaling (Kiranti, Sino-Tibetan) verbal inflectional system, together with a medium-scale lexicon. Our study shows that an explicit account for the so-called direct-inverse marking, based on concepts developed within PARSLI, allows for a more compact account of this inflectional system [42].

## 6.6. Extracting Derivational Relations from an Inflectional Lexicon

**Participants:** Marion Baranes, Benoît Sagot.

Derivational morphological can provide useful information for natural language processing tasks. Indeed, it can improve any application which has to deal with unknown words such as information extraction, spell-checking and others.

We define a morphological family as a set of semantically related lexical entries which differ by their prefix and/or suffix, thus limiting ourselves to concatenative derivational morphology. We shall denote as derivationally related two morphological lexical entries that belong to the same morphological family.

We have developed a system which performs an analogy-based unsupervised extraction of weighted transformation rules that relate derivationally related lexical entries, and use these rules for extracting derivational relations within an existing inflectional lexicon. Our transformation rules can also be used to infer morphological information (both inflectional and derivational) for wordforms unknown to the inflectional lexicon. Our system is language-independent, although restricted to concatenative derivational morphology. We have evaluated it on four languages: English, French, German and Spanish. Our results will be published at the LREC 2014 conference.

## 6.7. Improving post-OCR correction with shallow linguistic processing

**Participants:** Kata Gábor, Benoît Sagot.

Providing wider access to national cultural heritage by massive digitalization confronts the actors of the field to a set of new challenges. 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. While this quality may be adequate for indexing, documents intended for reading need to meet higher standards. A reduction of the error rate by a factor of 10 to 100 becomes necessary for the diffusion of digitalized books and journals through emerging technologies such as e-books.

Within the PACTE project, an “Investissements d’avenir” project led by the Numen company, we have worked on the automatic post-processing of digitalized documents in the aim of reducing the OCR error rate by using contextual information and linguistic processing, by and large absent from current OCR engines. At the current stage of the project, we are focusing on French texts coming from the archives of the French National Library (Bibliothèque Nationale de France).

We adopted a hybrid approach, making use of both statistical classification techniques and linguistically motivated modules to detect OCR errors and generate correction candidates. The technology is based on the noisy channel model, widely used in the field of machine translation and spelling correction and subsequently in OCR post-correction. As to linguistically enhanced models, POS tagging was successfully applied to spelling correction. However, to our knowledge, little work has been done to exploit linguistic analysis for post-OCR correction.

We have proposed to integrate a shallow processing module to detect certain types of named entities, and a POS tagger trained specifically to deal with NE-tagged input. Our studies demonstrate that linguistically informed processing can efficiently contribute to reduce the error rate by 1) detecting false corrections proposed by the statistical correction module, 2) detecting a certain amount of OCR errors not detected by the statistical correction module.

## 6.8. Named Entity Linking

**Participants:** Rosa Stern, Benoît Sagot.

The Ph.D. research work started in 2009 lead in 2013 to the development of a joint entity recognition and linking system for the processing of textual data at the Agence France Presse (AFP).

This system, Nomos, allows to use any existing named entity recognition system, as well as combinations of such systems; their results are passed to a linking module in charge of the association between each detected mention and a unique reference within an existing data inventory. The two tasks (recognition and linking) are jointly operated: the recognition module presents a set of possible detections, which are further disambiguated by the linking module concurrently to the search for the best linking solution to each mention. This joint approach is justified by the need to limit the error propagation between two such modules in a pipeline system.

Experiments were achieved in order to evaluate the performance of Nomos over AFP news wires. They showed that the joint approach, relatively to a purely sequential one, improves the system’s global precision, i.e. the linking accuracy as well as the named entity recognition task itself. A gain of 3 points (87,6) is observed for the recognition precision with a low recall loss, while a gain of 8 points (92,9) is observed when several recognition systems are combined - although with a more significant loss of recall.

The Nomos system also allows to anchor of the AFP’s textual production in the Linked Data network and the Semantic Web paradigm, since the annotations derived from the entity linking associate each entity to an identified resource in repositories such as Wikipedia, DBpedia, Geonames or the New York Times Linked Data.

## 6.9. Treebanking at Alpage

**Participants:** Djamé Seddah, Benoît Sagot, Marie-Hélène Candito, Corentin Ribeyre, Benoît Crabbé, Éric Villemonte de La Clergerie, Virginie Moulleron, Vanessa Combet.

Since the advents of supervised methods for building accurate statistical parsing models, treebank engineering has become of crucial importance. In fact building a treebank, namely a set of carefully annotated syntactic parses with possibly different annotation layers and covering potentially different text domains, can be seen as providing a parser with both a grammar and a set of probabilities used for disambiguation. The main problem of such approaches lies in the nature of the lexical probabilities: they force the parsing model to be extremely sensitive to its training data and hence limit its performance to some low upper-bound when applied in out-of-domain scenario.



### 6.9.1. Written French Treebanks

Originating from the merging of two NLP teams specialized in grammar engineering and in which the creation of the first treebank for French was initiated [46], it is no wonder that we decided to increase the coverage of our French Treebank-based parsers by building out-of-domain treebanks: the Sequoia Corpus, [55], [18], made from Europarl, biomedical and wikipedia data, and the French Social Media Bank (outside English, the first data set covering Facebook, Twitter and other social media noisy text data) [95], [96]. We built those two corpus for two purposes: first, we wanted to evaluate the performance of our nlp chains (tokenization, tagging, parsing) on out-of-domain data, being noisy or not ; then we increased the coverage of our French treebank based models by simply adding those new data set to the canonical training set (using of-course many lexical variation, morphological clustering, brown clustering, etc.). We're also on the process of finalizing a new 2600 sentence data set, made essentially of questions, which are strikingly absent from all the treebanks we've been using and developing. So far, only one such data set exist and only for English: the Question-Bank [66]. Our very preliminary results show that simply adding a third of that corpus to the French Treebank greatly improve our parser performance.

Finally, Alpage is leading, in collaboration with the Nancy-based team Calligrame, a project to annotate the Sequoia corpus and the French Treebank with a richer, "deeper" syntactic layer, at the interface between syntax and semantics. A paper describing this effort is to appear at the LREC 2014 conference.

### 6.9.2. Spoken French Treebank

In collaboration with Anne Abeillé (LLF, CNRS), we have also contributed to the deign of a spoken treebank for French based on data produced in the ANR ETAPE. Contrary to other languages such as English, where spoken treebanks such as the Switchboard corpus treebank (Meteer, 1995), there is no sizable spoken corpus for French annotated for syntactic constituents and grammatical functions. Our project is to build such a resource which will be a natural extension of the Paris 7 treebank (Abeillé et al. 2003) for written French, in order to be able to compare with similar annotations written and spoken French. We have reused and adapted the parser (Petrov et al., 2006) which has been trained on the written treebank, with manual correction and validation. The first results are promising [32].

## 6.10. Linear time constituent parser

**Participant:** Benoît Crabbé.

We have designed an efficient and accurate lexicalized LR inspired discriminative parsing algorithm that recasts some current advances in dependency parsing to the constituency setting. We specifically designed and evaluated a Graph Structured Stack-based parser (Huang et al. 2010) using some additional specific approximate inference techniques such as the max violation update for the perceptron (Huang et al. 2012) . By contrast with dependency parsing however, lexicalized constituent parsing raises some additional correctness issues that motivate the explicit use of an LR automata instead of a simpler shift reduce framework.

The parsing model is linear in time and has been evaluated on French data, where it turns out to be state of the art on SPMRL 2013 datasets [29] both in time and in accuracy. The parsing framework has been designed to be further extended with compositional semantic representations and allows in principle an easy integration of ressources — such as those developped in the team — considered to be important for parsing morphologically rich languages.

## 6.11. Improving FRMG through partially supervised learning

**Participant:** Éric Villemonte de La Clergerie.

Since the emergence of several statistical parsers for French developed on the French TreeBank (FTB), including those developed at Alpage, it was important to be able to compare the symbolic meta-grammar-based parser FRMG with these statistical parsers on their native treebank, but also possibly to extend the comparison for other treebanks.

A first necessary step in this direction was a conversion from FRMG's native dependency scheme into FTB's dependency scheme, a tedious task highlighting the differences in design at all levels (segmentation, parts of speech, representation of the syntactic phenomena, etc.). A preliminary evaluation has shown that accuracy is good, but largely below the scores reached by the statistical parsers.

A challenge was then to explore if training on the FTB could be used to improve the accuracy of a symbolic parser like FRMG. However, the main difficulty arises from the fact that FTB's dependency scheme has little in common with FRMG's underlying grammar, and that no reverse conversion from FTB to FRMG structures is available. Such a conversion could be investigated but would surely be difficult to develop. Instead, we tried to exploit directly FTB data, using only very minimal assumptions, nevertheless leading to important gains and results close to those obtained by the statistical parsers [31]: it was possible to tune the disambiguation process of FRMG and strongly increase its accuracy, from 83% up to 87.17% (in terms of CONLL Labeled Attachment Score), a level comparable to those reached by statistical parsers trained on the FTB. Preliminary experiments show that (a) disambiguation tuning also improve the performances on other corpora and (b) that FRMG seems to be more stable than statistical parsers on corpora other than the FTB. Finer-grained comparison of FRMG wrt statistical parsers have been done that provide some insight for further improvements of FRMG.

The interest is that the technique should be easily adaptable for training data with different annotation schemes. Furthermore, our motivation was not just to improve the performances on the FTB and for the annotation scheme of FTB, for instance by training a reranker (as often done for domain adaptation), but to exploit the FTB to achieve global improvement over all kinds of corpora and for FRMG native annotation scheme.

## 6.12. Statistical parsing of Morphologically Rich Languages

**Participants:** Djamé Seddah, Marie-Hélène Candito, Éric Villemonte de La Clergerie, Benoît Sagot.

### 6.12.1. The SPMRL shared task

Since several years, Djamé Seddah, together with Marie-Hélène Candito and more generally the whole Alpage team, has played a major role in setting up and animating an international network of researchers focusing on parsing morphologically rich languages (MRLs).

In 2013, Djamé Seddah led the organization of the first shared task on parsing MRLs, hosted by the fourth SPMRL workshop and described in a 36-page overview paper that constitutes an in-depth state-of-the-art analysis and review of the domain [29]. 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. We compiled data for as many as 9 languages, which represents an immense scientific and technical challenge.

### 6.12.2. DyALog-SR

The SPMRL 2013 shared task was the opportunity to develop and test, with promising results, a simple beam-based shift-reduce dependency parser on top of the tabular logic programming system DYALOG. We used (Huang and Sagae, 2010) as the starting point for this work, in particular using the same simple arc-standard strategy for building projective dependency trees. The parser was also extended to handle ambiguous word lattices, with almost no loss w.r.t. disambiguated input, thanks to specific training, use of oracle segmentation, and large beams. We believe that this result is an interesting new one for shift-reduce parsing.

The current implementation scales correctly w.r.t. sentence length and, to a lesser extent, beam size. Nevertheless, for efficiency reasons, we plan to implement a simple C module for beam management to avoid the manipulation in DYALOG of sorted lists. Interestingly, such a module, plus the already implemented model manager, should also be usable to speed up the disambiguation process of DYALOG-based TAG parser FRMG (de La Clergerie, 2005a). Actually, these components could be integrated in a slow but on-going effort to add first-class probabilities (or weights) in DYALOG, following the ideas of (Eisner and Filardo, 2011) or (Sato, 2008).

### 6.12.3. The Alpage-LIGM French parser

The second Alpage system that participated to the SPMRL shared task, although on French language only, was developed in collaboration with Mathieu Constant (LIGM), based on the Bonsai architecture. This system is made of several single statistical dependency parsing systems whose outputs are combined into a reparser. We use two types of single parsing architecture: (a) pipeline systems; (b) “joint” systems.

The pipeline systems first perform multi-word expression (MWE) analysis before parsing. The MWE analyzer merges recognized MWEs into single tokens and the parser is then applied on the sentences with this new tokenization. The parsing model is learned on a gold training set where all marked MWEs have been merged into single tokens. For evaluation, the merged MWEs appearing in the resulting parses are expanded, so that the tokens are exactly the same in gold and predicted parses.

The “joint” systems directly output dependency trees whose structure comply with the French dataset annotation scheme. Such trees contain not only syntactic dependencies, but also the grouping of tokens into MWEs, since the first component of an MWE bears dependencies to the subsequent components of the MWE with a specific label. At that stage, the only missing information is the POS of the MWEs, which we predict by applying a MWE tagger in a post-processing step.

This parsing system obtains the best results for French, both for overall parsing and for MWE recognition, using a reparsing architecture that combines several parsers, with both pipeline architecture (MWE recognition followed by parsing), and joint architecture (MWE recognition performed by the parser).

## 6.13. Towards a French FrameNet

**Participants:** Marie-Hélène Candito, Marianne Djemaa, Benoît Sagot, Éric Villemonte de La Clergerie, Laurence Danlos.

The ASFALDA project <sup>1</sup> 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 [47], <sup>2</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*, FEs). The resulting resources will consist of :

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

In the first year of the project, we focused on the first of these objectives. A team of 10 active members, from Alpage, the Laboratoire de Linguistique Formelle (LLF), the MELODI team (IRIT - Toulouse) and the CEA-List partners achieved :

- the delimitation and adaptation to French of a set of FrameNet frames, in order to cover a set of specific notional domains (commercial transaction, communication, cognitive positions, judgment/evaluation, temporal relations, spatial position, causality).
- and the semi-automatic construction of a French lexicon in which French lexical units are associated with frames

The current resource contains 110 frames, and roughly 2500 lexical units / frame pairs. The next phase consists in automatic pre-annotation of semantic annotations, that will serve as basis for the manual validation phase.

Note that a publication describing the project and these first achievements shall be presented at the LREC 2014 conference.

<sup>1</sup><https://sites.google.com/site/anrasfalda/>

<sup>2</sup><https://framenet.icsi.berkeley.edu/>

## 6.14. Modelisation of discourse structures with DSTAG

**Participant:** Laurence Danlos.

This work was done within the ANR Polymnie, in collaboration with Sylvain Pogodalla and Philippe de Groot from LORIA.

Neg-Raising (NR) verbs form a class of verbs with a clausal complement that show the following behavior: when a negation syntactically attaches to the matrix predicate, it can semantically attach to the embedded predicate. Such an implication does not always hold. Some contexts make it impossible to consider the negation as having scope over the embedded predicate only. This corresponds to the non-NR reading of the predicate.

We have developed and published [20] an account of NR predicates within Tree Adjoining Grammars (TAG) that relies on a Montague-like semantics for TAG. The different properties of NR predicates are rendered at different levels: the ambiguity of the readings is modeled by lexical ambiguity; the scoping and cyclicity properties are modeled through the lexical semantics and the higher-order interpretation of adjunction nodes; spurious ambiguities are avoided using fine-grained types for terms representing derivation trees. This provides us with a base layer where to account for interactions with discourse connectives and discourse representation represented in DSTAG.

## 6.15. Annotation of discourse structures on the FTB

**Participants:** Laurence Danlos, Margot Colinet.

With the aim of annotating the French TreeBank (FTB, already annotated for syntax) with discourse information, we have been working on the first step of the project, namely identify all the occurrences of discourse connectives in the FTB. This raises problems for lexemes which are ambiguous with a discourse usage and other uses. In collaboration with Mathilde Dargnat (ATILF) and Grégoire Winterstein, we have been working on the preposition *pour* (around 1500 occurrences) and the adverb *alors* (300 occurrences). This work is the basis for a future annotation manual.

In parallel, we have been working on adverbial discourse connectives and published on the topic [17]. This paper focuses on the following question: does the only syntactic argument of an adverbial discourse connective correspond to its second semantic argument? It shows that this is not always the case, which is a problem for the syntax-semantics interface. This interface brings us to distinguish two classes of adverbial connectives we sketch the study of.

## 6.16. Pairwise coreference models

**Participant:** Emmanuel Lassalle.

In collaboration with Pascal Denis (Magnet, Inria), we have proposed a new method for significantly improving the performance of pairwise coreference models [34]. Given a set of indicators, our method learns how to best separate types of mention pairs into equivalence classes for which we construct distinct classification models. In effect, our approach finds an optimal feature space (derived from a base feature set and indicator set) for discriminating coreferential mention pairs. Although our approach explores a very large space of possible feature spaces, it remains tractable by exploiting the structure of the hierarchies built from the indicators.

In the framework of decision trees, this method can be seen as a pruning procedure and thus can be combined with different methods for expanding a decision tree. It can also be compared to polynomial kernels, but has the advantage of a lower computational complexity [21]. Our experiments on the CoNLL-2012 Shared Task English datasets (gold mentions) indicate that our method is robust relative to different clustering strategies and evaluation metrics, showing large and consistent improvements over a single pairwise model using the same base features. Our best system obtains a competitive 67.2 of average F1 over MUC, B3, and CEAF which, despite its simplicity, places it above the mean score of other systems on these datasets.

## 6.17. Identification of implicit discourse relations

**Participant:** Chloé Braud.

In collaboration with Pascal Denis (Magnet, Inria), we have developed a system for identifying “implicit” discourse relations (that is, relations that are not marked by a discourse connective) [33]. Given the little amount of available annotated data for this task, our system also resorts to additional automatically labeled data wherein unambiguous connectives have been suppressed and used as relation labels, a method introduced by Marcu and Echiabi (2002). As shown by Sporleder and Lascarides (2008) for English, this approach doesn’t generalize well to implicit relations as annotated by humans. We have shown that the same conclusion applies to French due to important distribution differences between the two types of data. In consequence, we propose various simple methods, all inspired from work on domain adaptation, with the aim of better combining annotated data and artificial data. We have evaluated these methods through various experiments carried out on the ANNODIS corpus: our best system reaches a labeling accuracy of 45.6%, corresponding to a 5.9% significant gain over a system solely trained on manually labeled data.

## PANAMA Project-Team

## 6. New Results

### 6.1. Recent results on sparse representations

Sparse approximation, high dimension, scalable algorithms, dictionary design, sample complexity

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the field of sparse representations, which is at the core of the ERC project PLEASE (projections, Learning and Sparsity for Efficient Data Processing, see section 8.2.1 ).

#### 6.1.1. A new framework for sparse representations: analysis sparse models

**Participants:** Rémi Gribonval, Nancy Bertin, Srdan Kitic, Cagdas Bilen.

*Main collaboration:* Mike Davies, Mehrdad Yaghoobi (Univ. Edinburgh), Michael Elad (The Technion).

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 [19]. 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 [100].

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 [39] and showed its efficiency in situations where usual methods fail [60]. It was also shown to be applicable to the hard declipping problem [61]. Application to EEG brain imaging was also investigated and a paper was submitted to ICASSP'14 (see below).

#### 6.1.2. Theoretical results on sparse representations

**Participants:** Rémi Gribonval, Anthony Bourrier, Pierre Machart.

*Main collaboration:* Charles Soussen (Centre de recherche en automatique de Nancy (CRAN)), Jérôme Idier (Institut de Recherche en Communications et en Cybernétique de Nantes (IRCCyN)), Cédric Herzet (Equipe-projet FLUMINANCE (Inria - CEMAGREF, Rennes)), Mehrdad Yaghoobi, Mike Davies (University of Edinburgh), Patrick Perez (Technicolor R&I France), Tomer Peleg (The Technion)

**Sparse recovery conditions for Orthogonal Least Squares :** We pursued our investigation of conditions on an overcomplete dictionary which guarantee that certain ideal sparse decompositions can be recovered by some specific optimization principles / algorithms. We extended Tropp's analysis of Orthogonal Matching Pursuit (OMP) using the Exact Recovery Condition (ERC) to a first exact recovery analysis of Orthogonal Least Squares (OLS). We showed that when ERC is met, OLS is guaranteed to exactly recover the unknown support. Moreover, we provided a closer look at the analysis of both OMP and OLS when ERC is not fulfilled. We showed that there exist dictionaries for which some subsets are never recovered with OMP. This phenomenon, which also appears with  $\ell^1$  minimization, does not occur for OLS. Finally, numerical experiments based on



our theoretical analysis showed that none of the considered algorithms is uniformly better than the other [21]. More recently, we obtained simpler coherence-based conditions [18] and pursued the analysis of unrecoverable subsets [43].

**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 relative 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 [42], [33]. Noiseless and noise-robust settings were both considered [56]. 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.

**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  $dP_X(x) \neq C \cdot \exp(-\phi(x))dx$ . This year, we extended this result to general inverse problems [30], [58], [47].

### 6.1.3. Algorithmic and theoretical results on dictionary learning

**Participants:** Rémi Gribonval, Nancy Bertin, Cagdas Bilen, Srdan Kitic.

*Main collaboration:* Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (Inria, Paris)), Martin Kleins-teuber, Matthias Seibert (TU-Munich), Mehrdad Yaghoobi, Mike Davies (University of Edinburgh),

**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 [86], [110] [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. Moreover, we established new sample complexity bounds of dictionary learning and other related matrix factorization schemes (including PCA, NMF, structured sparsity ...) [59].

**Analysis Operator Learning for Overcomplete Cospase Representations :** Besides standard dictionary learning, we also considered learning in the context of the cospase model. We consider the problem of learning a low-dimensional signal model from a collection of training samples. The mainstream approach would be 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.

We considered several approaches to learn an analysis operator from a training corpus [102]. For one of them, which uses a constrained optimization program based on  $\ell^1$  optimization, we derived a practical learning algorithm, based on projected subgradients, and demonstrated its ability to robustly recover a ground truth analysis operator, provided the training set is of sufficient size. A local optimality condition was derived, providing preliminary theoretical support for the well-posedness of the learning problem under appropriate conditions [24]. Extensions to deal with noisy data have been obtained as well [119].

In more specific situations, when prior information is available on the operator, it is also possible to express the operator on a parametric form, and learn this parameter. For instance, in the sound source localization problem, we showed that unknown speed of sound can be learned jointly in the process of cospase recovery, under mild conditions. This work was submitted to the iTwist'14 workshop.

## 6.2. Emerging activities on compressive sensing, learning and inverse problems

Compressive sensing, acoustic wavefields, audio inpainting,

### 6.2.1. Audio inpainting (SMALL FET-Open project)

**Participants:** Rémi Gribonval, Nancy Bertin, Corentin Guichaoua, Srdan Kitic.

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. 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.

METISS has initiated a series of works about audio inpainting, from its definition to methods to address it. This research has begun in the framework of the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL (Sparse Models, Algorithms and Learning for Large-Scale data) which began in January 2009. Rémi Gribonval was the coordinator of the project. The research on audio inpainting has been conducted by Valentin Emiya in 2010 and 2011.

The contributions consist of:

- defining audio inpainting as a general scheme where missing audio data must be estimated: it covers a number of existing audio processing tasks that have been addressed separately so far – click removal, declipping, packet loss concealment, unmasking in time-frequency;
- proposing algorithms based on sparse representations for audio inpainting (based on Matching Pursuit and on  $\ell^1$  minimization);
- addressing the case of audio declipping (*i.e.* desaturation): thanks to the flexibility of our inpainting algorithms, they can be constrained so as to include the structure of signals due to clipping in the objective to optimize. The resulting performance are significantly improved. This work appeared as a journal paper [63].
- addressing the case of audio declipping with the competitive cospase approach, with promising result especially when the clipping level is low. A contribution was submitted to the iTwist'14 workshop [61].



Current and future works deal with developing advanced 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).

### 6.2.2. *Blind Calibration of Compressive Sensing systems*

**Participants:** Rémi Gribonval, Cagdas Bilen.

*Main collaborations:* Gilles Chardon, Laurent Daudet (Institut Langevin), Gilles Puy (EPFL)

We consider the problem of calibrating a compressed sensing measurement system under the assumption that the decalibration consists in 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]. This year, we focused on extending the framework to phase-only decalibration, using techniques revolving around low-rank matrix recovery [27], [26], [34], [52], and to joint phase and gain decalibration [54].

### 6.2.3. *Compressive Gaussian Mixture estimation*

**Participants:** Rémi Gribonval, Anthony Bourrier.

*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 [38], [40], [29].

## 6.3. 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 [92]. It is noteworthy that such two matrices can be interpreted as both slices of a three-way array and their joint diagonalization can be viewed as Hitchcock's polyadic decomposition [89] of the associated three-way array. Other works followed discussing rank problems related to multi-way structures and properties of multi-way arrays. However, these 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 [115]. His model is now often referred to as the Tucker3 model. At the same moment, other authors focused on a particular case of the Tucker3 model, calling it PARAFAC for PARALLEL FACTOR analysis [88], and on the means to achieve such a decomposition, which will become the famous canonical decomposition [77]. 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 [94]. Consequently, a new class of efficient algorithms able to take into account the properties of tensors to be decomposed is needed.

### 6.3.1. A novel direct algorithm for CP decompositions

**Participant:** Laurent Albera.

*Main collaborations: Sepideh Hajipour (LTSI & BiSIPL), Isabelle Merlet (LTSI, France), Mohammad Bagher Shamsollahi (BiSIPL, Iran)*

Nowadays several techniques are available to solve the CP problem. They can be classified in three main groups [113]: alternating algorithms, which update only a subset of the parameters at each step; derivative-based methods, seeking for an update of all the parameters simultaneously by successive approximations; and direct procedures. The latter algorithms compute the CP decomposition by solving an alternative algebra problem of lower dimensions, but they do not provide a solution in terms of least squares contrarily to the alternating and derivative-based techniques.

We proposed a new direct algorithm to compute the CP decomposition of complex-valued multi-way arrays. The proposed algorithm is based on the Simultaneous Schur Decomposition (SSD) of particular matrices derived from the array to process. We also proposed a new Jacobi-like algorithm to calculate the SSD of several complex-valued matrices. Besides, we analysed our SSD and SSD-based CP techniques in terms of i) identifiability, ii) computational complexity and iii) estimation accuracy through a large number of scenarios including synthetic and real data in the context of CP decomposition. Computer results showed the efficiency of the proposed SSD-based CP method of dealing with some well-known difficult scenarios with swamp-like degeneracies. We also showed that the proposed method outperformed the classical CP algorithms in processing of Paatero multi-way arrays. Finally, the robustness of the proposed algorithm with respect to overfactoring was highlighted. This work was briefly presented at ICASSP'13 [31] while a journal paper for submission to IEEE Transactions on Signal Processing is in preparation.

### 6.3.2. CP decomposition of semi-symmetric semi-nonnegative three-way arrays

**Participant:** Laurent Albera.

*Main collaboration (line search methods): 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  $CG_{2steps}$ -like and LM-like algorithms.

This work was accepted for publication with minor revisions to the Elsevier Linear Algebra and Applications journal.

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 [109] and LM [113] algorithms without symmetry and nonnegativity constraints, and the ACDC algorithm [120] 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 submitted to the Elsevier Signal Processing journal.

### **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 subproblems using i) a square change of variables to impose nonnegativity and ii) LU 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 lower and upper triangular matrices, each one depending on only one parameter.

The first approach minimizes alternatively the classical least squares objective criterion with respect to each parameter of the two equal nonnegative loading matrices and each column of the third loading matrix. This work was published in the IEEE Signal Processing Letters journal [23]. The second technique reduces the previous optimization problem to the computation of the two equal nonnegative loading matrices. 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 briefly presented at EUSIPCO'13 [37] while a journal paper for submission to IEEE Transactions on Signal Processing is in preparation. Numerical experiments on simulated matrices emphasize the advantages of the proposed algorithms over classical CP and JDC techniques, especially in the case of degeneracies.

## 6.4. Source separation and localization

Source separation, sparse representations, tensor decompositions, semi-nonnegative independent component analysis, probabilistic model, source localization

### 6.4.1. A general framework for audio source separation

**Participants:** Frédéric Bimbot, Rémi Gribonval, Nancy Bertin.

*Main collaboration:* E. Vincent (EPI PAROLE, Inria Nancy); N.Q.K. Duong (Technicolor R&I France)

Source separation is the task of retrieving the source signals underlying a multichannel mixture signal. The state-of-the-art approach consists of representing the signals in the time-frequency domain and estimating the source coefficients by sparse decomposition in that basis. This approach relies on spatial cues, which are often not sufficient to discriminate the sources unambiguously. Recently, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [103], which generalizes a number of existing techniques including our former study on spectral GMMs [66]. This framework makes it possible to quickly design a new model adapted to the data at hand and estimate its parameters via the EM algorithm. As such, it is expected to become the basis for a number of works in the field, including our own.

Since the EM algorithm is sensitive to initialization, we devoted a major part of our work to reducing this sensitivity. One approach is to use some prior knowledge about the source spatial covariance matrices, either via probabilistic priors [82] or via deterministic subspace constraints [91]. The latter approach was the topic of the PhD thesis of Nobutaka Ito [90]. A complementary approach is to initialize the parameters in a suitable way using source localization techniques specifically designed for environments involving multiple sources and possibly background noise [74]. This year, we showed that the approach provides a statistically principled solution to the permutation problem in a semi-informed scenario where the source positions and certain room characteristics are known [15].

### 6.4.2. Towards real-world separation and remixing applications

**Participants:** Nancy Bertin, Frédéric Bimbot, Jules Espiau de Lamaestre, Jérémy Paret, Laurent Simon, Nathan Souviraà-Labastie, Joachim Thiemann.

*Shoko Araki, Jonathan Le Roux (NTT Communication Science Laboratories, JP), E. Vincent (EPI PAROLE, Inria Nancy)*

Following our founding role in the organization of the Signal Separation Evaluation Campaigns (SiSEC) [65], [101], our invited paper summarized the outcomes of the three first editions of this campaign from 2007 to 2010 [116]. While some challenges remain, this paper highlighted that progress has been made and that audio source separation is closer than ever to successful industrial applications. This is also exemplified by the ongoing i3DMusic project and the contracts with Canon Research Centre France and MAIA Studio.

Our involvement in evaluation campaigns and source separation community was reinforced by the recording and the public release of the DEMAND (Diverse Environments Multi-channel Acoustic Noise Database) database, which provides multichannel real-world indoor and outdoor environment noise [44] under Creative Commons licence.

In order to exploit our know-how for these real-world applications, we investigated issues such as how to implement our algorithms in real time [111], how to adapt EM rules for faster computation in multichannel setting [35], how to reduce artifacts [96], how our techniques compare to beamforming in realistic conditions [36], and (in the context of our collaboration with MAIA studios) how best to exploit extra information or human input. In addition, while the state-of-the-art quality metrics previously developed by METISS remain widely used in the community, we proposed some improvements to the perceptually motivated metrics introduced last year [117].

### 6.4.3. Exploiting filter sparsity for source localization and/or separation

**Participants:** Alexis Benichoux, Rémi Gribonval, Frédéric Bimbot.

*E. Vincent (EPI PAROLE, Inria Nancy)*

Estimating the filters associated to room impulse responses between a source and a microphone is a recurrent problem with applications such as source separation, localization and remixing.

We considered the estimation of multiple room impulse responses from the simultaneous recording of several known sources. Existing techniques were restricted to the case where the number of sources is at most equal to the number of sensors. We relaxed this assumption in the case where the sources are known. To this aim, we proposed statistical models of the filters associated with convex log-likelihoods, and we proposed a convex optimization algorithm to solve the inverse problem with the resulting penalties. We provided a comparison between penalties via a set of experiments which shows that our method allows to speed up the recording process with a controlled quality tradeoff [72], [71]. This was a central part of the Ph.D. thesis of Alexis Benichoux [12] defended this year. A journal paper including extensive experiments with real data has been submitted [69].

We also investigated the filter estimation problem in a blind setting, where the source signals are unknown. On a more theoretical side, we studied the frequency permutation ambiguity traditionally incurred by blind convolutive source separation methods. We focussed on the filter permutation problem in the absence of scaling, investigating the possible use of the temporal sparsity of the filters as a property enabling permutation correction. The obtained theoretical and experimental results highlight the potential as well as the limits of sparsity as an hypothesis to obtain a well-posed permutation problem. This work has been published in a conference [70] and as a journal paper [14].

Finally, we considered the problem of blind sparse deconvolution, which is common in both image and signal processing. To counter-balance the ill-posedness of the problem, many approaches are based on the minimization of a cost function. A well-known issue is a tendency to converge to an undesirable trivial solution. Besides domain specific explanations (such as the nature of the spectrum of the blurring filter in image processing) a widespread intuition behind this phenomenon is related to scaling issues and the nonconvexity of the optimized cost function. We proved that a fundamental issue lies in fact in the intrinsic properties of the cost function itself: for a large family of shift-invariant cost functions promoting the sparsity of either the filter or the source, the only global minima are trivial. We completed the analysis with an empirical method to verify the existence of more useful local minima [25].

#### **6.4.4. Semi-nonnegative independent component analysis**

**Participant:** Laurent Albera.

*Main collaborations: Lu Wang (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France), Huazhong Shu (LIST, China)*

Independent Component Analysis (ICA) plays an important role in many areas including biomedical engineering [93], [64], [95], [118], [106], [81], speech and audio [67], [68], [78], [75], radiocommunications [80] and document restoration [114] to cite a few.

For instance in [114], 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 [81] 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 [106], 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 [112]. Moreover, using the JADE algorithm [76] 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 [75] 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 [81], the thickness of each organ, which stands for the mixing coefficient, is real positive. Furthermore, reflectance indices in [114] for the background, the overwriting and the underwriting, which correspond to the mixing coefficients, are also nonnegative. Regarding tissue classification from MRS data, each observation is a linear combination of independent spectra with positive weights representing concentrations [87]; the mixing matrix is again nonnegative.

By imposing the nonnegativity of the mixing matrix within the ICA process, we 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 [98] 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 6.3.2. Hence we use the new methods described in section 6.3.2 to perform semi-nonnegative ICA. Performance results in audio and biomedical engineering were given in the different papers cited in section 6.3.2.

#### 6.4.5. Brain source localization

**Participants:** Laurent Albera, Srdan Kitic, Nancy Bertin, Rémi Gribonval.

*Main collaborations:* 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), Gwenaël Birot (FBML, Genève), Martin Haardt (TUI, Germany)

*Main collaborations:* Hanna Becker (GIPSA & LTSI, France), Pierre Comon (GIPSA, France), Isabelle Merlet (LTSI, France), Fabrice Wendling (LTSI, 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 [105], cortical LORETA (cLORETA) [104], and 4-ExSo-MUSIC [73]. 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 recently submitted to the Elsevier NeuroImage journal.

##### 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 submitted to ICASSP'14 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 submitted to ICASSP'14.

## 6.5. Audio and speech content processing

Audio segmentation, speech recognition, motif discovery, audio mining

### 6.5.1. Audio motif discovery

**Participants:** Frédéric Bimbot, Laurence Catanese.

*This work was performed in close collaboration with Guillaume Gravier from the Texmex 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.

Former work extended the principles of seeded discovery to near duplicate detection and spoken document retrieval from examples [99].

In 2012, the work achieved consisted in consolidating previously obtained results with the motif discovery algorithm and making implementation choices regardless of the structure and the code, in order to minimize the computation time. This has lead to the creation of a software prototype called MODIS.

After the code has been thoroughly optimised, further optimizations to improve the system performances was to change the method used for the search of similarities between patterns. A new functionality has been added to get rid of irrelevant patterns like silence in speech. New versions of dynamic time warping have been implemented, as well as the possibility to downsample the input sequence during the process, which allows a huge gain of computation time.

The principles of the MODIS software has been documented in details [48] and demonstrated during a Show & Tell session at the Interspeech 2013 conference [41].

This work has been carried out in the context of the Quaero Project.

### 6.5.2. Landmark-driven speech recognition

**Participant:** Stefan Ziegler.

*This work is supervised by Guillaume Gravier and Bogdan Ludusan from the Texmex project-team.*

Our previous studies indicate that acoustic-phonetic approaches to ASR, while they cannot achieve state-of-the-art ASR performance by themselves, can prevent HMM-based ASR from degrading, by integrating additional knowledge into the decoding.

In our previous framework we inserted knowledge into the decoding by detecting time frames (referred to as landmarks) which estimate the presence of the active broad phonetic class. This enables the use of a modified version of the viterbi decoding that favours states that are coherent with the detected phonetic knowledge [122].

In 2012 we focused on two major issues. First, we aimed at finding new ways to model and detect phonetic landmarks. Our second focus was on the extension of our landmark detector towards a full acoustic-phonetic framework, to model speech by a variety of articulatory features.

Our new approach for the classification and detection of speech units focuses on developing landmark-models that are different from existing frame-based approaches to landmark detection [121]. In our approach, we use segmentation to model any time-variable speech unit by a fixed-dimensional observation vector. After training any desired classifier, we can estimate the presence of a desired speech unit by searching for each time frame the corresponding segment, that provides the maximum classification score.

We used this segment-based landmark-detection inside a standalone acoustic-phonetic framework that models speech as a stream of articulatory features. In this framework we first search for relevant broad phonetic landmarks, before attaching each landmark with the full set of articulatory features.

Integrating these articulatory feature streams into a standard HMM-based speech recognizer by weighted linear combination improves speech recognition up to 1.5

Additionally, we explored the possibilities of using stressed syllables as an information to guide the viterbi decoding. This work was carried under the leadership of Bogdan Ludusan from the team TEXMEX at IRISA [97].

### 6.5.3. *Mobile device for the assistance of users in potentially dangerous situations*

**Participants:** Romain Lebarbenchon, 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.

Currently in its initial phase, work has mainly focused on functional specifications and performance requirements.

## 6.6. Music Content Processing and Music Information Retrieval

Acoustic modeling, non-negative matrix factorisation, music language modeling, music structure

### 6.6.1. *Music language modeling*

**Participants:** Frédéric Bimbot, Dimitri Moreau, Stanislaw Raczynski.

*Main collaboration:* S. Fukayama (University of Tokyo, JP), E. Vincent (EPI PAROLE, Inria Nancy), Intern: A. Aras

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 joint modeling of "horizontal" (sequential) and "vertical" (simultaneous) dependencies between notes by log-linear interpolation of the corresponding conditional distributions. We identified the normalization of the resulting distribution as a crucial problem for the performance of the model and proposed an exact solution to this problem [108]. We also applied the log-linear interpolation paradigm to the joint modeling of melody, key and chords, which evolve according to different timelines [107]. In order to synchronize these feature sequences, we explored the use of beat-long templates consisting of several notes as opposed to short time frames containing a fragment of a single note.



The limited availability of multi-feature symbolic music data is currently an issue which prevents the training of the developed models on sufficient amounts of data for the unsupervised probabilistic approach to significantly outperform more conventional approaches based on musicological expertise. We outlined a procedure for the semi-automated collection of large-scale multifeature music corpora by exploiting the wealth of music data available on the web (audio, MIDI, leadsheets, lyrics, etc) together with algorithms for the automatic detection and alignment of matching data. Following this work, we started collecting pointers to data and developing such algorithms.

Effort was dedicated to the investigation of structural models for improving the modeling of chord sequence. Preliminary results obtained during Anwaya Aras' internship show that using a matricial structure of time dependencies between successive chords improves the predictability of chord sequences as compared to a purely sequential model.

### 6.6.2. Music structuring

**Participants:** Frédéric Bimbot, Anaik Olivero, Gabriel Sargent.

*Main collaboration:* E. Vincent (EPI PAROLE, Inria Nancy), *Intern:* E. Deruty

The structure of a music piece is a concept which is often referred to in various areas of music sciences and technologies, but for which there is no commonly agreed definition. This raises a methodological issue in MIR, when designing and evaluating automatic structure inference algorithms. It also strongly limits the possibility to produce consistent large-scale annotation datasets in a cooperative manner.

Last year, our methodology for the *semiotic* annotation of music pieces has developed and concretized into a set of principles, concepts and conventions for locating the boundaries and determining metaphoric labels of music segments. The method relies on a new concept for characterizing the inner organization of music segments called the System & Contrast (S&C) model [2]. The annotation of 383 music pieces has been finalized, documented [28] and released to the MIR scientific community: <http://musicdata.gforge.inria.fr/structureAnnotation.html>.

For what concerns algorithmic approaches to music structure description [13], we have formulated the segmentation process as the optimization of a cost function which is composed of two terms: the first one corresponds to the characterization of structural segments by means of audio criteria; the second one relies on the regularity of the target structure with respect to a "structural pulsation period". In this context, we have compared several regularity constraints and studied the combination of audio criteria through fusion. We also considered the estimation of structural labels as a probabilistic finite-state automaton selection process : in this scope, we have proposed an auto-adaptive criterion for model selection, applied to a description of the tonal content. We also proposed a labeling method derived from the system-contrast model. We have evaluated and compared several systems for structural segmentation of music based on these approaches in the context of national and international evaluation campaigns (Quaero, MIREX).

As a follow-up to this work on music structure description, we are currently designing new models and algorithms for segmenting and labeling music into structural units. In one approach (Corentin Guichaoua's PhD), music structure is described as a hierarchical tree estimated by a grammar inference process whereas a second approach (Anaik Olivero's Post-doc) addresses music structure description as the estimation of a graph of similarity relationships.

## PAROLE Project-Team

# 6. New Results

## 6.1. Speech analysis and synthesis

**Participants:** Anne Bonneau, Vincent Colotte, Dominique Fohr, Yves Laprie, Joseph Di Martino, Slim Ouni, Agnès Piquard-Kipffer, Emmanuel Vincent, Utpala Musti.

Signal processing, phonetics, health, perception, articulatory models, speech production, learning language, hearing help, speech analysis, acoustic cues, speech synthesis

### 6.1.1. Acoustic-to-articulatory inversion

The acoustic-to-articulatory inversion from cepstral data has been evaluated on the X-ray database, i.e. X-ray films recorded with the original speech signal. A codebook is used to represent the forward articulatory to acoustic mapping and we designed a loose matching algorithm using spectral peaks to access it. This algorithm, based on dynamic programming, allows some peaks in either synthetic spectra (stored in the codebook) or natural spectra (to be inverted) to be omitted. Quadratic programming is used to improve the acoustic proximity near each good candidate found during codebook exploration. The inversion [40], [10] has been tested on speech signals corresponding to the X-ray films. It achieves a very good geometric precision of 1.5 mm over the whole tongue shape unlike similar works which limit the error evaluation at 3 or 4 points corresponding to sensors located at the front of the tongue.

#### 6.1.1.1. Construction of articulatory models

Articulatory models are intended to approximate the vocal tract geometry with a small number of parameters controlling linear deformation modes. Most of the models have been designed on images of vowels and thus offer a good coverage for vowels but are unable to provide a good approximation for consonants, especially in the region of the constriction. The first problem is related to the nature of contours used to derive linear components. When dealing with vowels there is no contact between the tongue and other fixed articulators (palate, teeth). Factor analysis used to determine linear modes of deformation of the tongue only takes into account the influence of the tongue muscles. This is no longer the case with consonants, since a contact is realized between the tongue and the palate, alveolar ridge or teeth for stops /k, g, t, d/ and the sonorant /l/ in French. The deformation factors thus incorporate the “clipping” effect of the palate. Following the idea of using virtual articulatory targets that lie beyond the positions that can be reached, here the palate, we edited delineated tongue contours presenting a contact with the palate. We chose a conservative solution which consists of keeping the tongue contour up to the contact point and extending it while guaranteeing a “natural shape”. These new contours do not cross the palate for more than 10 mm. As such, this modification alone is not sufficient, because the number of images corresponding to consonants is small even if the corpus used in this work is phonetically balanced. We thus duplicated a number of consonant X-ray images in order to increase the weight of deformation factors corresponding to the tongue tip which is essential for some consonants, /l/ for instance. This approach provides a very good fitting with original tongue contours, i.e. 0.83 mm in average with 6 components over the whole tongue contour and only 0.56 mm in the region of the main place of articulation, which is important with a view of synthesizing speech.

#### 6.1.1.2. Articulatory copy synthesis

Acoustic features and articulatory gestures have always been studied separately. Articulatory synthesis could offer a nice solution to study both domains simultaneously provided that relevant information can be fed into the acoustic simulation. The first step consisted of connecting the 2D geometry given by mediosagittal images of the vocal tract with the acoustic simulation. Last year we thus developed an algorithm to compute the centerline of the vocal tract, i.e. a line which is approximately perpendicular to the wave front. The centerline is then used to segment the vocal tract into elementary tubes whose acoustic equivalents are fed into the acoustic simulation. A new version of the centerline algorithm [53] has been developed in order to approximate the propagation of a plane wave more correctly.

The work on the development of time patterns used to pilot the acoustic simulation has been continued by improving the choice of relevant X-ray images and the temporal transitions from one image to the following. This procedure has been applied successfully to copy sentences and VCV for four X-ray films of the DOCVACIM database[52]. More difficult transitions, i.e. those corresponding to consonant clusters, will be investigated this year.

In addition to the control of the acoustic simulation we started an informal cooperation with the IADI laboratory [www.iadi-nancy.fr](http://www.iadi-nancy.fr) in order to record better static images of the vocal tract, and cineMRI, i.e. films, for a number of sentences.

### **6.1.2. Using articulatory data for speech animation**

We are continuously working on the acquisition and analysis of the articulatory data using electromagnetic articulography (EMA). This year, we have conducted research to use EMA as motion capture data and we showed that it is possible to use it for audiovisual speech animation. In fact, as EMA captures the position and orientation of a number of markers, attached to the articulators, during speech, it performs the same function for speech that conventional motion capture does for full-body movements acquired with optical modalities, a long-time staple technique of the animation industry. We have processed EMA data from a motion-capture perspective and applied to the visualization of an existing multimodal corpus of articulatory data, creating a kinematic 3D model of the tongue and teeth by adapting a conventional motion capture based animation paradigm. Such an animated model can then be easily integrated into multimedia applications as a digital asset, allowing the analysis of speech production in an intuitive and accessible manner. In this work [61], we have addressed the processing of the EMA data, its co-registration with 3D data from vocal tract magnetic resonance imaging (MRI) and dental scans, and the modeling workflow. We will continue our effort in the future to improve this technique.

### **6.1.3. Acoustic analyses of non-native speech**

Within the framework of the project IFCASL, we designed a corpus for the study of French and German, with both languages pronounced by French and German speakers, so as to put into light L1/L2 interferences. The corpus was constructed to control for several segmental and suprasegmental phenomena. German and French, for instance, show different kinds of voicing patterns. Whereas in French, the voicing opposition of stops is realized as voiced versus unvoiced, in German, the same difference is realized mostly as unaspirated versus aspirated. Furthermore, differences between the two language groups are expected with respect to the production of nasal vowels (absent in German), the realization of /h/ (not present in French, but in German). On the suprasegmental level, word stress and focus intonation are central to our investigation. Speakers produce both native and non-native speech, which allows for a parallel investigation of both languages.

We have conducted a pilot study on the realization of obstruents in word-final position -a typical example of L1-L2 interference on the segmental level-, which are subject to devoicing in German, but not in French. First results showed that German learners (beginners) had difficulties to voice French obstruents in this context, and, when listening to French realizations, tend to add a final schwa to achieve the expected realization.

### **6.1.4. Speech synthesis**

We recall that within the framework of the ViSAC project we have developed bimodal acoustic-visual synthesis technique that concurrently generates the acoustic speech signal and a 3D animation of the speaker's outer face. This is done by concatenating bimodal diphone units that consist of both acoustic and visual information. In the visual domain, we mainly focus on the dynamics of the face rather than on rendering. The proposed technique overcomes the problems of asynchrony and incoherence inherent in classic approaches to audiovisual synthesis. The different synthesis steps are similar to typical concatenative speech synthesis but are generalized to the acoustic-visual domain. This year we have performed an extensive evaluation of the synthesis system using perceptual and subjective evaluations. The overall outcome of the evaluation indicates that the proposed bimodal acoustic-visual synthesis technique provides intelligible speech in both acoustic and visual channels [22]. For testing purposes we have also added a simple tongue model that is controlled by the generated phonemes. The purpose is to improve the quality of the audiovisual speech intelligibility.

Moreover, we perform feature selection and weight tuning for a given unit-selection corpus to make the ranking given by the target cost function consistent with the ordering given by an objective dissimilarity measure. To find an objective metric highly correlated to perception we analyzed correlation between objective and subjective evaluation results. It shows interesting patterns which might help in designing better tuning metrics and objective evaluation techniques [55].

### **6.1.5. Phonemic discrimination evaluation in language acquisition and in dyslexia and dysphasia**

We keep working on a project concerning identification of early predictors of reading, reading acquisition and language difficulties, more precisely in the field of specific developmental disabilities : dyslexia and dysphasia. A fair proportion of those children show a weakness in phonological skills, particularly in phonemic discrimination. However, the precise nature and the origin of the phonological deficits remain unspecified. In the field of dyslexia and normal acquisition of reading, our first goal was to contribute to identify early indicators of the future reading level of children. We based our work on the longitudinal study - with 85 French children - of [90], [91] which indicates that phonemic discrimination at the beginning of kindergarten is strongly linked to success and specific failure in reading acquisition. We study now the link between oral discrimination both with oral comprehension and written comprehension. Our analyses are based on the follow up of a hundred children for 4 years from kindergarten to end of grade 2 (from age 4 to age 8) [98].

### **6.1.6. Enhancement of esophageal voice**

#### *6.1.6.1. Pitch detection*

Over the last two years, we have proposed two new real time pitch detection algorithms (PDAs) based on the circular autocorrelation of the glottal excitation, weighted by temporal functions, derived from the CATE [85] original algorithm (Circular Autocorrelation of the Temporal Excitation), proposed initially by J. Di Martino and Y. Laprie. In fact, this latter algorithm is not constructively real time because it uses a post-processing technique for the Voiced/Unvoiced (V/UV) decision. The first algorithm we developed is the eCATE algorithm (enhanced CATE) that uses a simple V/UV decision less robust than the one proposed later in the eCATE+ algorithm. We propose a recent modified version called the eCATE++ algorithm which focuses especially on the detection of the F0, the tracking of the pitch and the voicing decision in real time. The objective of the eCATE++ algorithm consists in providing low classification errors in order to obtain a perfect alignment with the pitch contours extracted from the Bagshaw or Keele databases by using robust voicing decision techniques. This algorithm has been published in *Signal, Image and Video Processing*, [14].

#### *6.1.6.2. Real-time pitch detection for application to pathological voices*

The work first rested on the CATE algorithm developed by Joseph Di Martino and Yves Laprie, in Nancy, 1999. The CATE (Circular Autocorrelation of the Temporal Excitation) algorithm is based on the computation of the autocorrelation of the temporal excitation signal which is extracted from the speech log-spectrum. We tested the performance of the parameters using Bagshaw database, which is constituted of fifty sentences, pronounced by a male and a female speaker. The reference signal is recorded simultaneously with a microphone and a laryngograph in an acoustically isolated room. These data are used for the calculation of the contour of the pitch reference. When the new optimal parameters from the CATE algorithm were calculated, we carried out statistical tests with the C functions provided by Paul BAGSHAW. The results obtained were very satisfactory and a first publication relative to this work was accepted and presented at the ISIVC 2010 conference [79]. At the same time, we improved the voiced / unvoiced decision by using a clever majority vote algorithm electing the actual F0 index candidate. Recently Fadoua Bahja developed a new algorithm based on wavelet transforms applied to the cepstrum excitation. The preliminary results obtained were satisfactory and a complete description of this latter study is under a submission process in an international journal.

#### *6.1.6.3. Voice conversion techniques applied to pathological voice repair*

Voice conversion is a technique that modifies a source speaker's speech to be perceived as if a target speaker had spoken it. One of the most commonly used techniques is the conversion by GMM (Gaussian Mixture Model). This model, proposed by Stylianou, allows for efficient statistical modeling of the acoustic space of a

speaker. Let “x” be a sequence of vectors characterizing a spectral sentence pronounced by the source speaker and “y” be a sequence of vectors describing the same sentence pronounced by the target speaker. The goal is to estimate a function F that can transform each source vector as nearest as possible of the corresponding target vector. In the literature, two methods using GMM models have been developed: In the first method (stylianou,98), the GMM parameters are determined by minimizing a mean squared distance between the transformed vectors and target vectors. In the second method (kain,98), source and target vectors are combined in a single vector “z”. Then, the joint distribution parameters of source and target speakers is estimated using the EM optimization technique. Contrary to these two well known techniques, the transform function F, in our laboratory, is statistically computed directly from the data: no needs of EM or LSM techniques are necessary. On the other hand, F is refined by an iterative process. The consequence of this strategy is that the estimation of F is robust and is obtained in a reasonable lapse of time. Recently, we realized that one of the most important problems in speech conversion is the prediction of the excitation. In order to solve this problem we developed a new strategy based on the prediction of the cepstrum excitation pulses. Another very important problem in voice conversion concerns the prediction of the phase spectra. This study is under progress in the framework of an Inria ADT which began in September 2013.

#### 6.1.6.4. Signal reconstruction from short-time Fourier transform magnitude spectra

Joseph Di Martino and Laurent Pierron developed in 2010 an algorithm for real-time signal reconstruction from short-time Fourier magnitude spectra [86]. Such an algorithm has been designed in order to enable voice conversion techniques we are developing in Nancy for pathological voice repair. Recently Mouhcine Chami, an assistant-professor of the INPT institute at Rabat (Morocco) proposed a hardware implementation of this algorithm using FPGAs. This implementation has been published in the SIIE 2012 conference [81]. Maryem Immassi, a PhD student of Mouhcine Chami, is comparing this algorithm with the state of the art RTISI-LA algorithm in the framework of a hardware implementation.

#### 6.1.7. Audio source separation

Audio source separation is the task of extracting one or more target source signals from a given mixture signal. It 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. We studied the impact of sparsity penalties over the mixing filters [38] and we defined probabilistic priors [20] and deterministic subspace constraints [45] over the spatial covariance matrices. We also wrote a review paper about guided audio source separation for *IEEE Signal Processing Magazine* [28].

This paper highlighted that many guided separation techniques now exist that are closer than ever to successful industrial applications, as exemplified by the ongoing industrial collaborations of the team. In order to exploit our know-how for these real-world applications, we investigated issues such as the impact of audio coding [59], artifact reduction [21], real-time implementation [62], and latency [70]. Two patents have been filed [77], [76]. We also started a new research track on the fusion of multiple source separation techniques [46].

Finally, we pursued our long-lasting efforts on the evaluation of audio source separation by collecting the first-ever publicly available dataset of multichannel real-world noise recordings [71] and by conducting an experimental comparison of the two main families of techniques used for source separation [63].

## 6.2. Automatic speech recognition

**Participants:** Dominique Fohr, Jean-Paul Haton, Irina Illina, Denis Jovet, Odile Mella, Emmanuel Vincent, Arseniy Gorin, Luiza Orosanu, Dung Tran.

stochastic models, acoustic models, language models, automatic speech recognition, speech transcription, training, robustness

### 6.2.1. 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.



The use of larger speech training corpora allows us increasing the size of the acoustic models (more parameters through more Gaussians components per density, and more shared densities) and this leads to improved performance. However, in such approaches, Gaussian components are estimated independently for each density. Thus, after having investigated last year the usage of multiple modeling approaches for better constraining the acoustic decoding space, recent studies have focused on enriching the acoustic models themselves in view of handling trajectory and speaker consistency in decoding.

This year a new modeling approach was developed that takes benefit 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, some Gaussian components are estimated on the data of each class. 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. The method allows us to better parameterize GMM-HMM without increasing significantly the number of model parameters. The experiments on French radio broadcast news data demonstrate the improvement of the accuracy with such parameterization compared to the models with similar, or even larger number of parameters [43].

Current experiments deal with stranded HMM. The objective of such an approach is to introduce in the GMM-HMM modeling some extra parameters to take into account the transition between the Gaussian components when moving from one frame to the next.

### 6.2.2. Noise-robust speech recognition

In many real-world conditions, the speech signal is overlapped with noise, including environmental sounds, music, or undesired extra speech. Source separation may then be used as a pre-processing stage to enhance the desired speech signal [64]. In practice, the enhanced signal always includes some distortions compared to the original clean signal. It is important to quantify which parts of the enhanced signal are reliable in order not to propagate these distortions to the subsequent feature extraction and decoding stages. A number of heuristic statistical uncertainty estimators and propagators have been proposed to this aim. We started some work aiming to improve the accuracy of these estimators and propagators. We also showed how to exploit uncertainty in order to train unbiased acoustic models directly from noisy data [24].

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. This challenge aims to recognize small or medium-vocabulary speech mixed with noise recorded in a real family home over the course of several weeks. We analyzed the outcomes of the first edition [16] which led to a special issue of *Computer Speech and Language* [15] and we organized a second edition in 2013 [66] which illustrated the progress made in two years over small-vocabulary speech and the remaining challenges towards robust recognition of medium-vocabulary speech [65].

### 6.2.3. Linguistic modeling

Usually the lexicon used by a speech recognition system refers to word entries, where each entry in the pronunciation lexicon specifies a possible pronunciation of a word, and the associated language model specifies the probability of a word knowing preceding words. However, whatever the size of the lexicon is, the size is always finite, and the speech recognition system cannot recognize properly words that are not present in the lexicon. In such cases, the unknown word is typically replaced by a sequence of short words which is acoustically similar to the unknown speech portion.

#### 6.2.3.1. Random indexing

This year we studied the introduction of semantic information through the Random Indexing paradigm (RI) in statistical language models used in speech recognition. Random Indexing is a scalable alternative to LSA (Latent Semantic Analysis) for analyzing relationships between a set of documents and the terms they contain. We determined the best methods and parameters by minimizing the perplexity of a realistic corpus of 290000 words. We investigated 4 methods for training RI matrices, 4 weighting functions, several matrix sizes and how balancing the 4-gram and RI language model. We only obtained a relative gain of 3% [42].

### 6.2.3.2. Continuous language models

Language modeling plays an important role in automatic speech recognition because it constrains the decoder to search the most likely sequences of words according to a given language and a given task. A limitation of N-grams models is that they represent the words in a discrete space. It would be interesting to represent words in a continuous space where semantically close words would be projected in the same region of space. This projection can be achieved by recurrent neural networks. Moreover they are able to learn long-term dependencies with the recurrent layer that can store a record of the past. During his master internship, Othman Zennaki integrated this new language model in our speech recognition system ANTS.

### 6.2.3.3. Linguistic units for embedded systems

In the framework of the RAPSODIE project, speech recognition is to be used to help communication with hard of hearing people. Because of requirements on memory and CPU (almost real time processing), various modeling approaches have been investigated with respect to linguistic units. The first approach has focused on analyzing the achieved phonetic decoding performance of various linguistic units (phonemes, syllables, words). Best phonetic decoding performance is achieved using word units and associated tri-gram language model, but at the expense of large CPU and memory requirements. Using directly phoneme units leads to the smallest models and requires little CPU, however, this also leads to the worst performance. The proposed approach relying on syllable units provides results which are rather close to the word based approach, but requires much less CPU [58], [57].

Further experiments are now focusing on combining word and syllable units, in view of having frequent words covered by the word units, and using syllables for decoding unknown words.

### 6.2.3.4. OOV proper name retrieval

Proper name recognition 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.

In the framework of the ContNomina project, we focus on increasing the vocabulary coverage of a speech transcription system by automatically retrieving proper names from contemporary diachronic text documents. We proposed methods that dynamically augment the automatic speech recognition system vocabulary, using lexical and temporal features in diachronic documents. We also studied different metrics for proper name selection in order to limit the vocabulary augmentation and therefore the impact on the ASR performances. Recognition results show a significant reduction of the word error rate using augmented vocabulary [56].

## 6.2.4. Speech transcription

The first complete version of the speech transcription system ANTS (see section 5.5 ) has been initially developed in the framework of the Technolangu project ESTER, and since then, the system has been regularly enriched through the integration of research results. The latest version can handle either HTK-based acoustic models through the Julius decoder, or Sphinx-based acoustic models with the CMU Sphinx decoders. In the last version, a Perl script encapsulates all the calls to the various tools used for diarization, model adaptation and speech recognition, and takes benefit of the multiple CPU available on the computer for parallelizing the different tasks as much as possible.

### 6.2.4.1. Combining recognizers

Last year in the context of the ETAPE speech transcription evaluation campaign, the Sphinx-based and Julius-based decoders have been further improved, and it was observed that combining the recognition outputs of several Sphinx-based and Julius-based decoder lead to a significant word error rate reduction compared to the best individual system.

More controlled experiments have then been performed to understand what was the main reason of the large performance improvement observed when combining Julius-based and Sphinx-based transcription system results. The Sphinx decoder processes the speech data in a forward pass, whereas the Julius decoder ends its decoding process by a backward pass. The Sphinx training and decoding scripts have been modified to process the speech material in a reverse time order; and various systems were developed by using different

sets of acoustic features and different sets of acoustic units. It was then observed that combining several Sphinx-forward and several Sphinx-reverse decoders lead to much better results than combining the same amount of only Sphinx-forward decoders or only Sphinx-reverse decoders; and the achieved word error rate was consistent with the one obtained by combining the Sphinx-based (forward) and Julius-based (backward) decoders [49]. Hence, the improvement is mainly due to the fact that forward-based and backward-based processing are combined. Because heuristics are applied during decoding to limit the acoustic space that is explored, some hypotheses might be wrongly pruned when processing the data one way, and may be kept in the active beam search when processing the other way. This is corroborated by the analysis of the word graph which show a large dissimilarity in the distribution of the number of words starting and ending in each frame [48].

Experiments have also shown that when the forward and backward decoders yield the same word hypothesis, this word is likely to be a correct answer. Recent experiments are investigating how far such behavior could help for unsupervised learning of acoustic models.

#### 6.2.4.2. Spontaneous speech

During his master internship, Bruno Andriamiarina focuses on the new challenges brought by this spontaneity of the speech, making it difficult to be transcribed by the existing automatic speech recognition systems. He studied how to improve global performance of automatic speech recognition systems when dealing with spontaneous speech by adapting language model and pronunciation dictionary to this particular type of speech. He also studied the detection of disfluent speech portions (produced by spontaneous speech) in speech signal using a Gaussian Mixture Model (GMM)-based classifier trained on prosodic features covering the main prosodic characteristics (duration, fundamental frequency and energy).

#### 6.2.4.3. Towards a structured output

The automatic detection of the prosodic structure of speech utterances has been investigated. The algorithm relies on a hierarchical representation of the prosodic organization of the speech utterances, and detects prosodic boundaries whether they are followed or not by pause. The detection of the prosodic boundaries and of the prosodic structures is based on an approach that integrates little linguistic knowledge and mainly uses the amplitude of the F0 slopes and the inversion of the F0 slopes as well as phone durations. The approach was applied on a corpus of radio French broadcast news and also on radio and TV shows which are more spontaneous speech data. The automatic prosodic segmentation results were then compared to a manual prosodic segmentation made by an expert phonetician [37].

Further work has focused on analyzing the links between manually set punctuation marks and this automatically detected prosodic structure, in view of using the prosodic structure for helping an automatic punctuation process.

### 6.2.5. Speech/text alignment

#### 6.2.5.1. Alignment with non-native speech

Non-native speech alignment with text is one critical step in computer assisted foreign language learning. 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 - too short or too long -). 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.

In the case of French speakers learning English, we conducted a detailed analysis that has showed the benefit of taking into account non-native variants, and lead to determining the classes of phonemes whose temporal boundaries are the most accurate and which should be favored in the design of exercises for language learning[18].



In the framework of the IFCASL project, we proposed to use a two-step approach for automatic phone segmentation. The first step consists in determining the phone sequence that best explains the learner's utterance. This is achieved by force aligning the learner's speech utterance with a model representing the various possible pronunciation variants of the current sentence (both native and non-native variants need to be considered). In this step detailed acoustic Hidden Markov Models (HMMs) are used, with a rather large number of Gaussian components per mixture density. This kind of detailed acoustic models is the one that provides the best performance in automatic speech recognition. The second step consists in determining the phone boundaries. This is also achieved through a forced alignment process, but this time, the sequence of phones is known (as determined in the first step), and phone acoustic models with only a few Gaussians components per mixture density are used because it has been shown that they provide better temporal precision than detailed acoustic models. For the training of the models used for both forced alignment steps, the speech of native and non-native speakers could be used, either directly or by MLLR (Maximum Likelihood Linear Regression) adaptation.

#### 6.2.5.2. *Alignment with spontaneous speech*

In the framework of the ANR ORFEO, 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. As regards overlapping speech, the orthographic transcription of the audio files only provides a rather imprecise time information of the overlapping speech segment. As a first approach, 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 with the same time boundaries.

### 6.3. Machine translation and language modeling

**Participants:** Kamel Smaïli, David Langlois, Denis Jouvét, Emmanuel Vincent, Motaz Saad, Cyrine Nasri.

machine translation, statistical models

#### 6.3.1. *Language modeling*

##### 6.3.1.1. *Vocabulary selection*

In the framework of the ETAPE evaluation campaign a new machine learning based process was developed to select the most relevant lexicon to be used for the transcription of the speech data (radio and TV shows). The approach relies on a neural network trained to distinguish between words that are relevant for the task and those that are not. After training, the neural network (NN) is applied to each possible word (text tokens extracted from a very large text corpus). Then the words that have the largest NN output score are selected for creating the speech recognition lexicon. Such an approach can handle counts of occurrences of the words in various data subsets, as well as other complementary information, and thus offer more perspectives than the traditional unigram-based selection procedures [50].

##### 6.3.1.2. *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 log-linear interpolation of multiple conditional distributions. We applied it to the joint modeling of "horizontal" (sequential) and "vertical" (simultaneous) dependencies between notes for polyphonic pitch estimation [26] and to the joint modeling of melody, key and chords for automatic melody harmonization [25]. 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 [74].

### **6.3.2. Quality estimation of machine translation**

In the scope of Confidence Measures, we participated to the World Machine Translation evaluation campaign for the second year (WMT2013 <http://www.statmt.org/wmt13/quality-estimation-task.html>). More precisely, we proposed a Quality Estimation system to the Quality Estimation shared task. The goal was to predict the quality of translations generated by an automatic system. Each translated sentence is given a score between 0 and 1. The score is obtained by using several numerical or boolean features calculated according to the source and target sentences. We performed a linear regression of the feature space against scores in the range [0 ;1], to this end, we use a Support Vector Machine with 66 features. In this new participation, we proposed to increase the size of the training corpus. For that, we decided to use the post-edited and reference corpora in the training step after assigning a score to each sentence of these corpora. Then, we tune these scores on a development corpus. This leads to an improvement of 10.5% on the development corpus, in terms of Mean Average Error (average difference between reference and predicted scores), but achieves only a slight improvement on the test corpus. This work has been published in [51].

### **6.3.3. Comparable corpora and multilingual sentiment analysis**

In the PhD Thesis of Motaz Saad, we work on collecting comparable corpora. For that purpose we presented a method which extracts and aligns comparable corpora at the article level from Wikipedia encyclopedia based on interlanguage links. To evaluate the closeness of corpora we proposed several comparability measures. Our evaluations show that the proposed comparability measures are able to capture the comparability degree of any comparable corpora [60]. We go further on the comparability of multilingual corpora by studying their comparability in terms of sentiment. The final objective is to propose a multilingual press review concerning a given topic. This review should use several multilingual resources (electronic newspapers), and should class resources according to the including sentiments (fear, joy...about the subject), polarity (against or not to the subject)...This conducts to study opinions across different languages by comparing the underlying messages written by different people having different opinions. We propose "Sentiment based Comparability Measures" to compare opinions in multilingual comparable articles without translating source/target into the same language [27].

### **6.3.4. Machine translation of arabic dialect**

The translation of Arabic dialect constitutes a real challenge since it is an under-resourced language. In fact, Modern Standard Arabic is as any other evaluated language, it means it could be processed by the available tools but unfortunately in Arabic countries people speak an Arabic language which is inspired from the standard one but is different. Our objective is then to propose a speech to speech system converting modern standard Arabic to Algerian dialect. After collecting corpus, we decided to propose a method allowing to diacritize dialects in order to be able in the following to develop an acoustic model. For that, we considered the issue of diacritization as a machine translation issue, and we have developed a statistical machine translation which learns to transform an undiacritized corpus into a diacritized one [44].

## SÉMAGRAMME Project-Team

# 6. New Results

## 6.1. Syntax-Semantics Interface

### 6.1.1. TAG, Dependency Grammars, and ACG

Aleksandre Maskharashvili and Sylvain Pogodalla gave an ACG account of [41]’s process of transformation of the derivation trees of Tree Adjoining Grammar (TAG) into dependency trees. They made explicit how the requirement of keeping a direct interpretation of dependency trees into strings results into lexical ambiguity. Since the ACG framework has already been used to provide a logical semantics from TAG derivation trees, it results in a unified picture where derivation trees and dependency trees are related but independent equivalent ways to account for the same surface–meaning relation. This result has been published in [15].

### 6.1.2. Semantics of Neg-Raising Predicates in TAG

Laurence Danlos, Philippe de Groote, and Sylvain Pogodalla proposed a lexical semantic interpretation of Neg-Raising (NR) predicates that heavily relies on a Montague-like semantics for TAG and on higher-order types. NR verbs form a class of verbs with a clausal complement that show the following behavior: when a negation syntactically attaches to the matrix predicate, it can semantically attach to the embedded predicate, as the implication of (2) by (1) shows. This corresponds to the NR reading of this predicate.

- *Marie ne pense pas que Pierre partira.*
- *Marie pense que Pierre ne partira pas.*

As a base case, the approach lexically provides both NR and non-NR readings to NR predicates. The proposal is implemented in the ACG framework as it offers a fairly standard interface to logical formal semantics for TAG. This result has been published in [13].

### 6.1.3. Intensionalization

Makoto Kanazawa and Philippe de Groote have defined a general intensionalization procedure that turns an extensional semantics for a language into an intensionalized one that is capable of accommodating truly intensional lexical items without changing the compositional semantic rules [10]. They have proved some formal properties of this procedure and have clarified its relation to the procedure implicit in Montague’s PTQ.

## 6.2. Lexical Disambiguation

Guy Perrier adapted the methods of lexical disambiguation presented in Mathieu Morey’s PhD thesis [49] to the formalism of Tree Adjoining Grammar (TAG) in a common work with Claire Gardent, Yannick Parmentier and Sylvain Schmitz [24].

More precisely, the algorithm of lexical disambiguation for TAG uses the one-to-one relations between substitution nodes and roots of elementary trees in the parsing process and it takes also into account the position of the substitution nodes with respect to the anchors in elementary trees, to discard lexical selections that do not respect some constraints. These constraints are implemented through a polarization of the elementary trees and for sake of efficiency, the lexical selections are represented in a compact way with automata.

A major default of the methods of lexical disambiguation presented in Mathieu Morey’s PhD thesis is that they ignore local contexts. To overcome this default, Guy Perrier proposed an algorithm to foresee the elementary structures of the grammar that can be inserted between two words that will interact in the parsing process [20]. This algorithm applies to lexicalized grammars, in which the elementary structures are trees.

## 6.3. Linguistic Resources

### 6.3.1. Large Scale Grammatical Resources

Guy Perrier and Bruno Guillaume continued to develop FRIGRAM<sup>2</sup> a French grammar with a large coverage, written in the formalism of Interaction Grammars [16].

A major challenge in this task is to guarantee and to maintain the consistency of the grammar while aiming at the largest coverage. For this, they resorted an original property coming from the polarization of the elementary structures of an interaction grammar : the *companion property*. It is possible to determine all elementary structures (the *companions*) that are able to interact with a given elementary structure, in a static computation on the whole non anchored grammar, using the systeme of polarities. The knowledge of the companions of every elementary structure is very useful to check the linguistic consistency of a grammar.

Guy Perrier wrote a detailed documentation on FRIGRAM illustrated with a lot of examples [26].

### 6.3.2. 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 started a project of annotating the Sequoia French Treebank with deep syntax dependencies.

The Sequoia French Treebank [33] is a 3 200 sentence 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, which abstracts away from surface syntactic variation and diathesis alternations. 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. In Nancy, Line Heckler, Mathilde Huguin and Alice Kneip produced a double annotation of the corpus and Guy Perrier was in charge of the adjudication.

At the beginning of the project, a mini reference was selected randomly, composed of 250 sentences from the Sequoia Corpus. Its annotation was conducted in parallel to the production of the annotation guide, in order to get feedback for the guide. Each team separately produced an initial annotated version of the mini reference. The final version, resulting from several iterations and adjudications, is already available<sup>3</sup>.

The full version of the Sequoia French Treebank with deep syntax dependencies and its annotation guide will be released during Spring 2014.

### 6.3.3. Agile Annotation

In [19], Bruno Guillaume and Karën Fort present a methodology, inspired from the agile development paradigm, that helps preparing an annotation campaign. The idea behind the methodology is to formalize as much as possible the instructions given in the guidelines, in order to automatically check the consistency of the corpus being annotated with the guidelines, as they are being written. To formalize the guidelines, the authors use a graph rewriting tool, that allows to use a rich language to describe the instructions. This formalization allows to spot the rightfully annotated constructions and, by contrast, those that are not consistent with the guidelines. In case of inconsistency, an expert can either correct the annotation or update the guidelines and rerun the process.

<sup>2</sup><http://wikilligramme.loria.fr/doku.php?id=frigram:frigram>

<sup>3</sup>[http://talc2.loria.fr/mini\\_sequoia/](http://talc2.loria.fr/mini_sequoia/)

### **6.3.4. Integration of Multiple Constraints in ACG**

In [14], Jiri Marsik and Maxime Amblard present a first step toward the integration of multiple constraints in ACG. However, all of the known treatments only consider tiny fragments of languages. We are interested in building a wide-coverage grammar which integrates and reconciles the existing formal treatments of discourse and allows us to study their interactions and to build discourse representations automatically.

This proposal is a first step towards a wide-coverage Abstract Categorical Grammar (ACG) that could be used to automatically build discourse-level representations. We focus on the challenge of integrating the treatment of disparate linguistic constraints in a single ACG and propose a generalization of the formalism: Graphical Abstract Categorical Grammars.

## **6.4. Graph Rewriting**

Guillaume Bonfante and Bruno Guillaume studied formal properties of the Graph Rewriting in [12]. It is well-known that some linguistic phenomena do not cope properly with trees as the core mathematical structure to represent linguistic informations. In a former paper, the authors showed the benefit of encoding linguistic structures by graphs and of using graph rewriting rules to compute on those structures.

The Graph Rewriting formalism they consider is a formalization of the system which is implemented in the Grew software. Justified by some linguistic considerations, this Graph Rewriting formalization is characterized by two features: first, there is no node creation along computations and second, there are non-local edge modifications. Under these hypotheses, the article shows that uniform termination is undecidable and that non-uniform termination is decidable. Two termination techniques based on weights are described and a complexity bound on the derivation length for these rewriting systems is given.

## **6.5. Discourse in Pathological context**

Maxime Amblard, Manuel Rebuschi and Michel Musiol continue to analyze in fine details pathological dialogues from the SLAM project. They present all these results in [22] [21] and [11]. 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.

## COPRIN Project-Team

## 6. New Results

### 6.1. Robotics

#### 6.1.1. Cable-driven parallel robots (CDPR)

##### 6.1.1.1. Analysis of Cable-driven parallel robots

**Participants:** Laurent Blanchet, Jean-Pierre Merlet [correspondant], Yves Papegay, Rémy Ramadour.

We are still investigating the extremely complex analysis of the kinematics [24] of CDPRs assuming either rigid [21] [20], elastic or sagging cables.

We have also started an analysis of *cable configuration* of redundantly actuated CDPRs for control purposes. Indeed we have shown that for robot with rigid cables it is impossible to have, in a given pose, more than 6 cables in tension simultaneously: the set of cables under tension is called the cable configuration. However at a pose there may be different sextuplets of under tension cables that satisfy the kinematico-static equations. Each of these sextuplets exhibits different performances (e.g. maximal tension in the cables or sensitivity of the positioning to errors in the cable lengths). Hence it may be interesting for control purposes to select one of the sextuplet that is optimal with respect to a performance criteria and to enforce this configuration by letting *voluntary* the cables that are not in the sextuplet being slack (i.e. adjusting their lengths to be larger than the one required for the pose).

We have generalized this approach for a trajectory of a 4 cables CDPR with all cables attached to the same point of the platform. In that case only up to 3 cables may be under tension at the same time. We have designed an algorithm that determine the optimal cable configuration on the whole trajectory.

Simultaneously we have addressed part of an ambitious goal: a full simulation tool for CDPR. We assume a high level motion planning loop that calculate a motion order every  $\Delta t_1$  second and send this command to an inner motor control loop that execute it by sending a command to the motor every  $\Delta t_2$  second. Then we have a continuous time model of the motor that determine its velocity. The whole purpose is to calculate the pose of the platform together with the tensions in the cables. This simulation is extremely demanding and cannot be performed with classical software because of the changes in the cable configuration that have to be detected for determining the platform pose and cable tensions. We have succeeded for CDPR with rigid and elastic cables, furthermore introducing random errors in the cable length measurements. This tool has allowed us to show that cable tensions are very sensitive: for example a high level loop that is designed to minimize  $\sum \tau_j^2$ , where  $\tau$  are the cable tensions, exhibits large difference with the objective as soon as discrete time-control is taken into account.

##### 6.1.1.2. Certified Calibration of a Cable-Driven Robot Using Interval Contractor Programming

**Participants:** Julien Alexandre Dit Sandretto, David Daney, Gilles Trombettoni.

An interval based approach is proposed to rigorously identify the model parameters of a parallel cable-driven robot. The studied manipulator follows a parallel architecture having 8 cables to control the 6 DOFs of its mobile platform. This robot is complex to model, mainly due to the cable behavior. To simplify it, some hypotheses on cable properties (no mass and no elasticity) are done. An interval approach can take into account the maximal error between this model and the real one. This allows us to work with a simplified although guaranteed interval model. In addition, a specific interval operator makes it possible to manage outliers. A complete experiment validates our method for robot parameter certified identification and leads to interesting observations [9], [16], [15].

##### 6.1.1.3. Tool for Agencement Analysis and Synthesis of CDPRs

**Participants:** Laurent Blanchet, Jean-Pierre Merlet [correspondant].

In the frame of FP7 project CABLEBOT, we are developing a methodology to analyze or synthesize a Cable Driven Parallel Robots configuration i.e. either to determine the performances of a given CDPR (e.g. maximal wire tensions over a given workspace) or, being given a list of requirements, to determine what what are **all** possible CDPR geometries that are guaranteed to satisfy the requirements. This tool relies heavily on our analysis of the CDPRs and on interval analysis.

To illustrate this approach we have developed a software that can be used to illustrate the workings/operating procedure of interval analysis through a 3D visualization. This software sets up a scenario of a CDPR in a warehouse and computes in real time its workspace under different constraints.

#### 6.1.1.4. Visual-servoing of a parallel cable-driven robot

**Participants:** Rémy Ramadour, Jean-Pierre Merlet [correspondant], François Chaumette [correspondant].

MARIONET-ASSIST is a parallel cable-driven robot designed to move through large rooms in order to provide services such as walking-aid, lifting people or manipulating heavy loads. In order to experiment, a full-scaled flat with a crane robot has been built. Adding one or several low-cost cameras (the cost being here a fundamental constraint), visual-servoing control is used to provide a whole new set of useful services such as grasping objects in order to bring them to the end-user (if they are too heavy, too far, high or low), or cleaning the table after lunch. Using a parallel crane robot, we are able to cover a large workspace, the vision-control allowing us to obtain the precision required by the manipulation of daily-life objects. The collaborative implementation of the vision and the kinematic control of the robot gives us a way to make best use of the advantages of both parts, while overcoming their respective drawbacks.

This project is supported by the large-scale initiative PAL.

Experimentation showed that we are able to provide a much better accuracy and repeatability using visual-servoing. However, the velocity of the process is slowed because of several encountered problems :

- when there are changes in the distribution of tension between the wires, oscillations are occurring on the end-effector, affecting the movement of the camera in such a way that we can not rely on the measurements
- the methods used to first detect the object are not satisfactory. Also, the actual segmentation is not robust to luminance changes, the target may thus be lost during the process.

In order to overcome the first problem, we are working on an algorithm able to determinate the best sequence of configurations (distribution of tension) : we can avoid singularities and provide a more stable trajectory. The second problem has yet to be solved : we are at the moment looking into several methods, using for example k-nearest neighbors algorithms with different color spaces, gradient-based information and morphological preprocessing.

Finally, we experimented our device with others technologies developed within the context of PAL, in a full-scaled apartment located in Nancy (Loria-Inria).

### 6.1.2. 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 [22], [23], [19].

#### 6.1.2.1. Assessment of elderly frailty

**Participants:** Karim Bakal, Jean-Pierre Merlet.

The assessment of elderly frailty is a difficult concept because it involves the physical capacities of a person and its environment (health-care services, families, funds...). To evaluate the physical abilities, biomechanics tests can be underwent on the upper limb, lower limb or the whole body. In particularly, the motricity of the upper limb can be measured in terms of range of motion, velocity, acceleration or forces.



To analyze the velocity of the loads in the upper limb, a polytope interpretation is used. Currently the force polytope at the hand is calculated from the torques  $\tau$  measured at each joint (shoulder, elbow and wrist) by a dynamometer (Biodex III, Biodex Medical Systems). But because of the redundancy of the upper limb (7 degrees of freedom), the dynamic equation ( $\tau = \mathbf{J}^T F$ ) is difficult to solve. To find the minimal and maximal forces  $F$  that can be exerted at the hand from the measured torques, we may use the jacobian pseudo-inverse with the method of Chiacchio but this method is not well suited to manage the large uncertainties in the measurements. In the a reverse approach, the force at the hand will be measured by a 6-axis load sensor and the minimal and maximal joint torques will be computed by using interval analysis and compared with the measurements of the Biodex.

Moreover this analysis of the force capacities in the upper limb need to be connected to the daily activities or usual motion test monitored by the medical services. Therefore, a review of tests and questionnaires regularly used to measure the physical capacities has been performed. This review gather the type of mark, the exercises and the used sensors that can be employed in future experimentation. Also, this review will be discussed with medical staff to highlight relevant activities.

#### 6.1.2.2. *Walking analysis*

**Participants:** Claire Dune, Ting Wang, Jean-Pierre Merlet [correspondant].

In the period 2009-2013 we have conducted in collaboration with Nice hospital a large experiment involving 54 subjects (30 elderly and 24 young adults) for determining walking pattern of elderly people using our instrumented walker ANG-light. We have started the processing of this large amount of data we some interesting results [25]:

- a classical walking test is the 10 meter walking test: the subject is asked to perform a 10m straight line trajectory and the result is the total time. Such test may have large consequences as it is used to determine the autonomy level and the resulting financial aid. Our test has surprisingly shown that when using a walker elderly people are usually faster than young adults
- on the other hand the maximal deviation with respect to the desired trajectory is much smaller for young adults than for elderly one. Furthermore few elderly have the same deviation and it may be considered as a signature of the walking pattern that is worth measuring

Our objective is now to analyze the maneuvers (half-turn and round-about) and to compare/complement the data with the one obtained with a Kinect. A long term objective is also to implement a model of a human walking with a walker and to use this model for an inverse calculation: measuring walking patterns indicators with the walker and calculating these indicators when not using the walker.

#### 6.1.3. *Experimental calibration of a high-accuracy space telescope*

**Participants:** Thibault Gayral, David Daney, Jean-Pierre Merlet.

A collaborative work began in October 2010 with Thales Alenia Space on the calibration of the mechanical structure of a space telescope. Its architecture is based on a parallel manipulator (type active wrist 6-PUS) used to correct the relative position of two mirrors. The aim is to reach a micrometer accuracy in order to obtain a suitable quality of the images provided by the telescope. Thus, a complete model of the space telescope needs to be developed and validated through calibration. Since high velocity is not required in such an application, the dynamic effects can be neglected and only geometric and/or static calibration has to be considered.

For the geometric models, measurements for calibration were performed in a clean room under controlled pressure, temperature and humidity conditions to minimize the influence of the non-geometric errors. Thus, two possible static inaccuracy sources were identified and modeled: one from the deformation of the mobile platform and the other resulting from the behavior of the flexure joints. Three incremental models of the flexure joints were developed and compared: a spherical joint model, a model issued from the beam theory and a stiffness model. Results of calibration using an accurate measurement system of photogrammetry showed that the flexure joints can be modeled by perfect spherical joints due to the small workspace of the telescope. Concerning the mobile platform deformation, two models were developed. With those models, a positioning accuracy of some micrometers was finally reached after calibration with only position and orientation measurements of the mobile platform.



Then, opto-mechanical models were developed considering experimental measurements by imaging on the prototype of the space telescope. The optical defects were analyzed considering Zernike polynomials. The aim of optical calibration was to minimize the coefficients of the Zernike polynomials in order to improve the optical properties of the space telescope. Results of calibration were studied in order to perform a proper choice of the opto-mechanical models. Finally, the optical quality was improved after calibration. This validates the fact that the telescope can be calibrated directly in space, after its deployment, with only the provided information. A second campaign of measurements by imaging was programmed to finely adjust the opto-mechanical model parameters.

## 6.2. Miscellaneous results

### 6.2.1. Symbolic tools for modeling and simulation

**Participant:** Yves Papegay.

This activity is the main part of a long-term ongoing collaboration with Airbus whose goal is to directly translate the conceptual work of aeronautics engineers into digital simulators to accelerate aircraft design.

An extensive modeling and simulation platform has been designed which includes a dedicated modeling language for the description of aircraft dynamics models in term of formulae and algorithms, and a symbolic compiler producing as target an efficient numerical simulation code ready to be plugged into a flight simulator, as well as a formatted documentation compliant with industrial requirements of corporate memory.

Implementation of this platform is a modeling and simulation environment based on symbolic computation tools. It contains several components :

- a model editor, that makes it possible and easy to enter the whole set of equations describing large and complex industrial models,
- an highly interactive and modular evaluation workbench allowing to simulate the models and to visualize the results inside the modeling environment with the benefits for the designer of being able to directly use all its computational functionalities.
- a C code generator which, using these models, automatically generates the numerical real-time simulation engines
- a technical documentation generator

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.

However, in 2013, we have worked on several enhancements and extension of functionalities, namely to ease the integration of our environment into the airbus toolbox. Developer level know-how has been transferred to a software company in charge of industrialization and maintenance of the modeling and simulation environment.

### 6.2.2. Multi-agent aircraft design

**Participant:** Yves Papegay.

The modeling environment described in the previous section is used, in collaboration with other teams at Airbus, in the framework of the ID4CS project founded by ANR and dedicated to multi-agent optimization of large scale system.

Several models of aircraft engines and of aircrafts have been developed as user cases for the project.

2013 is the last year of the project when agent code based on models has been used to solve several practical optimization problems based on these models.

### 6.2.3. Equilibrium strategies for linked Electricity and CO2 markets

**Participant:** Odile Pourtallier.

In collaboration with M. Bossy (Inria -TOSCA Team) and N. Maïzi (CMA - Mines Paristech) O. Pourtallier we have pursued our work on CO2 and electricity market coupling.

The aim of this work is to develop analytic tools, in order to design a relevant mechanism for carbon markets, where relevant refers to emission reduction. In the context of electricity, the number of producers is limited, a standard game theory approach applies. The producers are considered as players behaving on the two financial markets represented here by carbon and electricity. We establish a Nash equilibrium for this non-cooperative  $J$ -player game through a coupling mechanism between the two markets.

The original idea comes from the French electricity sector, where the spot electricity market is often used to satisfy peak demand. Producers behavior is demand driven and linked to the maximum level of electricity production. Each producer strives to maximize its market share. In the meantime, it has to manage the environmental burden associated with its electricity production through a mechanism inspired by the EU ETS (European Emission Trading System) framework : each producer emission level must be balanced by a permit or through the payment of a penalty. Emission permit allocations are simulated through a carbon market that allows the producers to buy the allowances at an auction.

Based on a static elastic demand curve (referring to the times stages in an organized electricity market, mainly day-ahead and intra-day), we solve the local problem of establishing a non-cooperative Nash equilibrium for the two coupled markets.

## IMARA Project-Team

### 6. New Results

#### 6.1. ABV

**Participants:** Hao Li, Paulo Lopes Resende, Evangeline Pollard, Joshué Pérez Rastelli, Fawzi Nashashibi.

The ABV project builds on the HAVEit philosophy (a previous IMARA project for high speed automation) by offering higher levels of automation on highways and organizing the cooperation between human and system along novel automation levels. It differs from HAVEit by focusing on congested traffic at speeds below 50 km/h and adding fully automated driving to the automation spectrum. By automatically following congested traffic, the ABV system relieves the human driver from monotonous tasks. During fully automated driving, the human driver is not required to monitor the system, but has to take over control at the end of the application zone. Real experiments on a prototype vehicle have been conducted. The experiment objective was to realize several use-cases: lane following, changing of lane, overtaking, ACC and emergency braking. All these maneuvers have been successfully conducted several times on the Satory tracks (cf. [46], [41] for more details) during the final event of the project which took place late March 2013.

#### 6.2. Urban Autonomous Driving

**Participants:** Evangeline Pollard, Guillaume Tréhard, Fawzi Nashashibi.

Beyond low speed automation, IMARA is tackling a very important issue for autonomous driving on open roads, which is: dealing with intersections. In collaboration with Valeo, Imara wants to provide innovative way to safely cross any kind of intersections for an autonomous vehicle in a urban context and without communication. The goal is to deal with intersection with different shapes, (roundabout, T junctions , X junctions, *etc.*), with different rules, specific (traffic lights, main road...) or not ("priority to the right" in France), with different traffic (busy or empty).

#### 6.3. 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 questions of whether the Wi-Fi-based V2P communications meet the requirements of the pedestrian safety application. This year, we studied the performances of the V2P communications especially 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.

#### 6.4. Visible light communications for platooning control

**Participants:** Mohammad Abualhoul, Oyunchimeg Shagdar, Mohamed Marouf, Fawzi Nashashibi.

While V2V communications is requisite for platooning stability, the existing radio communications technologies suffer from poor performance in highly dense road scenarios, which are exactly to be created for platooning. Targeting this issue, we study the applicability of visible light communications (VLC) for information exchange between the platoon members [20], [35]. Because the existing studies on VLC mainly focus on indoor applications or for communications from traffic light to vehicle, the performances of VLC for V2V is not clear. In this work, we develop a complete VLC channel and noise model by taking account of the key parameters including background noise and incidence angle. Our studies show that it is feasible to achieve up to 7 meters line of sight communication range even in the presence of optical noise at significant levels and with up to 60 degree of road curvature.

## 6.5. ITS-G5 for road safety and efficiency applications

**Participants:** Oyunchimeg Shagdar, Younes Bouchaala, Mohammad Abualhoul, Manabu Tsukada, Thierry Ernst.

To support V2V and V2I communications for road safety and efficiency applications, ETSI standardized ITS-G5 technology. One of key objectives of the SCORE@F project is to study the performance of ITS-G5 in real-world scenarios and demonstrate its applicability to road safety and efficiency applications. Under the scope of the SCORE@F project, we studied the performances of ITS-G5 for both the V2V and V2I communications based on field tests and theoretical studies with emphasis on the effects of channel in combination with MAC and some parameters of car traffic [32]. An important insight achieved from the study is that in addition to the distance dependent pathloss, the signal fading and road traffic characteristics provide significant impacts on the reliability of ITS-G5.

We also study the performance of the ITS-G5 medium access control protocol for realistic autonomous driving applications especially to seek answers to the questions of whether the IEEE 802.11p can support merging control and how the communications performance is translated into that of CACC (Cooperative Adaptive Cruise Control) [33]. The study discloses several useful insights including packet inter-arrival time and throughput but not packet delivery ratio, gives good indications of the CACC performance; the V2I communications structure is preferred over the V2V structure for CACC.

Finally, we demonstrate the low latency video streaming over ITS-G5 to support platoon and reverse parking maneuvers [21].

## 6.6. Cooperative driving

**Participants:** Joshué Pérez Rastelli, Fawzi Nashashibi.

In the scope of the French project “Co-Drive” one task assigned to Inria was the development of a smart controller capable of driving the vehicle, allowing it to perform optimal traversal of traffic lights in order to reduce vehicle accelerations and thus the gas emissions. This controller needs remote information regarding the traffic lights’ status, the distance to it and the time needed to reach it.

Three input variables, which are the traffic light times, red light, green light and the distance to interception (DTI), were defined in fuzzy logic tool [37].

Two variables are used for the traffic light (Red and Green), where each of them has defined two completely symmetrical membership functions covering all the possible inputs. In this application the time cycle of the lights are 30 seconds for green and 20 seconds for red. The values of input membership functions were defined considering these times.

The DTI membership function (see Figure 1 ) gives more weight to the distance when the vehicle is closer at the intersection. In this situation, the vehicle can be inside the *short* or the *middle* label, because in these cases the response has to be faster than in the case where the vehicle is in the *long* label. The cross rule base, based on driver knowledge when the vehicle is arriving to an intersection, are defined using natural language.

Some Simulations were performed to validate the controller. However, the final implementation will be presented in 2014 during the final event of Co-Drive Project.



Figure 1. Codrive: input variables for the speed reference fuzzy controller

## 6.7. Intelligent Planning algorithm using Bezier curves

**Participants:** Joshué Pérez Rastelli, Fawzi Nashashibi.

The Bezier curve is the heart of the Local Planning, which allows a fast trajectory computation in order to send the trajectory in real-time to the controller stage. This method has been recently used in robot mobile solutions due to its versatility and simplicity for intersections.

We have proposed a novel method for the generation of control points for two distinct road configurations: roundabouts and a standard intersections. If an intersection is being dealt with, the control points will be generated based on the reference path given by the Global Planner.

The experiments we made presented several urban intersections. Figure 2 shows the whole generated path with four intersections and a roundabout, using the global map. A comparison with different methods is drawn. The first one (thin line) is based on the static method used in [22], which sets the control points by hand. In this case we can see how sometimes the path passes over the sidewalk. The second experiment (dotted line) is using the same previous method, but modifying the distance used to position the control points, in order to obtain a path into the road. The third method (thick line) is the Intelligent Planning algorithm. As we can see in the figure, the automatic algorithm sets the control points of Bezier (based on the convex hull property) achieving a smooth path, without going over sidewalks or obstacles.

## 6.8. Ontologies

**Participants:** Evangeline Pollard, Philippe Morignot, Fawzi Nashashibi.

Full autonomy of ground vehicles is a major goal of the ITS (Intelligent Transportation Systems) community. However, reaching such highest autonomy level in all situations (weather, traffic, . . . ) is seen as impossible in practice, despite recent results regarding driverless cars (e.g., Google Cars). In addition, an automated vehicle should also self-assess its own perception abilities, and not only perceive its environment. In this new research axis, we propose an intermediate approach towards full automation, by defining a spectrum of automation



*Figure 2. Generated path using different methods*

levels, from fully manual (the car is driven by a driver) to fully automated (the car is driven by a computer), based on an ontological model for representing knowledge. We also propose a second ontology for situation assessment (what does the automated car perceive?), including the sensors/actuators state, environmental conditions and driver's state. Finally, we also define inference rules to link the situation assessment ontology to the automation level one [24].

## 6.9. Communications and Management Control for Cooperative Vehicular Systems

**Participants:** Ines Ben Jemaa, Oyunchimeg Shagdar, Arnaud de La Fortelle.

One of the attractive applications of electric autonomous vehicles is electric automated Car Sharing service, where on-demand passenger transportation is provided by a set of automated vehicles and a control center, which is installed in the Internet. Data transmission from the control center to the set of vehicles requires an efficient multicast data delivery, i.e. multi-cast routing. The conventional multicast routing in the Internet is based on protocols such as Protocol Independent Multicast (PIM), which relies on a tree structure to deliver packets from the source to the destinations. Thanks to the fixed topology of the Internet, it is possible to build a large and stable multicast trees. However, due to the highly mobile nature of vehicular networks, it is not clear how stable and large can be such trees in vehicular environments. This year, we studied the stability of multicast trees for data flows from the Internet to a set of vehicles [38], [36]. Our study shows that the stability of multicast tree largely depends on the relative velocity (inter-vehicle) and the road density but not directly on the road shape or moving direction. Based on our study we are developing a mobility aware multicast routing protocol, which constructs its tree based on the vehicles' mobility dynamics and the road condition.

## 6.10. New urban transportation platforms: Inria's Cybus

**Participants:** François Charlot, Joshué Pérez Rastelli, Fawzi Nashashibi, Paulo Lopes Resende, Michel Parent, Armand Yvet.

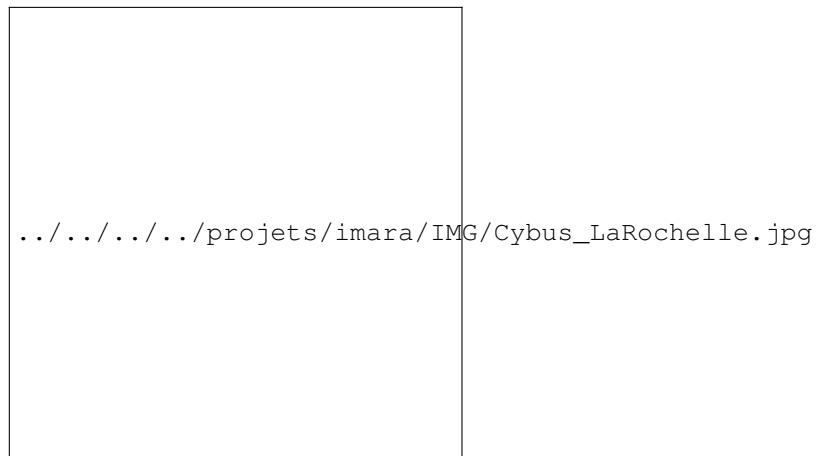


Figure 3. The Cybus operated at La Rochelle City during 3 months as a free transport service.

*Cybus* is the best known prototyping and demonstration platform designed at Inria. Apart from the chassis and engines, the whole hardware and software systems were developed thanks to IMARA's researchers and engineers talents. These electric vehicles are based on a Yamaha chassis but the embedded intelligence is the result of two years of development.

Much of the perception and control software has been improved. New guidance functionalities were developed this year, mainly with the introduction of stereovision-based SLAM, and Bezier curve in path planning generation. The platforms developed here (*Cybus*) will be demonstrated in the context of the EU CityMobil-2 project. This time real operational mobility services demonstrations will be extended to 6-12 months in selected European cities! Other showcases are expected to take place in Asian cities in 2014.

### 6.11. Real-time visual perception: detection and localization of static and moving objects from a moving stereo rig

**Participants:** Benjamin Lefaudeux, Fawzi Nashashibi.

Perception of the surrounding environment is one of the many tasks an automated vehicle has to achieve in complex and ever-changing surroundings. This, typically includes several distinct sub-tasks, such as map-building, localization, static obstacles and moving objects detection and identification. Some of these tasks are nowadays very well known, such as the map-building process which has been extensively investigated in the last decade ; whereas the perception, localization and classification of moving objects from an equally moving vehicle are in many aspects a work in progress. The objective of the PhD thesis of Benjamin Lefaudeux was to propose a vision-based approach built on the extensive tracking of numerous visual features over time, from a stereo-vision pair.

Through on-the-fly environment 3D reconstruction, based on visual clues, we proposed an integrated method to detect and localize static and moving obstacles, whose position, orientation and speed vector is estimated. Our implementation runs in real-time depending on the number of processed points, and should in the future be enclosed in a more complete, probabilistic pipeline. The complete achievements are described in the thesis of Benjamin Lefaudeux ([8] defended on September 30th) with very interesting and competitive results obtained with international benchmarks (cf. Figure 4) and on the real vehicles of IMARA.

### 6.12. Belief propagation inference for traffic prediction

**Participants:** Cyril Furtlehner, Jean-Marc Lasgouttes, Victorin Martin.



Figure 4. Left: A single camera view from the KITTI sequence. Right: A bird view of the scene as modeled by the system: point cloud and estimated trajectory.

This work [55] 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's highlights are

- Victorin Martin has defended his PhD thesis entitled “Modélisation probabiliste et inférence par l’algorithme Belief Propagation” [9] at Mines-ParisTech on May 23.
- The work about the theoretical aspects of encoding real valued variables into a binary Ising model has been published as a research report [44] and submitted for publication.

### 6.13. Sparse covariance inverse estimate for Gaussian Markov Random Field

**Participants:** Cyril Furtlehner, Jean-Marc Lasgouttes, Victorin Martin.

We investigate the problem of Gaussian Markov random field (GMRF) selection under the constraint that the model is suitable for Gaussian belief propagation (GaBP) inference. We develop a method based on iterative proportional scaling (IPS) to incrementally select optimal GMRF factors, while maintaining GaBP compatibility. Besides the intrinsic sparsity-inducing capability, the proposed method is indeed sufficiently flexible to incorporate various spectral constraints like e.g. walk summability (WS) to insure the compatibility of the solutions with Gaussian Belief Propagation inference. Experimental tests on various datasets with refined  $L_0$  or  $L_1$  regularized sparse inverse estimate indicate that this approach is competitive and provides us with useful alternatives to traditional sparsity-inducing penalizations norms, giving more freedom in the graph structure selection process with no additional computational cost.



## 6.14. Evaluation of dual mode transport system by event-driven simulation

**Participants:** Arnaud de La Fortelle, Jean-Marc Lasgouttes, Thomas Liennard.

The European project CATS — City Alternative Transport System — is developing and evaluating a new vehicle system using a single type of vehicle for two different usages: individual use or collective transport. Real experiments will necessarily take place with a limited number of vehicles and stations. Hence, there is a need for evaluation using simulations.

We are developing a discrete events simulator for that purpose, which model relies on an adapted events/decision graph. The new feature of this model is the way we deal with two modes that can be extended to many other modes. This work therefore shows on a concrete example a method to efficiently merge multiple modes into one model.

This year has seen a partial rewrite of the simulator in order to make it more generic and handle the new setting of the CATS project with automated vehicles.

## 6.15. 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* [11].

## 6.16. Analytic properties of random walks in the quarter plane

**Participant:** Guy Fayolle.

In collaboration with K. Raschel (CNRS, Université F. Rabelais à Tours), we pursued the works initiated these last three years in two main directions.

### 6.16.1. The group and zero drift case

In several recent studies on random walks with small jumps in the quarter plane, it has been noticed that the so-called *group of the walk* governs the behavior of a number of quantities, in particular through its *order*. When the *drift* of the random walk is equal to 0, we have provided an effective criterion (see RA 2012) giving the order of this group. More generally, we showed that in all cases where the *genus* of the algebraic curve defined by the so-called *kernel* is 0, the group is infinite, except precisely for the zero drift case, where finiteness is quite possible.

This year, we investigated new proofs of this results, which could lead to an explicit tractable criterion for the finiteness of the group, which a priori, as shown in [2] involves a ratio of elliptic integrals.

### 6.16.2. Counting and asymptotics

The enumeration of planar lattice walks is a classical topic in combinatorics. For a given set  $\mathcal{S}$  of allowed unit jumps (or steps), it is a matter of *counting the number of paths* starting from some point and ending at some arbitrary point in a given time, and possibly restricted to some regions of the plane.

Like in the probabilistic context, a common way of attacking these problems relies on the following analytic approach. Let  $f(i, j, k)$  denote the number of paths in  $\mathbb{Z}_+^2$  starting from  $(0, 0)$  and ending at  $(i, j)$  at time  $k$ . In the case of small jumps (size at most one), the corresponding CGF

$$F(x, y, z) = \sum_{i, j, k \geq 0} f(i, j, k) x^i y^j z^k$$

satisfies the functional equation

$$K(x, y, z)F(x, y, z) = c(x)F(x, 0, z) + \tilde{c}(y)F(0, y, z) + c_0(x, y),$$

where  $x, y, z$  are complex variables,  $K(x, y, z)$  is a polynomial of degree 2 (both in  $x$  and  $y$ ), and linear in the time variable  $z$  which plays somehow the role of a parameter. The question of the type of the associated counting generating functions, rational, algebraic, or holonomic (i.e. solution of a linear differential equation with polynomial coefficients), was solved whenever the group is *finite* (see RA 2010). When the group is infinite, the problem is still largely open.

The nature of the singularities of the function  $F$  plays a key role for this classification. Starting from our study [54], we proved in various cases that the first singularities of  $F(1, 0, z)$  are either polar or correspond to a value  $z_g$  for which the genus of the algebraic curve  $K(x, y, z) = 0$  passes from 1 to 0 (i.e. a torus becomes a sphere).

### 6.16.3. Harmonic functions and more general jumps

The determination of Martin boundaries in the case of random walks is a longstanding problem, solved only in special situations. For homogeneous random walks in the quarter plane, stopped on the boundary (the axes), with upward jumps of size 1, and arbitrary downward jumps of size  $d$ , it turns out that the computation of harmonic functions is here plainly equivalent to find a positive function  $H$  satisfying a functional equation of the form

$$L(x, y)H(x, y) = L(x, 0)H(x, 0) + L(0, y)H(0, y) - L(0, 0)H(0, 0).$$

Here the chief difficulty to make the reduction to a boundary value problem is to analyze the algebraic curve  $L(x, y) = 0$ , which might be of arbitrary genus. Some examples lead us to conjecture the existence of a *single real cut* inside the unit disk, which should allow to get integral form solution.

### 6.16.4. Correction of papers

Guy Fayolle found important errors in several articles dealing with models involving random walks in the quarter plane. This is the object of the letter to the editors [10]. The Concerned authors are currently preparing corrected versions.

## E-MOTION Project-Team

# 5. New Results

## 5.1. Perception and Situation Awareness in Dynamic Environments

### 5.1.1. Sensor Fusion for state parameters identification

**Participants:** Agostino Martinelli, Chiara Troiani.

#### 5.1.1.1. General theoretical results

We continued to investigate the visual-inertial structure from motion problem by further addressing the two important issues of observability and resolvability in closed form. Regarding the first issue, we extended our previous results published last year on the journal of Transaction on Robotics [44] by investigating the case when the visual sensor is not extrinsically calibrated. In order to deal with this case, we must augment the state to be estimated by including all the parameters that characterize the extrinsic camera calibration, i.e., the six parameters that describe the relative transformation between the frame attached to the camera and the frame attached to the Inertial Measurement Unit (IMU). On the other hand, because of the larger size of the resulting state, it became prohibitive a direct application of the method that we introduced two years ago (see [43]) in order to discover the observability properties for this new state. For this reason, our first novel contribution during this year was the introduction of new methodologies able to significantly reduce the computational burden demanded by the implementation of the method in [43]. These methodologies have been published in [22] and a deeper description of their use is currently under revision on the journal Foundations and Trends in Robotics. The new results obtained by using these methodologies basically state that also the new six parameters that describe the camera extrinsic calibration are observable. Finally, we started a new research that in the literature is known as the Unknown Input Observability (UIO) and it is investigated by the automatic control community. We started this new research since we investigated the observability properties of the visual inertial structure from motion as the number of inertial sensors is reduced. Specifically, instead of considering the standard formulation, which assumes a monocular camera, three orthogonal accelerometers and three orthogonal gyroscopes, the considered sensor suit only consists of a monocular camera and one or two accelerometers. This analysis has never been provided before. A preliminary investigation seems to prove 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, as the camera is extrinsically calibrated, the system maintains the same observability properties as in the standard case. This contribution clearly shows that the information provided by a monocular camera, three accelerometers and three gyroscopes is redundant. Additionally, it provides a new perspective in the framework of neuroscience to the process of vestibular and visual integration for depth perception and self motion perception. Indeed, the vestibular system, which provides balance in most mammals, consists of two organs (the *utricle* and the *sacculle*) able to sense the acceleration only along two independent axes (and not three). In order to analyze these systems with a reduced number of inertial sensors, we had to consider control systems where some of the inputs are unknown. Indeed, the visual-inertial structure from motion problem can be characterized by a control system where the inputs are known thanks to the inertial sensors. Hence, to deal with the visual-inertial structure from motion as the number of inertial sensors is reduced, we had to introduce a new method able to address the more general UIO problem. We believe that our solution to the UIO is general and this is the reason because we started this new research domain in control theory. Preliminary results are currently under revision on the journal Foundations and Trends in Robotics and we also plan to present them at the next ICRA conference. Regarding the second issue, i.e., the problem resolvability in closed form, a new simple closed form solution to visual-inertial structure from motion has been derived. This solution expresses the structure of the scene and the motion only in terms of the visual and inertial measurements collected during a short time interval. This allowed us to introduce deterministic algorithms able to simultaneously determine the structure

of the scene together with the motion without the need for any initialization or prior knowledge. Additionally, the closed-form solution allowed us to identify the conditions under which the visual-inertial structure from motion has a finite number of solutions. Specifically, it is shown that the problem can have a unique solution, two distinct solutions or infinite solutions depending on the trajectory, on the number of point-features and on their arrangement in the 3D space and on the number of camera images. All the results have been published on the international journal of Computer Vision [15].

#### 5.1.1.2. Applications with a Micro Aerial Vehicle

We introduced a new method to localize a micro aerial vehicle (MAV) in GPS denied environments and without the usage of any known pattern [26]. The method exploits the planar ground assumption and only uses the data provided by a monocular camera and an inertial measurement unit. It is based on a closed solution which provides the vehicle pose from a single camera image, once the roll and the pitch angles are obtained by the inertial measurements. Specifically, the vehicle position and attitude can uniquely be determined by having two point features. However, the precision is significantly improved by using three point features. The closed form solution makes the method very simple in terms of computational cost and therefore very suitable for real time implementation. Additionally, because of this closed solution, the method does not need any initialization. Results of experimentation show the effectiveness of the proposed approach.

We proposed a novel method to estimate the relative motion between two consecutive camera views, which only requires the observation of a single feature in the scene and the knowledge of the angular rates from an inertial measurement unit, under the assumption that the local camera motion lies in a plane perpendicular to the gravity vector [27]. Using this 1-point motion parametrization, we provide two 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. We test the proposed approaches on both synthetic and real data, using video footage from a small flying quadrotor. We show that our methods outperform standard RANSAC-based implementations by up to two orders of magnitude in speed, while being able to identify the majority of the inliers.

#### 5.1.2. A new formulation of the Bayesian Occupancy Filter : an hybrid sampling based framework

**Participants:** Lukas Rummelhard, Amaury Nègre.

The Bayesian Occupancy Filter (BOF) is a discretized grid structure based bayesian algorithm, in which the environment is subdivided 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 developed. 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 main idea of this new representation is to mix two forms of sampling : a uniform one, represented as a grid, for the static objects and the empty areas, and a non uniform one, based on particles drawn from dynamic regions. The motion field in a cell will be 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, as well as 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 first results are stimulating, in term of occupancy evaluation and mostly in term of velocity prediction, being way more accurate and responsive than the older version. Those improvements will soon be presented in detail in upcoming papers, one being currently in the process of redaction.

### 5.1.3. DATMO

**Participants:** Dung Vu, Mathias Perrollaz, Amaury Nègre.

In the current work, we have been developing a general framework for tracking multiple targets from lidar data.

In the past decades, multiple target tracking has been an active research topic. When object observations are known, object tracking becomes a data association (DA) problem. Among popular DA methods, multiple hypothesis tracking (MHT) is widely used. MHT is a multi-frame tracking method that is capable of handling ambiguities in data association by propagating hypotheses until they can be solved when enough observations are collected. The main disadvantage of MHT is its computational complexity since the number of hypotheses grows exponentially over time. The joint probabilistic data association (JPDA) filter is more efficient but prone to make erroneous decision since only single frame is considered and the association made in the past is not reversible. Other sequential approaches using particle filters share the same weakness that they cannot reverse time back when ambiguities exist. All DA approaches mentioned above requires a strong assumption of one-to-one mapping between targets and observations which is usually violated in real environments. For instance, a single object can be seen by several observations due to occlusion, or multiple moving objects can be merged into a single observation when moving closely.

In the research, we propose a new data association approach that deals with split/merge nature of object observations. In addition, our approach also tackles ambiguities by taking into account a sequence of observations in a sliding window of frames. To avoid the high computational complexity, a very efficient Markov Chain Monte Carlo (MCMC) technique is proposed to sample and search for the optimum solution in the spatio-temporal solution space. Moreover, various aspects including prior information, object model, motion model and measurement model are explicitly integrated in a theoretically sound framework.

### 5.1.4. Visual recognition for intelligent vehicles

**Participants:** Alexandros Makris, Mathias Perrollaz, Christian Laugier.

We have developed an object class recognition method. The method uses local image features and follows the part-based detection approach. It fuses intensity and depth information in a probabilistic framework. The depth of each local feature is used to weigh the probability of finding the object at a given distance. To train the system for an object class, only a database of images annotated with bounding boxes is required, thus automatizing the extension of the system to different object classes. We apply our method to the problem of detecting vehicles from a moving platform. The experiments with a data set of stereo images in an urban environment show a significant improvement in performance when using both information modalities.

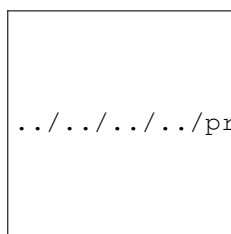
In 2013, the method has been published in IEEE Transactions on Intelligent Transportation Systems [14].



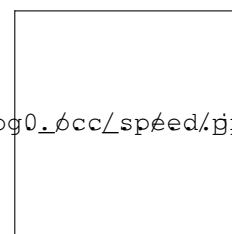
(a) Camera image, two-way street



(b) Camera image, highway



(c) Occupancy and velocity result, two-way street



(d) Occupancy and velocity result, highway

Figure 1. Results of the new algorithm : the free cells are black, the occupied ones are white. Unknown areas are grey. When a cell is seen as dynamic, a vector representing the average of the associated motion field is drawn in red.

### 5.1.5. Experimental platform for road perception

#### 5.1.5.1. Experimental platform material description

Our experimental platform for road perception is shown in Figure 2 . This platform is a commercial Lexus car LS600h equipped with a variety of sensor, including two IBEO Lux lidars placed toward the edges of the front bumper, a TYZX stereo camera plus a high resolution color camera situated behind the windshield, and an Xsens MTi-G inertial sensor with GPS. To do online data computation and data acquisition, a standard computer is located on the back of the car.

This platform allows us to conduct experimentation and data acquisition in various road environments (country roads, downtown and highway), at different time of the day, with various driving situations (light traffic, dense traffic, traffic jams).



Figure 2. Lexus LS600h car equipped with two IBEO Lux lidars, a stereo plus a monocular camera, and an Xsens MTi-G inertial sensor with GPS.

#### 5.1.5.2. Software architecture

The perception and situation awareness software architecture is integrated in the ROS framework. ROS (<http://www.ros.org>) is an open source robotics middleware designed to be distributed and modular. For the Lexus platform, we developed a set of ROS module for each sensor and for each perception component. Each

perception module can be dynamically connected with the required drivers or other perception modules. The main architecture of the perception components is illustrated on Figure 3 .



Figure 3. Architecture of the main perception components in the Lexus platform.

#### 5.1.6. Software and Hardware Integration for Embedded Bayesian Perception

**Participants:** Mathias Perrollaz, Christian Laugier, Qadeer Baig, Dizan Vasquez, Lukas Rummelhard, Amaury Nègre.

The objective of this recently started research work is to re-design in a highly parallel fashion our Bayesian Perception approach for dynamic environments (based on the BOF concept), in order to deeply integrate the software components into new multi-processor hardware boards. The goal is to miniaturize the software/hardware



perception system (i.e., to reduce the size, the load, the energy consumption and the cost, while increasing the efficiency of the system).

For supported this research, we began to work in the “Perfect” project. This project, included in the IRT-Nano program, involves the CEA-LETI DACLE lab and ST-Microelectronics. Perfect is focusing onto the second integration objectives (6 years) and the development of integrated open platforms in the domain of transportation (vehicle and infrastructure) and in a second step in the domain of health sector (mobility of elderly and handicapped people, monitoring of elderly people at home...). The objective of e-Motion in this project is to transfer and port its main Bayesian perception modules from traditional computing system to embedded low-power multi-processors board. The targeted board is a STHorm from ST Microelectronics which is has a many-core architecture with a very low consumption. In 2013 we worked with the CEA to obtain a first implementation of the Bayesian occupancy grid filter on STHorm. Those preliminary results demonstrated the feasibility of the concepts but highlighted some key points to improve such as the memory footprint we need to reduce to obtain real-time accurate results.



*Figure 4. First objective for software/hardware of Bayesian perception : developing and using multiple processor board from mobile technologies. The approach will be validated with real demonstrators..*

## 5.2. Dynamic Change Prediction and Situation Awareness

### 5.2.1. Vision-based Lane Tracker

**Participants:** Mathias Perrollaz, Amaury Nègre.

In the field of vehicle risk assessment system, vehicle to road localization is an essential information to predict drivers behaviors as well as collision risk evaluation. To achieve this task, we have developed a vision based lane tracker to estimate the geometry of the lane using the line markers. Previous development was based on a Monte-Carlo particle filter to estimate simultaneously the road plane orientation, the lane curvature and the camera position. To perform the particle evaluation, the algorithm first process a ridge extraction of the camera image and then projected the left and right marker line represented by the particle on the ridge image.

The first improvement of the tracker consists on dynamically adapting the scale of the ridge filter to improve the efficiency and the precision of the particle evaluation. The second improvement is capability the perform multi-lane tracking for example in highway environment. To solve this problem, the position of the lane is added to the state and the particle evaluation consider the total number of line marker (a-priori known). Figure 5 shows the results of the lane tracker program on a highway environment.

### 5.2.2. Vision-based Lane Change Prediction

**Participants:** Suryansh Kumar, Dizan Vasquez, Mathias Perrollaz, Stephanie Lefevre, Amaury Nègre, Maiwen Gault.

For both Advanced Driving Assistance Systems and Autonomous Vehicles, it is very important to have the capability of predicting and understanding the driver's behavior. This work addresses this subject in a bottom-up fashion by first detecting low-level "atomic" maneuvers which can be used as a building block for more complex behavior. Concretely, we have developed a learning-based approach that uses lane tracking data to predict lane changes.

Most works in the literature address this as a classification problem, and often use some version of Support Vector Machines (SVM) to solve it. The problem with this approach is that it is sensitive to noise and can yield high-frequency oscillations in the obtained predictions, moreover, they do not provide any information concerning the Time To Change (TTC). Other approaches use a filtering approaches, using Hidden Markov Models (HMM), for example. Although they produce smoother predictions and, in some cases, even a TTC estimate, some studies [33] suggest that HMM-based approaches are less accurate than those based on SVMs.

Our work combines the advantages of both approaches in a hierarchical fashion. First, lane tracking data (i.e. lane-relative yaw, lane-relative lateral position and their first-order derivatives) are used as an input of a multi-class SVM. Then, the Bradley-Terry model is used to translate the SVM output into a probability which is used as the observation model of a Bayesian filter (Fig. 6).

This work has been published in the Intelligent Vehicles conference [18]. Since then, we have worked on an improved HMM-only approach which addresses the shortcomings of similar approaches by using a continuous observation model. In our preliminary experiments, this approach leads to improved predictions over the hybrid one.

### 5.2.3. Feature-based human behavior modeling

**Participants:** Suryansh Kumar, Yufeng Yu, Dizan Vasquez.

When it comes to modeling and learning complex human behavior, the preferred approach in the literature is to try to learn the typical motion patterns that people or vehicles get engaged into in a given environment. This has, however, a major drawback: the learned patterns only apply to the environment where they have been learned. This means that, for a new environment, previous knowledge cannot be used and patterns should be learned from scratch.

This situation has been recently addressed [49] by a family of approaches which rely on two complementary hypotheses:

- *Agents behave like planners.* The idea is that when people or other intelligent agents move around, they plan in order to minimize a cost function. Thus, if this function is known, it fully determines the agent's behavior.



(a) Camera image



(b) Particles cloud

(c) Lane tracker result

*Figure 5. Visual Particle based lane tracking.*



*Figure 6. Overview of our hybrid lane change prediction approach.*

- *The cost is a function of local features.* This extends the previous hypothesis by assuming that the cost function does not only depends on the agent's state but also in a number of external features (e.g. local people density, nearby traffic signs). Since the cost function depends on the features, it is possible to compute it even for previously unseen environments, as long as they contain the same kind of features.

Under these assumptions, the problem becomes that of learning the unknown cost function by observing how people move. This is often called *apprenticeship learning* and, when the underlying planning model is a Markov Decision Process, *inverse reinforcement learning*.

We have been working on a software library and evaluation testbed for different features and cost-function learning algorithms. We have conducted, in collaboration with the University of Freiburg, a first round of experiments concerning people moving in crowds. The results will be the subject of a submission to the IROS 2014 conference.

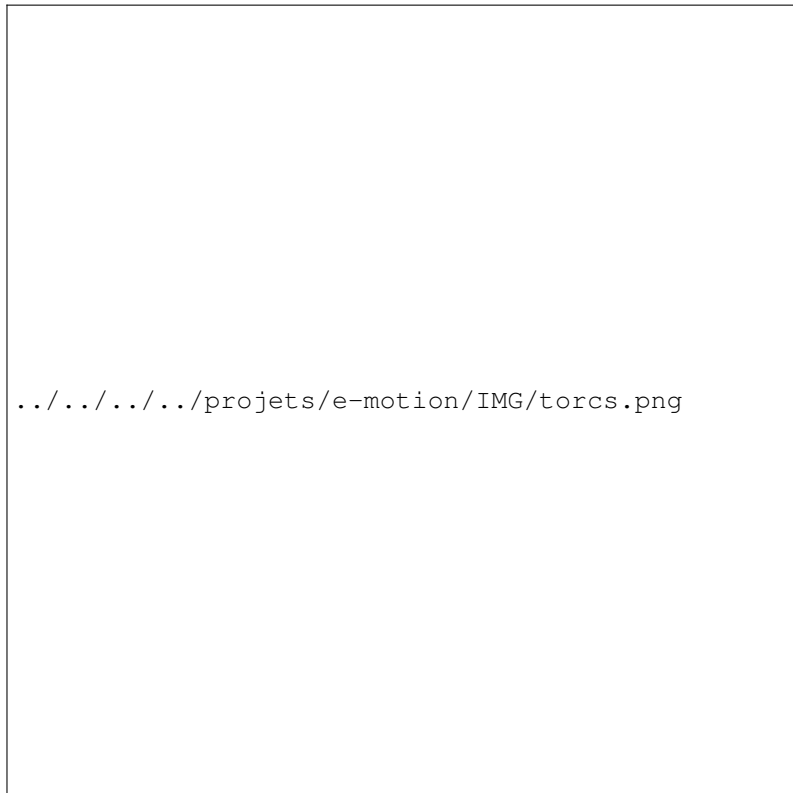


Figure 7. The TORCS racing simulator.

We have also been working on an application for intelligent vehicles and ADAS. As a first step, we have developed a ROS interface for the TORCS racing simulator, as well as a road simulation using the same platform (Fig. 7). This work has been the product of a collaboration with the Beijing University and IIT Hyderabad.

#### 5.2.4. Safety applications at road intersections for connected vehicles

**Participants:** Stéphanie Lefèvre, Christian Laugier.

From a safety perspective, road intersections are the most dangerous areas in the road network. They are also the most complex. Because of the extended situational awareness that they provide, wireless vehicular communications (or Vehicle-to-X communications, V2X) could greatly reduce the rate of intersection accidents. However, numerous research challenges remain before the full use of this technology can be achieved. A PhD was started on this topic in 2009 in collaboration with Renault, and was successfully defended in 2012 [42]. The purpose was to formulate and develop a probabilistic reasoning framework which would allow combining the information shared by the vehicles to estimate the situation and the associated risk as a vehicle negotiates an intersection. The first contribution of the PhD was to model the motion of vehicles using a Dynamic Bayesian Network where the maneuvers of different vehicles influence each other via an “expected maneuver”. This “expected maneuver” represents what a driver is expected to do given the state of the other vehicles in the area and the traffic rules which apply at the intersection. Thanks to the use of a probabilistic framework, uncertainties related to sensor errors and interpretation ambiguities are handled. The second contribution was a novel approach to risk estimation based on the comparison between what drivers intend to do and what they are expected to do. The reasoning is carried out by performing inference on the Dynamic Bayesian Network introduced earlier, using a particle filter. The approach was validated with field trials using Renault passenger vehicles equipped with vehicle-to-vehicle wireless communication modems [41], and in simulation [40]. The results show that the algorithm is able to detect dangerous situations early and complies with real-time constraints. We also developed a theoretical extension of the model to generalize it to arbitrary traffic situations [29]. This work is still ongoing thanks to an Inria@SiliconValley fellowship granted to S. Lefevre at the end of her PhD. Since January 2013 she is working in the Teleimmersion group at the University of California Berkeley, as a postdoctoral researcher. The research conducted there led to two new developments on the topic of “Safety applications at road intersections for connected vehicle”.

The first development concerns probabilistic decision making for Collision Avoidance (CA) systems. In the processing chain of a CA system, the “Decision making” module follows the “Risk assessment” module. The research done during the PhD stopped at the “Risk assessment” module, and we now address the challenges present in the “Decision making” module. We identified two main challenges:

1. The first one is that the decision making module has to make decisions based on uncertain knowledge. Sensors provide noisy measurements, digital maps contain errors, and interpreting a vehicle’s motion in terms of driver intention is uncertain. These uncertainties propagate to the risk assessment module and to the decision making module, but the latter is still required to make a decision from that uncertain knowledge.
2. The second one is that the timing of interventions is critical. If an intervention is triggered at a time when the uncertainty about the occurrence of a collision is too large, there is a chance that it will end up being a false alarm. High false alarm rates are detrimental to the driver acceptance of safety systems and can lead to the user losing trust in the system. On the other hand, if the system waits until the last moment (certainty about the occurrence of a collision) to trigger an intervention, it might be too late to avoid the accident.

In recent work [20] we proposed to introduce the possibility for a CA system to postpone making a decision. Our objective is to implement the fact that in some situations the new observations obtained by waiting will reduce the uncertainty about the occurrence of a collision, therefore the decision will be more reliable if it is made later using this additional information. The important question to solve is whether the potential gain brought by the additional information outweighs the cost of waiting. In order to answer this question, our decision making approach runs a *preposterior analysis* to determine the expected value and cost of the additional information. The value of the additional information can be quantified by means of the Expected Value of Sample Information (EVSI). It corresponds to the additional expected payoff possible through knowledge of the additional information and is computed by subtracting the expected costs of deciding with and without additional information. The cost of the additional information is quantified by means of the Expected Cost of Waiting (ECW). It is computed as the difference between the probability that the CA system will be able to avoid the potential collision if it intervenes now and if it intervenes at time  $t+1$ . Our decision making strategy is to postpone the decision making process to time  $t+1$  if and only if the EVSI is positive and the ECW is null. The algorithm was tested in simulation at a two-way stop intersection for collision scenarios

and no-collision scenarios involving two vehicles. A comparative evaluation with a decision making strategy which does not allow postponing decisions showed that our approach generates fewer false alarms and avoids as many collisions. These results were published at the conference IEEE IROS'13 [20] and a patent application was filed with UC Berkeley [32].

The second development addresses the relationship between privacy strategies for V2X and safety applications which rely on V2X. User privacy is a requirement for wireless vehicular communications, and a number of privacy protection strategies have already been developed and standardized. In particular, methods relying on the use of temporary pseudonyms and silent periods have proved their ability to confuse attackers who would attempt to track vehicles. However, these privacy protection schemes are not without consequences for safety applications. Such applications make decisions (e.g. warning drivers of an upcoming danger) based on their current estimation of the state of the real world, and this representation is created from the information contained in beacons received from other vehicles. Therefore, interruptions in the transmission of information will impact the decision-making process. If a silent period is scheduled to start at a safety-critical moment, it could result in safety systems not intervening when they should have, namely a “missed intervention”. From a user and safety perspective, this is not acceptable. In this work we address this issue and evaluate the impact of pseudonym change strategies on V2X-based Intersection Collision Avoidance (ICA) system. We use the ICA system developed during the PhD and simulated 3 different privacy protection strategies:

1. The “*Fixed ID*” strategy assigns a fixed pseudonym to a vehicle for the entire duration of a trip (i.e. a new pseudonym is assigned to the vehicle every time it starts). Testing this case gives us a reference for how well the collision avoidance system performs when there is no pseudonym change and no silent period during a trip, which is what was assumed in our PhD work.
2. The “*Baseline*” strategy follows the recommendations of the SAE J2735 standard for V2X communications. Pseudonyms are changed every 120 seconds and are followed by a silent period of random duration.
3. The “*Adaptive*” strategy is a modified version of the *Baseline* strategy where the risk of the situation is taken into account to decide whether or not a vehicle should be allowed to change pseudonym at time  $t$ . It relies on the estimation of the current ability of the collision avoidance system to keep the vehicle on a collision-free trajectory. The idea here is to authorize a pseudonym change and silent period only if it will not affect the performance of the safety application.

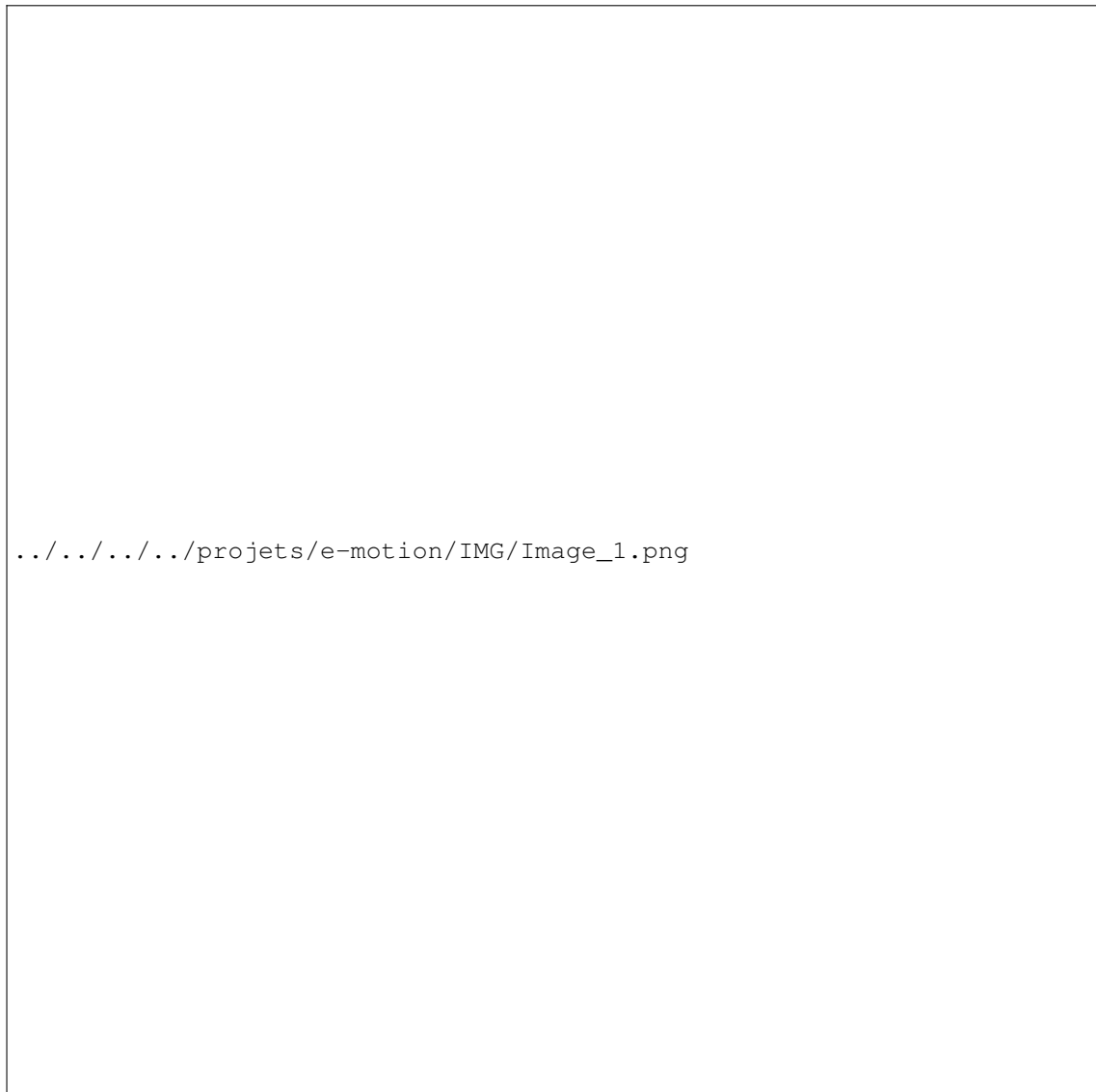
Simulations were conducted using the same simulator and the same scenarios as the tests run during the PhD. The performance of the three privacy strategies was evaluated both in terms of privacy and in terms of successful interventions of the ICA system. The results show that the ICA application requires silent periods to be shorter than two seconds in order to operate correctly in conjunction with the SAE J2735 standard. They also indicate that the addition of simple rules which authorize or not a pseudonym change depending on the context leads to major safety improvements compared to the SAE J2735 standard alone (see Figure 8). These results, which were published at the conference IEEE VNC'13 [21], highlight the necessity of a joint design. That is, the requirements of safety applications should be taken into account when designing privacy strategies, and pseudonym change schemes should be accounted for when designing safety applications which rely on V2X communications. This collaboration is necessary in order to ensure that vehicular communications and safety applications do not neutralize each other, but instead, work together toward safer roads.

### 5.2.5. Guidance for Uncertain shooting domain

**Participant:** Emmanuel Mazer.

This study is made in collaboration with MBDA (Monsieur Le Menec) and Probayes (Monsieur Laurent Saroul) under the ITP framework financed by the British MOD and the French DGA.

From N°51 of Vector Magazine (copyright MBDA) *It's an enduring question facing those in military conflict: 'when do you pull the trigger?' However, for pilots in air combat there is an added question: 'Once you've pulled the trigger, when do you break the link with your missile?' A new answer to that problem was a highlight of the MCM-ITP conference at Lille in May (see Vector 50). Entitled 'Guidance in Uncertain Shooting Domains', this joint project between MBDA and French company Probayes has produced a set of algorithms*



*Figure 8. Percentage of missed interventions, avoided collisions, and failed interventions as a function of the duration of the silent period for the Baseline strategy (left) and the Adaptive strategy (right).*



to help a pilot decide when to break a telemetry link. It's a decision fraught with danger. As pilots approach a target, the longer they keep the link, the greater the chance of their missile finding its target. However, closing in on the target usually means entering the enemy's own kill zone. So, a calculation of the best trade-off between mission success and pilot safety needs to be made.

"At the moment, when a pilot is engaging an enemy aircraft, he's obliged to do sums in his head," explained Graham Wallis, MBDA UK's Chief Technologist. "What we're trying to do is to take that away, and hide it behind a probability display." The problem arises because a seeker's scanning range is often far shorter than the maximum travelling distance, or kinematic range, of a missile. Though less of an issue at short range, it's a problem for medium-range weapons – where the target is likely to manoeuvre beyond the initial seeker scan area; hence the need for guidance from the pilot, who may have the target in radar sight.

**SUCCESS PROBABILITY** Stepping back, there are two sides to the firing equation. As customers require an air-to-air missile that will find its target almost 100 of the time, MBDA provides distance data (known as Launch Success Zone tables) to achieve that. They are understandably conservative. Clearly, though, the final decision to break the datalink can only be taken by the pilot, who is also making mental calculations to minimise his or her exposure to deadly risk. GUSD effectively offers a 'probability meter' to help reconcile those two imperatives.

Physically, GUSD could be a display with four bars and the circle of a pie-chart. Each bar represents the probability in percentage terms of an enemy pilot adopting one of four typical behaviours during air combat: flying head-on to attack; turning tail and heading for home, and either turning left or right – along with the probabilities that the enemy has just launched its own missile. The pie chart gives a single percentage – of the MBDA missile's chances of hitting its target. "The figures displayed are effectively our computer trying to read the mind of the enemy pilot," said Graham Wallis, whose team also drew on the experience of MBDA's former air force pilots.

**QUICK CALCULATIONS** Not surprisingly, the computations behind GUSD are hugely complex. The main input is the realtime radar tracking data of the enemy aircraft, although other elements such as seeker acquisition data, missile dynamics and the missile's inertial navigation errors are also included. A set of algorithms (and their associated mathematical methods, see box) then process this input – with the three key algorithms covering target behaviour and identification (Hidden Markov Models); the generation of bundles of trajectories (Markov Chains and Monte Carlo techniques) and trajectory collision checking (R-Tree).

Currently at Technology Readiness Level 3, GUSD's future depends on being incorporated into a programme. The generic data used to date would be replaced by real missile and seeker data, requiring GUSD to move out of the open forum of MCM-ITP and into the area covered by defence secrecy. "Implementation of research and technology is particularly difficult when it comes to jet fighters," Graham Wallis declared. "But I'm optimistic it will get into a future upgrade of current aircraft, and could even replace the Launch Success Zone tables as a firing cue for pulling that trigger."

**MATHEMATICAL MODELS** Named after 19th century Russian mathematician Andrey Markov, the Markov Chain uses a set of rules to predict what will happen next in a situation, when all the variables are known. In a game of bridge, if all the hands were known to all players and they had no discretion over which card to play next, every card laid down could be predicted as a Markov Chain. With a Hidden Markov Model, a player doesn't know what an opponent holds but can infer that, for example, they no longer have any trump cards because of how they have played their last hand. The Model takes multiple states e.g. 'Opponent Has Trumps' or 'Opponent Doesn't Have Trumps' and establishes a percentage weighting of probability to each – which can then be used for a decision over the best card to play. Though it is only an analogy, GUSD's algorithms use the mass of input data to infer the probability of an enemy pilot's tactical moves. R-trees are a way of handling multi-dimensional information - in GUSD's case by using geometric models that drastically cut the computation load.

**OTHER APPLICATIONS** Even before its own future is decided, GUSD could already lead to a successor MCM-ITP project with Probayes in the area of mission planning for long-range missiles. Called Rapid Mission

*Planning and Rehearsal, it could automate much of what is currently a labour-intensive process and to provide the customer with different options for mission routes – each with its own probability rating for success.*



*Figure 9. artistic view of GUSD (copyright MBDA)*

## 5.3. Human Centered Navigation in the physical world

### 5.3.1. 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 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. We accomplish this by appointing to humans distinct, skew-normal density functions that quantify social sensitivity and by using them in the sequel to induce a training set for regressing a collective density function of social sensitivity (see fig. 10 ). Finally, we extract level-sets of constant social sensitivity levels within the social map by which we can effectively and efficiently analyse individual as well as shared interaction zones of varying shape and size. Extensive experiments on human interaction scenarios demonstrate the feasibility and utility of the proposed approach in diverse conditions and promote its application to social mapping of human-populated environments. This work was published at IROS [23] and submitted to RAS journal.



*Figure 10. A representative example of mapping human social interactions using the proposed methodology. The regressed global sociality density along with isocontours of constant social comfort are superimposed on the corresponding scene.*

### 5.3.2. Goal oriented risk based navigation in dynamic uncertain environment

**Participants:** Anne Spalanzani, Jorge Rios-Martinez, Arturo Escobedo-Cabello, Procopio Silveira-Stein, Gregoire Vignon, Alejandro Dizan Vasquez Govea, Christian Laugier.

Navigation in large dynamic spaces has been addressed often using deterministic representations, fast updating and reactive avoidance strategies. However, probabilistic representations are much more informative and their use in mapping and prediction methods improves the quality of obtained results. 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 [36]. Moving obstacles are supposed to move along typical motion patterns represented by Gaussian Processes or Growing HMM. The likelihood of the obstacles' future trajectory and the probability of occupation are used to compute the risk of collision. The proposed planning algorithm, call RiskRRT (see Figure 11 for an illustration), is a sampling-based partial planner guided by the risk of collision. Results concerning this work were published in [37] [38] [39]. [47] and [48]. In 2013, Jorge Rios defended his PhD on this topic. We obtained an Inria ADT to optimize and share the RiskRRT algorithm.



Figure 11. Social navigation example. RiskRRT selected a plan (red line) to the goal (red arrow).

### 5.3.3. 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 select and follow 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. To accomplish this proposition, novel approaches are developed in the area of leader selection, where two methods are explored. The first uses motion prediction approaches while the second uses a machine learning method, to evaluate the leader quality of subjects, which is trained with real examples. Finally, the leader selection methods are integrated with motion planning algorithms and experiments are conducted in order to validate the proposed techniques. One of the most relevant application is navigation among crowds. Figure 12 illustrates the concept.



Figure 12. Navigation among crowds. The robot (rectangle) needs to reach the right side of the corridor. 2 groups of people (one in yellow, the other in blue) are crossing in this corridor. A classical motion planning would not find a path. The robot choses a leader (represented by a green circle around a person) and follows him.

The work is published in [25] [24] and Procopio Stein defended his PhD the 11th of december 2013 at the Aveiro University (pHD co-directed by Anne Spalanzani and Vitor Santos).

#### 5.3.4. 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 [46]. 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.

Simple improvements of the classical powered wheelchair can often diminish several difficulties while driving. This idea of comfort has emerged as a design goal in autonomous navigation systems, designers are becoming more aware of the importance of the user when scheming solution algorithms. This is particularly important when designing services or devices intended to assist people with some disability. In order for the robot to have a correct understanding of the intention of the user (when moving around) it is necessary to create a model of the user that takes into account his habits, type of disability and environmental information. The ongoing research project is centered in the understanding of the intentions of the user while driving an autonomous wheelchair, so that we can use this information to make this task easier.

During the last two years the work was centered in the improvement of the usability of the system. A review of the state of the art in user's intention estimation algorithms was made and a new model to infer the intentions of the user in a known environment was presented [34], [35].

The algorithm models the intention of the user as 2D topological goals in the environment. Those places are selected according to how frequently they are visited by the user (user habits). The system was designed so that the user can give orders to the wheelchair by using any type of interface, as long as he can show the direction of the intended movement (joystick, head tracking, brain control, etc). As shown in figure13 , the chosen

approach uses a Dynamic Bayesian Network to model and infer the intentions. The main contribution of this work is to model the intention of the user as topological goals instead of normal trajectory-based methods, therefore the model is simpler to deal with.

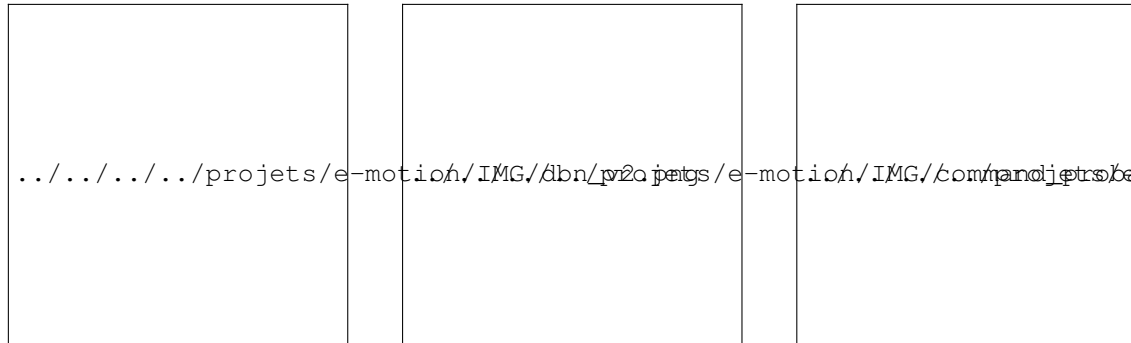


Figure 13. User's Intention Estimation Algorithm Left: User's intention model. The Bayesian network used to estimate the current user's intended; Center: The probability value for a given command  $C_t$  (big arrow) is proportional to the angle  $\alpha_i$  formed respect to each goal  $g_i$  in the environment. Right: The user is looking to the left (in the direction of his desired goal). Once that the user's intention estimation system computes the goal with the highest probability, the autonomous navigation module plans the path and controls the movement of the wheelchair to take the user to the destination.

In 2013 the results of the user intention destination method were published in the IROS conference [16]. The presented work covered the following aspects:

**User Intention Estimation:** Two different methods to drive the wheelchair were compared, a semi-autonomous and a manual mode. In semi-autonomous mode the user's intention is estimated from the position of the face and the wheelchair takes care of all the planning to arrive there while avoiding obstacles. In manual mode the wheelchair is driven using the face without assistance from the robotic controller.

The Bayesian estimator shown in Fig. 13 was used to do the inference of the desired destination of the user in semi-autonomous mode. The user's intention was modeled as a set of destinations commonly visited by the user and the task consisted in finding the destination targeted by the user.

**Interfaces:** People with motor disabilities and elders often have problems using joysticks and other standard control devices. Under this consideration our experimental platform was equipped with different types of user-interfaces to provide a multi-modal functionality as described in [35]. A face pose interface allows to control the wheelchair's motion by changing the face direction, while voice recognition interface is used to guarantee an adequate control of the wheelchair for those commands that otherwise would be difficult to give by only using the face (Stop, start, etc). This exploitation of more natural and easy-to-use human machine interfaces was one of the main contributions of the work presented in [16].

**Experimental evaluation:** Experiments were done in the hall of the Inria Rhône-Alpes laboratory. People in the scene were tracked to detect the most visited destinations in this setting (red circles). Those typical destinations were then placed in the map used by the robotic wheelchair. Each destination has a probability value related to the number of times that it is visited by people. The extracted typical destinations and related probabilities are used as prior knowledge when inferring the user's desired destination. In Fig. 14 (Right) the spheres represent the typical destinations placed in the internal map of the wheelchair and the size of the sphere represents the probability of being the desired destination of the user in the wheelchair given its position in the map and direction of the face (blue arrow).

To evaluate the performance of the method different persons were asked to drive the wheelchair in both “manual” and “semi-autonomous” mode. The trajectories of followed by the wheelchair were recorded and evaluated in [16], In Fig.15 one sample of those trajectories is presented where it can be appreciated how those trajectories executed with assistance of the robot are considerably softer than those obtained in manual mode as explained in [16].

**Human aware navigation:** Current work is being done in the construction of a social cost-map that is able to work with different open source path planning algorithms. This plug-in was developed using the method presented by Rios in [45].

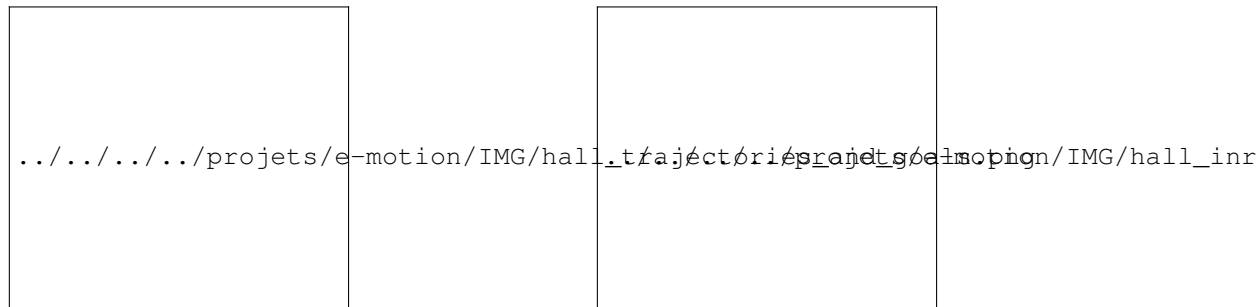


Figure 14. Left: The users of a normal environment move between typical destinations that can be learned. Right: The typical destinations marked in the map used by the wheelchair, The probability for each destination given the position of the wheelchair and direction of the command is proportional to the size of the sphere.

### 5.3.5. Bayesian modelling to implement and compare different theories of speech communication

**Participants:** Raphael Laurent, Pierre Bessière, Julien Diard, Jean-Luc Schwartz.

A central issue in speech science concerns the nature of representations and processes involved in communication. The search for phoneme or syllable specific invariants led to three major sets of approaches: motor, auditory and perceptuo-motor theories. They have been widely argued for and against, but the theoretical debate appears to be stagnating. It is our belief that computational models designed within a rigorous mathematical framework may allow to put forward new arguments to support either theory, and new ideas for experiments to be carried out on human subjects.

We have designed an integrative Bayesian model which allows to study auditory, motor and perceptuo-motor aspects of speech production and perception. This model was tested on perception tasks on evaluation corpora with more and more variability compared to the learning corpus. This showed a really high robustness of the purely motor model, which contained more information than it is the case in practise, due to unrealistic learning methods. The work was then focused on more realistic learning algorithms, where speech motor gestures are unsupervisedly learned through imitation, by generating motor gestures trying to reach auditory targets, and memorising the acoustics corresponding to these motor commands. This work was published in 2013 [19]. Raphael Laurent defended his PhD in 2013.

### 5.3.6. Bayesian computing

**Participants:** Emmanuel Mazer, Pierre Bessière.

A book and the compagnon software on bayesian programming have been released this year :

The book “Bayesian Programming” is available at <http://www.crcpress.com/product/isbn/9781439880326> and the associated software at <http://www.probayes.com/fr/Bayesian-Programming-Book/>



*Figure 15. Experimental evaluation of the user's intention method. Some samples of the resulting trajectories are presented. (a) and (c) show the results when using the assistance of the user's intention estimation system. (b) and (d) were achieved by driving the wheelchair using the face without any assistance. Here we can appreciate how those trajectories executed with assistance of the robot are considerably softer than those obtained in manual mode.*



## Features

Presents a new modeling methodology and inference algorithms for Bayesian programming Explains how to build efficient Bayesian models Addresses controversies, historical notes, epistemological debates, and tricky technical questions in a dedicated chapter separate from the main text Encourages further research on new programming languages and specialized hardware for computing large-scale Bayesian inference problems Offers an online Python package for running and modifying the Python program examples in the book

## Summary

Probability as an Alternative to Boolean Logic While logic is the mathematical foundation of rational reasoning and the fundamental principle of computing, it is restricted to problems where information is both complete and certain. However, many real-world problems, from financial investments to email filtering, are incomplete or uncertain in nature. Probability theory and Bayesian computing together provide an alternative framework to deal with incomplete and uncertain data.

## Decision-Making Tools and Methods for Incomplete and Uncertain Data

Emphasizing probability as an alternative to Boolean logic, Bayesian Programming covers new methods to build probabilistic programs for real-world applications. Written by the team who designed and implemented an efficient probabilistic inference engine to interpret Bayesian programs, the book offers many Python examples that are also available on a supplementary website together with an interpreter that allows readers to experiment with this new approach to programming.

## Principles and Modeling

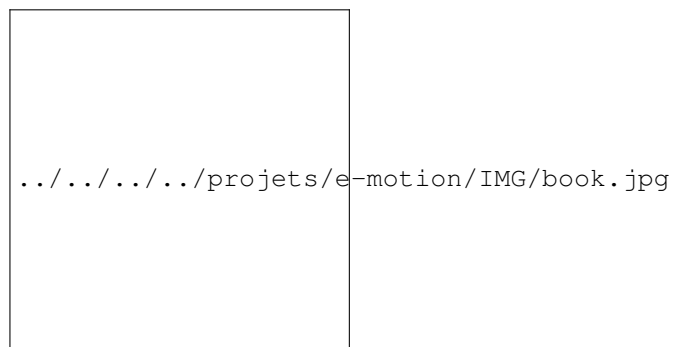
Only requiring a basic foundation in mathematics, the first two parts of the book present a new methodology for building subjective probabilistic models. The authors introduce the principles of Bayesian programming and discuss good practices for probabilistic modeling. Numerous simple examples highlight the application of Bayesian modeling in different fields.

## Formalism and Algorithms

The third part synthesizes existing work on Bayesian inference algorithms since an efficient Bayesian inference engine is needed to automate the probabilistic calculus in Bayesian programs. Many bibliographic references are included for readers who would like more details on the formalism of Bayesian programming, the main probabilistic models, general purpose algorithms for Bayesian inference, and learning problems.

FAQs Along with a glossary, the fourth part contains answers to frequently asked questions. The authors compare Bayesian programming and possibility theories, discuss the computational complexity of Bayesian inference, cover the irreducibility of incompleteness, and address the subjectivist versus objectivist epistemology of probability.

The First Steps toward a Bayesian Computer A new modeling methodology, new inference algorithms, new programming languages, and new hardware are all needed to create a complete Bayesian computing framework. Focusing on the methodology and algorithms, this book describes the first steps toward reaching that goal. It encourages readers to explore emerging areas, such as bio-inspired computing, and develop new programming languages and hardware architectures.



*Figure 16. Bayesian Programming book cover*

## FLOWERS Project-Team

# 6. New Results

## 6.1. Robotic and Computational Models of Human Development

### 6.1.1. Computational models of information-seeking, curiosity and attention

**Participants:** Pierre-Yves Oudeyer, Manuel Lopes.

An associated team, called Neurocuriosity, was created between Flowers and the Cognitive Neuroscience lab of Jacqueline Gottlieb at Univ. Columbia, NY. The goal of this associated team is to investigate mechanisms of spontaneous exploration and learning in humans by setting up experiments allowing to confirm or falsify predictions made by computational models previously developed by the team. This constitutes a crucial collaboration between developmental robotics and cognitive neuroscience. This joint work already led to a major publication on curiosity and information seeking, in the prestigious Trends in Cognitive Science journal (impact factor: 16.5). [27]

Abstract: Intelligent animals devote much time and energy to exploring and obtaining information, but the underlying mechanisms are poorly understood. We review recent developments on this topic that have emerged from the traditionally separate fields of machine learning, eye movements in natural behavior, and studies of curiosity in psychology and neuroscience. These studies show that exploration may be guided by a family of mechanisms that range from automatic biases toward novelty or surprise to systematic searches for learning progress and information gain in curiosity-driven behavior. In addition, eye movements reflect visual information searching in multiple conditions and are amenable for cellular-level investigations. This suggests that the oculomotor system is an excellent model system for understanding information-sampling mechanisms.

#### 6.1.1.1. Formalizing Imitation Learning

**Participants:** Thomas Cederborg, Pierre-Yves Oudeyer.

An original formalization of imitation learning was elaborated. Previous attempts to systematize imitation learning has been limited to categorizing different types of demonstrator goals (for example defining success in terms of the sequential joint positions of a dance, or in terms of environmental end states), and/or been limited to a smaller subset of imitation (such as learning from tele-operated demonstrations). The formalism proposed attempts to describe a large number of different types of learning algorithms using the same notation. Any type of algorithm that modifies a policy based on observations of a human, is treated as an interpretation hypothesis of this behavior. One example would be an update algorithm that updates a policy, partially based on the hypothesis that the demonstrator succeeds at demonstrations with probability 0.8, or an update algorithm that assumes that a scalar value is an accurate evaluation of an action compared to the latest seven actions. The formalism aims to give a principled way of updating these hypotheses, either rejecting some of a set of hypotheses regarding the same type of behavior, or set of parameters of an hypothesis. Any learning algorithm that modifies policy based on observations of a human that wants an agent to do something or act in some way, is describable as an interpretation hypothesis. If the learning algorithm is static, this simply corresponds to an hypothesis that is not updated based on observations. A journal article [26].

#### 6.1.1.2. Self-Organization of Early Vocal Development in Infants and Machines: The Role of Intrinsic Motivation

**Participants:** Clément Moulin-Frier, Sao Mai Nguyen, Pierre-Yves Oudeyer.

We bridge the gap between two issues in infant development: vocal development and intrinsic motivation. We propose and experimentally test the hypothesis that general mechanisms of intrinsically motivated spontaneous exploration, also called curiosity-driven learning, can self-organize developmental stages during early vocal learning and explain several aspects observed in infants (Figure 20). We introduce a computational model of intrinsically motivated vocal exploration, which allows the learner to autonomously structure its own vocal experiments, and thus its own learning schedule, through a drive to maximize competence progress. This model relies on a physical model of the vocal tract, the auditory system and the agent's motor control, as well as vocalizations of social peers. We present computational experiments that show how such a mechanism can explain the adaptive transition from vocal self-exploration with little influence from the speech environment, to a later stage where vocal exploration becomes influenced by vocalizations of peers (Figure 21). Within the initial self-exploration phase, we show that a sequence of vocal production stages self-organizes, and shares properties with data from infant developmental psychology: the vocal learner first discovers how to control phonation, then focuses on vocal variations of unarticulated sounds, and finally automatically discovers and focuses on babbling with articulated proto-syllables (Figure 22). As the vocal learner becomes more proficient at producing complex sounds, imitating vocalizations of peers starts to provide high learning progress explaining an automatic shift from self-exploration to vocal imitation.

This work has been recently accepted in the journal *Frontiers in Psychology, Cognitive Science* [30].

#### 6.1.1.3. *Emergent Proximo-Distal Motor Development through Adaptive Exploration, applied to Reaching and Vocal Learning*

**Participants:** Freek Stulp, Pierre-Yves Oudeyer, Jules Brochard, Clément Moulin-Frier.

Life-long robot learning in the high-dimensional real world requires guided and structured exploration mechanisms. In this developmental context, we have investigated the use of the  $PI^2$ -CMAES episodic reinforcement learning algorithm, which is able to learn high-dimensional motor tasks through adaptive control of exploration. By studying  $PI^2$ -CMAES in a reaching task on a simulated arm, we observe two developmental properties. First, we show how  $PI^2$ -CMAES autonomously and continuously tunes the global exploration/exploitation trade-off, allowing it to re-adapt to changing tasks. Second, we show how  $PI^2$ -CMAES spontaneously self-organizes a maturational structure whilst exploring the degrees-of-freedom (DOFs) of the motor space. In particular, it automatically demonstrates the so-called *proximo-distal maturation* observed in humans: after first freezing distal DOFs while exploring predominantly the most proximal DOF, it progressively frees exploration in DOFs along the proximo-distal body axis. These emergent properties suggest the use of  $PI^2$ -CMAES as a general tool for studying reinforcement learning of skills in life-long developmental learning contexts. This work was published in the *Paladyn Journal of Behavioral Robotics* [36].

This model of emergent developmental freezing and unfreezing of degrees of freedom was then applied to infant vocal development. For this aim, we used an articulatory synthesizer which is a computer model of the human vocal tract and the ear. While testing different possibilities, the algorithm eventually creates learning structures, which are more efficient than random motor babbling. Using the algorithm with a vocal synthesizer, we show that it can reproduce a babbling infant's characteristic: the predominance of the jaw over the other articulators, namely the canonical babbling.

This is the first study to our knowledge of emergent maturation in speech. Without presupposing any biological or social constraint, we give a new explanation of the jaw predominance in babbling, based on freezing and freeing the degrees of freedom in an adaptive maturation scheme to improve learning. This provides an original hypothesis regarding the emergence of canonical babbling in infant vocal development.

This last work was performed during the internship of Jules Brochard in 2013 and a journal article is currently being written.

#### 6.1.2. *COSMO (“Communicating about Objects using Sensory-Motor Operations”): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems*

**Participants:** Clément Moulin-Frier, Jean-Luc Schwartz, Julien Diard, Pierre Bessi re.



*Figure 20. Rapid view of the first year of infant vocal development.*



*Figure 21. Our model displays an adaptive transition from vocal self-exploration with little influence from the speech environment, to a later stage where vocal exploration becomes influenced by vocalizations of peers.*



Figure 22. Within the self-exploration phase, our model first discovers how to control phonation, then focuses on vocal variations of unarticulated sounds, and finally automatically discovers and focuses on babbling with articulated proto-syllables.

This work began with the PhD thesis of Clément Moulin-Frier at GIPSA-Lab, Grenoble, France, supervised by Jean-Luc Schwartz (GIPSA-Lab, CNRS), Julien Diard (LPNC, CNRS) and Pierre Bessière (College de France, CNRS). A few papers were finalized during his post-doc at FLOWERS in 2012. Firstly, an international journal paper based on the PhD thesis work of Raphael Laurent (GIPSA-Lab), extending Moulin-Frier’s model, was published [108], as well as a commentary in *Behavioral and Brain Sciences* [97]. Both these papers provide computational arguments based on a sensory-motor cognitive model to feed the age-old debate of motor vs. auditory theories of speech perception. Secondly, in another journal paper under the submission process, we attempt to derive some properties of phonological systems (the sound systems of human languages) from the mere properties of speech communication. We introduce a model of the cognitive architecture of a communicating agent, called COSMO (for “Communicating about Objects using Sensory-Motor Operations”) that allows expressing in a probabilistic way the main theoretical trends found in the speech production and perception literature. This allows a computational comparison of these theoretical trends, helping to identify the conditions that favor the emergence of linguistic codes. We present realistic simulations of phonological system emergence showing that COSMO is able to predict the main regularities in vowel, stop consonant and syllable systems in human languages.

This work is currently under consideration as a target article for a special issue in an international journal. Pierre-Yves Oudeyer joined this process as a member of the editing committee.

### 6.1.3. Recognizing speech in a novel accent: the Motor Theory of Speech Perception reframed

**Participants:** Clément Moulin-Frier, Michael Arbib.

Clément Moulin-Frier engaged this work with Michael Arbib during his 6-month visit in 2009 at the USC Brain Project, University of Southern California, Los Angeles, USA, during his PhD thesis at Gipsa-Lab, Grenoble. He continues to write a journal article during his post-doc in the Flowers team in 2012-2013. This paper has been published recently in *Biological Cybernetics* [29], in which we offer a novel computational

model of foreign-accented speech adaptation, together with a thorough analysis of its implications with respect to the motor theory of speech perception.

## 6.2. Life-Long Robot Learning and Development of Motor and Social Skills

### 6.2.1. Active Learning and Intrinsic Motivation

#### 6.2.1.1. Active Learning of Inverse Models with Goal Babbling

**Participants:** Adrien Baranes, Pierre-Yves Oudeyer.

We have continued to elaborate and study our Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity (SAGG-RIAC) architecture as an intrinsically motivated goal exploration mechanism which allows active learning of inverse models in high-dimensional redundant robots. Based on active goal babbling, this allows a robot to efficiently and actively learn distributions of parameterized motor skills/policies that solve a corresponding distribution of parameterized tasks/goals. The architecture makes the robot sample actively novel parameterized tasks in the task space, based on a measure of competence progress, each of which triggers low-level goal-directed learning of the motor policy parameters that allow to solve it. For both learning and generalization, the system leverages regression techniques which allow to infer the motor policy parameters corresponding to a given novel parameterized task, and based on the previously learnt correspondences between policy and task parameters.

We have conducted experiments with high-dimensional continuous sensorimotor spaces in three different robotic setups: 1) learning the inverse kinematics in a highly-redundant robotic arm, 2) learning omnidirectional locomotion with motor primitives in a quadruped robot [2324](#), 3) an arm learning to control a fishing rod with a flexible wire. We show that 1) exploration in the task space can be a lot faster than exploration in the actuator space for learning inverse models in redundant robots; 2) selecting goals maximizing competence progress creates developmental trajectories driving the robot to progressively focus on tasks of increasing complexity and is statistically significantly more efficient than selecting tasks randomly, as well as more efficient than different standard active motor babbling methods; 3) this architecture allows the robot to actively discover which parts of its task space it can learn to reach and which part it cannot.

This work was published in the journal *Robotics and Autonomous Systems* [\[25\]](#).

#### 6.2.1.2. Learning Exploration Strategies in Model-based Reinforcement Learning

**Participants:** Manuel Lopes, Todd Hester, Peter Stone, Pierre-Yves Oudeyer.

We studied how different exploration algorithms can be combine and selected at runtime. Typically the user must hand-tune exploration parameters for each different domain and/or algorithm that they are using. We introduced an algorithm called leo for learning to select among different exploration strategies on-line. This algorithm makes use of bandit-type algorithms to adaptively select exploration strategies based on the rewards received when following them. We show empirically that this method performs well across a set of five domains. In contrast, for a given algorithm, no set of parameters is best across all domains. Our results demonstrate that the leo algorithm successfully learns the best exploration strategies on-line, increasing the received reward over static parameterizations of exploration and reducing the need for hand-tuning exploration parameters [\[46\]](#).

#### 6.2.1.3. Active Inverse Reinforcement Learning through Generalized Binary Search

**Participants:** Manuel Lopes, Francisco Melo.

We contributed the first aggressive active learning algorithm for nonseparable multi-class classification. We generalize an existing active learning algorithm for binary classification [\[116\]](#) to the multi-class setting, and identify mild conditions under which the proposed method provably retains the main properties of the original algorithm, namely consistency and sample complexity. In particular, we show that, in the binary case, our method reduces to the original algorithm of [\[116\]](#). We then contribute an extension of our method to multi-label settings, identify its main properties and discuss richer querying strategies. We conclude the paper with two illustrative application examples. The first application features a standard text-classification problem. The second application scenario features a learning from demonstration setting. In both cases we demonstrate the advantage of our active sampling approach against random sampling. We also discuss the performance of the proposed approach in terms of the derived theoretical bounds.





*Figure 23. Experimenting SAGG-RIAC for learning an inverse model for omnidirectional locomotion of a quadruped robot. The quadruped robot is controlled using 24 dimensional motor synergies parameterized with 24 continuous values : 12 for the amplitudes and 12 others for the phases of a sinusoid tracked by each motor. Experiments consider a task space  $u, v, \alpha$  which corresponds to the 2D position and orientation of the quadruped.*



*Figure 24. Evolution of the quality of the learnt inverse model for the quadruped robot experiment, depending on various exploration strategies (measured as mean error over a set of uniformly distributed goals generated independantly from learning trials).*

#### 6.2.1.4. Exploration strategies in developmental robotics: a unified probabilistic framework

**Participants:** Clément Moulin-Frier, Pierre-Yves Oudeyer.

We present a probabilistic framework unifying two important families of exploration mechanisms recently shown to be efficient to learn complex non-linear redundant sensorimotor mappings. These two explorations mechanisms are: 1) goal babbling, 2) active learning driven by the maximization of empirically measured learning progress. We show how this generic framework allows to model several recent algorithmic architectures for autonomous exploration. Then, we propose a particular implementation using Gaussian Mixture Models, which at the same time provides an original empirical measure of the competence progress. Finally, we perform computer simulations on two simulated setups: the control of the end effector of a 7-DoF arm and the control of the formants produced by an articulatory synthesizer. We are able to reproduce previous results from [25] with the advantages of a clean and compact probabilistic framework to efficiently express, implement and compare various exploration strategies on developmental robotics setups.

This work was published in three international conferences [54], [56], [55].

#### 6.2.1.5. Autonomous Reuse of Motor Exploration Trajectories

**Participants:** Fabien Benureau, Pierre-Yves Oudeyer.

We developed an algorithm for transferring exploration strategies between tasks that share a common motor space in the context of lifelong autonomous learning in robotics. In such context sampling is costly, and exploration can take a long time before finding interesting, learnable data about a task. Our algorithm shows that we can significantly reduce sampling by reusing past data of other learned tasks, with no need of external knowledge or specific task structure. The algorithm does not transfer observations, or make assumptions about how the learning is conducted. Instead, only selected motor commands are transferred between tasks, chosen autonomously according to an empirical measure of learning progress. We show that on a wide variety of variations from a source task, such as changing the object the robot is interacting with or altering the morphology of the robot, this simple and flexible transfer method increases early performance significantly in the new task. We also investigate the limitation of this algorithm on specific situations.

This work has been published at ICDL, in Osaka [40].

### 6.2.2. Learning and optimization of motor policies

#### 6.2.2.1. Off-Policy Actor-Critic

**Participants:** Thomas Degris, Martha White, Richard Sutton.

Actor-critic architectures are an interesting candidate for learning with robots: they can represent complex stochastic policies suitable for robots, they can learn online and incrementally and their per-time-step complexity scales linearly with the number of learned weights. Moreover, interesting connections have been identified in the existing literature with neuroscience. Until recently, however, practical actor-critic methods have been restricted to the on-policy setting, in which the agent learns only about the policy it is executing.

In an off-policy setting, on the other hand, an agent learns about a policy or policies different from the one it is executing. Off-policy methods have a wider range of applications and learning possibilities. Unlike on-policy methods, off-policy methods are able to, for example, learn about an optimal policy while executing an exploratory policy, learn from demonstration, and learn multiple tasks in parallel from a single sensory-motor interaction with an environment. Because of this generality, off-policy methods are of great interest in many application domains.

We have presented the first actor-critic algorithm for off-policy reinforcement learning. Our algorithm is online and incremental, and its per-time-step complexity scales linearly with the number of learned weights. We have derived an incremental, linear time and space complexity algorithm that includes eligibility traces and empirically show better or comparable performance to existing algorithms on standard reinforcement-learning benchmark problems. This work was reproduced independently by Saminda Abeyruwan from the University of Miami.

#### 6.2.2.2. Auto-Actor Critic

**Participant:** Thomas Degris.

As mentioned above, actor–critic architectures are an interesting candidate for robots to learn new skills in unknown and changing environments. However, existing actor–critic architectures, as many machine learning algorithms, require manual tuning of different parameters to work in the real world. To be able to systematize and scale-up skill learning on a robot, learning algorithms need to be robust to their parameters. The Flowers team has been working on making existing actor–critic algorithms more robust to make them suitable to a robotic setting. Results on standard reinforcement learning benchmarks are encouraging. This work will be submitted to international conference related with reinforcement learning. Interestingly, the methods developed in this work also offer a new formalism to think about different existing themes of Flowers research such as curiosity and maturational constraints.

#### 6.2.2.3. Deterministic Policy Gradient Algorithms

Thomas Degris and colleagues from UCL and Deepmind have considered deterministic policy gradient algorithms for reinforcement learning with continuous actions. The deterministic policy gradient has a particularly appealing form: it is the expected gradient of the action-value function. This simple form means that the deterministic policy gradient can be estimated much more efficiently than the usual stochastic policy gradient. To ensure adequate exploration, we introduce an off-policy actor-critic algorithm that learns a deterministic target policy from an exploratory behaviour policy. We demonstrate that deterministic policy gradient algorithms can significantly outperform their stochastic counterparts in high-dimensional action spaces. [58]

#### 6.2.2.4. Relationship between Black-Box Optimization and Reinforcement Learning

**Participant:** Freek Stulp.

Policy improvement methods seek to optimize the parameters of a policy with respect to a utility function. There are two main approaches to performing this optimization: reinforcement learning (RL) and black-box optimization (BBO). In recent years, benchmark comparisons between RL and BBO have been made, and there has been several attempts to specify which approach works best for which types of problem classes.

We have made several contributions to this line of research by: 1) Defining four algorithmic properties that further clarify the relationship between RL and BBO. 2) Showing how the derivation of ever more powerful RL algorithms displays a trend towards BBO. 3) Continuing this trend by applying two modifications to the state-of-the-art  $PI^2$  algorithm, which yields an algorithm we denote  $PI^{BB}$ . We show that  $PI^{BB}$  is a BBO algorithm, and, more specifically, that it is a special case of the state-of-the-art CMAES algorithm. 4) Demonstrating that the simpler  $PI^{BB}$  achieves similar or better performance than  $PI^2$  on several evaluation tasks. 5) Analyzing why BBO outperforms RL on these tasks. These contributions have been published on HAL [129], and in Paladyn: Journal of Behavioral Robotics [36].

#### 6.2.2.5. Probabilistic optimal control: a quasimetric approach

**Participants:** Steve N’Guyen, Clément Moulin-Frier, Jacques Droulez.

During his previous post-doc at the Laboratoire de Physiologie de la Perception et de l’Action (College de France, Paris), Clément Moulin-Frier joined Jacques Droulez and Steve N’Guyen to work on an alternative and original approach of probabilistic optimal control called *quasimetric*. A journal paper was published in PLoS ONE in December 2013 [31], where the authors propose a new approach dealing with decision making under uncertainty.

### 6.2.3. Social learning and intrinsic motivation

#### 6.2.3.1. Socially Guided Intrinsic Motivation for Skill Learning

**Participants:** Sao Mai Nguyen, Pierre-Yves Oudeyer.

We have explored how social interaction can bootstrap the learning of a robot for motor learning. We first studied how simple demonstrations by teachers could have a bootstrapping effect on autonomous exploration with intrinsic motivation by building a learner who uses both imitation learning and SAGG-RIAC algorithm [25], and thus designed the SGIM-D (Socially Guided Intrinsic Motivation by Demonstration) algorithm [32], [24], [32] [114], [111]. We then investigated on the reasons of this bootstrapping effect [113], to show that demonstrations by teachers can both enhance more tasks to be explored, as well as favor more easily generalized actions to be used. This analysis is generalizable for all algorithms using social guidance and goal-oriented exploration. We then proposed to build a strategic learner who can learn multiple tasks and with multiple strategies. An overview and theoretical study of multi-task, multi-strategy Strategic Learning is presented in [99]. We also forsook to build a learning algorithm for more natural interaction with the human users. We first designed the SGIM-IM algorithm so that it can determine itself when it should ask for help from the teacher while trying to explore autonomously as long as possible so as to use as little of the teacher's time as possible [112]. After tackling with the problem of how and when to learn, we also investigated an active learner who can determine who to ask for help: in the case of two teachers available, SGIM-IM can determine which strategy to adopt between autonomous exploration and learning by demonstration, and which teacher enhances most learning progress for the learner [115], and ask him for help.



Figure 25. Illustration of SGIM-D and SGIM-IM algorithms

While the above results have been shown in simulation environments: of a simple deterministic air hockey game (fig. 25), and a stochastic fishing experiment with a real-time physical simulator (fig. 26), we are now building the experimental setup of the fishing experiment in order to carry out the experiments with naive users.

#### 6.2.3.2. Adaptive task execution for implicit human-robot coordination



*Figure 26. Illustration of SGIM-D and SGIM-IM algorithms*



*Figure 27. Illustration of SGIM-D and SGIM-IM algorithms*

**Participants:** Ievgen Perederieiev, Manuel Lopes, Freek Stulp.

We began a project which goal is to study how computational models of multi-agent systems can be applied in situations where one agent is a human. We aim at applications where robots collaborate with humans for achieving complex tasks..

A very important capability for efficient collaborative work is the mutual agreement of a task and the ability to predict the behavior of others. We address such aspect by studying methods that increase the predictability of the robot actions. An efficient motor execution becomes the one that not just optimize speed and minimizes energy but also the one that improves the reliability of the team behavior. We are studying policy gradient methods and working on policy improvement algorithms ( $PI^2$ ,  $CEM$  and  $CMAES$ ). A feasibility study will consider a simple task between a robot and a person where the goal is to coordinate the way a set of three colored buttons is pressed.

## 6.2.4. Unsupervised learning of motor primitives

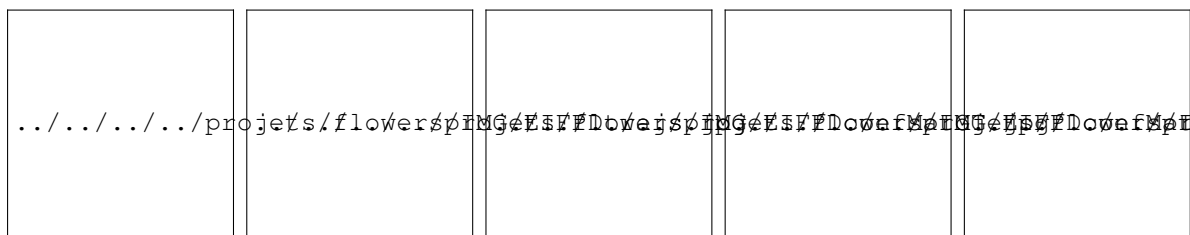
### 6.2.4.1. Clustering activities

**Participants:** Manuel Lopes, Luis Montesano, Javier Almingol.

Learning behaviors from data has applications in surveillance and monitoring systems, virtual agents and robotics among others. In our approach, we assume that in a given unlabeled dataset of multiple behaviors, it is possible to find a latent representation in a controller space that allows to generate the different behaviors. Therefore, a natural way to group these behaviors is to search a common control system that generate them accurately.

Clustering behaviors in a latent controller space has two major challenges. First, it is necessary to select the control space that generate behaviors. This space will be parameterized by a set of features that will change for different behaviors. Usually, each controller will minimize a cost function with respect to several task features. The latent representation is in turn defined by the selected features and their corresponding weight. Second, an unknown number of such controllers is required to generate different behaviors and the grouping must be based on the ability of the controller to generate the demonstrations using a compact set of controllers.

We propose a Dirichlet Process based algorithm to cluster behaviors in a latent controller space which encodes the dynamical system generating the observed trajectories. The controller uses a potential function generated as a linear combination of features. To enforce sparsity and automatically select features for each cluster independently, we impose a conditional Laplace prior over the controller parameters. Based on this models, we derive a sparse Dirichlet Process Mixture Model (DPMM) algorithm that estimates the number of behaviors and a sparse latent controller for each of them based on a large set of features [38].



(a)

(b)

(c)

(d)

(e)

Figure 28. EIFPD dataset. (a) Trajectories of the EIFPD to be clustered (color is non-informative). (b-d) correspondence matrix for the 474 trajectories for the labeled ground truth, the KMeans in measurement space and the DPMM, respectively. (e) Reconstructed trajectories from the initial point using the estimated parameters of the DPMM algorithm. Due to the large number of clusters (37), colors are repeated for different clusters.

#### 6.2.4.2. *Learning the Combinatorial Structure of Demonstrated Behaviors with Inverse Reinforcement Control*

**Participants:** Olivier Mangin, Pierre-Yves Oudeyer.

We have elaborated and illustrated a novel approach to learning motor skills from demonstration. This approach combines ideas from inverse reinforcement learning, in which actions are assumed to solve a task, and dictionary learning. In this work we introduced a new algorithm that is able to learn behaviors by assuming that the observed complex motions can be represented in a smaller dictionary of concurrent tasks. We developed an optimization formalism and show how we can learn simultaneously the dictionary and the mixture coefficients that represent each demonstration. We presented results on a toy problem and shown that our algorithm finds an efficient set of primitive tasks where naive approaches such as PCA and using a dictionary built from random examples fail to achieve the same tasks. These results that were presented as [60], extend the ones from [103].

#### 6.2.4.3. *Interaction of Maturation and Intrinsic Motivation for Developmental Learning of Motor Skills in Robots*

**Participants:** Adrien Baranes, Pierre-Yves Oudeyer.

We have introduced an algorithmic architecture that couples adaptively models of intrinsic motivation and physiological maturation for autonomous robot learning of new motor skills. Intrinsic motivation, also called curiosity-driven learning, is a mechanism for driving exploration in active learning. Maturation denotes here mechanisms that control the evolution of certain properties of the body during development, such as the number and the spatio-temporal resolution of available sensorimotor channels. We argue that it is useful to introduce and conceptualize complex bidirectional interactions among these two mechanisms, allowing to actively control the growth of complexity in motor development in order to guide efficiently exploration and learning. We introduced a model of maturational processes, taking some functional inspiration from the myelination process in humans, and show how it can be coupled in an original and adaptive manner with the intrinsic motivation architecture SAGG-RIAC (Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity algorithm), creating a new system, called McSAGG-RIAC. We then conducted experiments to evaluate both qualitative and quantitative properties of these systems when applied to learning to control a high-dimensional robotic arm, as well as to learning omnidirectional locomotion in a quadruped robot equipped with motor synergies. We showed that the combination of active and maturational learning can allow to gain orders of magnitude in learning speed as well as reach better generalization performances. A journal article is currently being written.

### 6.3. Autonomous and Social Perceptual Learning

#### 6.3.1. *The Impact of Human-Robot Interfaces on the Learning of Visual Objects*

**Participants:** Pierre Rouanet, Pierre-Yves Oudeyer, Fabien Danieau, David Filliat.

We have continued and finalized a large-scale study of the impact of interfaces allowing non-expert users to efficiently and intuitively teach a robot to recognize new visual objects. We identified challenges that need to be addressed for real-world deployment of robots capable of learning new visual objects in interaction with everyday users. We argue that in addition to robust machine learning and computer vision methods, well-designed interfaces are crucial for learning efficiency. In particular, we argue that interfaces can be key in helping non-expert users to collect good learning examples and thus improve the performance of the overall learning system. Then, we have designed four alternative human-robot interfaces: three are based on the use of a mediating artifact (smartphone, wiimote, wiimote and laser), and one is based on natural human gestures (with a Wizard-of-Oz recognition system). These interfaces mainly vary in the kind of feedback provided to the user, allowing him to understand more or less easily what the robot is perceiving, and thus guide his way of providing training examples differently. We then evaluated the impact of these interfaces, in terms of learning efficiency, usability and user's experience, through a real world and large scale user study. In this experiment, we asked participants to teach a robot twelve different new visual objects in the context of a robotic game. This game happens in a home-like environment and was designed to motivate and engage users in an interaction where using the system was meaningful. We then analyzed results that show significant differences among



interfaces. In particular, we showed that interfaces such as the smartphone interface allows non-expert users to intuitively provide much better training examples to the robot, almost as good as expert users who are trained for this task and aware of the different visual perception and machine learning issues. We also showed that artifact-mediated teaching is significantly more efficient for robot learning, and equally good in terms of usability and user's experience, than teaching thanks to a gesture-based human-like interaction.

This work was published in the IEEE Transactions on Robotics [34].



*Figure 29. Smartphone Interface. To make the robot collect a new learning example, users have to first draw the robot's attention toward the object they want to teach through simple gestures. Once the robot sees the object, they touch the head of the robot to trigger the capture. Then, they directly encircle the area of the image that represents the object on the screen. The selected area is then used as the new learning example. The combination of the video stream and the gestures facilitate the achievement of joint attention.*

### **6.3.2. Developmental object learning through manipulation and human demonstration**

**Participants:** Natalia Lyubova, David Filliat.



*Figure 30. Wiimote + laser pointer interface. With this interface users can draw the robot's attention with a laser pointer toward an object. The laser spot is automatically tracked by the robot. They can ensure that the robot detects the spot thanks to haptic feedback on the Wiimote. Then, they can touch the head of the robot to trigger the capture of a new learning example. Finally, they encircle the object with the laser pointer to delimit its area which will be defined as the new learning example.*



*Figure 31. The real world environment designed to reproduce a typical living room. Many objects were added in the scene in order to make the environment cluttered.*



*Figure 32. iCub performing curiosity-driven exploration and active recognition of visual objects in 3D*

The goal of this work is to design a visual system for a humanoid robot. We used a developmental approach that allows a humanoid robot to continuously and incrementally learn entities through interaction with a human partner in a first stage before categorizing these entities into objects, humans or robot parts and using this knowledge to improve objects models by manipulation in a second stage. This approach does not require prior knowledge about the appearance of the robot, the human or the objects. The proposed perceptual system segments the visual space into proto-objects, analyses their appearance, and associates them with physical entities. Entities are then classified based on the mutual information with proprioception and on motion statistics. The ability to discriminate between the robot's parts and a manipulated object then allows to update the object model with newly observed object views during manipulation. We evaluate our system on an iCub robot, showing the independence of the self-identification method on the robot's hands appearances by wearing different colored gloves. The interactive object learning using self-identification shows an improvement in the objects recognition accuracy with respect to learning through observation only [52], [51].

### 6.3.3. *A Comparison of Geometric and Energy-Based Point Cloud Semantic Segmentation*

#### *Methods*

**Participants:** Mathieu Dubois, Alexander Gepperth, David Filliat.

The software we developed for object segmentation and recognition rely on a geometric segmentation of the space. We tested alternative methods for this semantic segmentation task in which the goal is to find some relevant classes for navigation such as wall, ground, objects, etc. Several effective solutions have been proposed, mainly based on the recursive decomposition of the point cloud into planes. We compare such a solution to a non-associative MRF method inspired by some recent work in computer vision.

The results [42] shows that the geometric method gives superior results for the task of semantic segmentation in particular for the object class. This can be explained by the fact that it incorporates a lot of domain knowledge (namely that indoor environments are made of planes and that objects lie on top of them). However, MRF segmentation gives interesting results and has several advantages. First most of its components can be used for other purpose or in other, less constrained, environments where domain knowledge is not available. For instance we could try to recognize more precisely the objects. Second it requires less tuning since most parameters are learned from the database. Third, it uses the appearance information which could help to identify different types of ground or wall (this was one of the goal in the CAROTTE challenge). Last but not least, as it gives a probabilistic output, it allows the robot to draw hypothesis on the environment and adapt its behavior. Therefore we think it is interesting to investigate improvements to improve the exploitation of the structure of the point clouds.

### 6.3.4. *Efficient online bootstrapping of sensory representations*

**Participant:** Alexander Gepperth.

This work [86] is a simulation-based investigation exploring a novel approach to the open-ended formation of multimodal representations in autonomous agents. In particular, we addressed here the issue of transferring (bootstrapping) features selectivities between two modalities, from a previously learned or innate reference representation to a new induced representation. We demonstrated the potential of this algorithm by several experiments with synthetic inputs modeled after a robotics scenario where multimodal object representations are bootstrapped from a (reference) representation of object affordances, focusing particularly on typical challenges in autonomous agents: absence of human supervision, changing environment statistics and limited computing power. We proposed an autonomous and local neural learning algorithm termed PROPRED (projection-prediction) that updates induced representations based on predictability: competitive advantages are given to those feature-sensitive elements that are inferable from activities in the reference representation, the key ingredient being an efficient online measure of predictability controlling learning. We verified that the proposed method is computationally efficient and stable, and that the multimodal transfer of feature selectivity is successful and robust under resource constraints. Furthermore, we successfully demonstrated robustness to noisy reference representations, non-stationary input statistics and uninformative inputs.

### 6.3.5. *Simultaneous concept formation driven by predictability*

**Participants:** Alexander Gepperth, Louis-Charles Caron.

This work [83] was conducted in the context of developmental learning in embodied agents who have multiple data sources (sensors) at their disposal. We developed an online learning method that simultaneously discovers meaningful concepts in the associated processing streams, extending methods such as PCA, SOM or sparse coding to the multimodal case. In addition to the avoidance of redundancies in the concepts derived from single modalities, we claim that meaningful concepts are those who have statistical relations across modalities. This is a reasonable claim because measurements by different sensors often have common cause in the external world and therefore carry correlated information. To capture such cross-modal relations while avoiding redundancy of concepts, we propose a set of interacting self-organization processes which are modulated by local predictability. To validate the fundamental applicability of the method, we conducted a plausible simulation experiment with synthetic data and found that those concepts which are predictable from other modalities successively "grow", i.e., become overrepresented, whereas concepts that are not predictable become systematically under-represented. We additionally explored the applicability of the developed method to real-world robotics scenarios.

### 6.3.6. *The contribution of context: a case study of object recognition in an intelligent car*

**Participants:** Alexander Gepperth, Michael Garcia Ortiz.

In this work [84], we explored the potential contribution of multimodal context information to object detection in an "intelligent car". The used car platform incorporates subsystems for the detection of objects from local visual patterns, as well as for the estimation of global scene properties (sometimes denoted scene context or just context) such as the shape of the road area or the 3D position of the ground plane. Annotated data recorded on this platform is publicly available as the "HRI RoadTraffic" vehicle video dataset, which formed the basis for the investigation. In order to quantify the contribution of context information, we investigated whether it can be used to infer object identity with little or no reference to local patterns of visual appearance. Using a challenging vehicle detection task based on the "HRI RoadTraffic" dataset, we trained selected algorithms (context models) to estimate object identity from context information alone. In the course of our performance evaluations, we also analyzed the effect of typical real-world conditions (noise, high input dimensionality, environmental variation) on context model performance. As a principal result, we showed that the learning of context models is feasible with all tested algorithms, and that object identity can be estimated from context information with similar accuracy as by relying on local pattern recognition methods. We also found that the use of basis function representations [1] (also known as "population codes" allows the simplest (and therefore most efficient) learning methods to perform best in the benchmark, suggesting that the use of context is feasible even in systems operating under strong performance constraints.

### 6.3.7. *Co-training of context models for real-time object detection*

**Participant:** Alexander Gepperth.

In this work [85], we developed a simple way to reduce the amount of required training data in context-based models of real-time object detection and demonstrated the feasibility of our approach in a very challenging vehicle detection scenario comprising multiple weather, environment and light conditions such as rain, snow and darkness (night). The investigation is based on a real-time detection system effectively composed of two trainable components: an exhaustive multiscale object detector (signal-driven detection), as well as a module for generating object-specific visual attention (context models) controlling the signal-driven detection process. Both parts of the system require a significant amount of ground-truth data which need to be generated by human annotation in a time-consuming and costly process. Assuming sufficient training examples for signal-based detection, we showed that a co-training step can eliminate the need for separate ground-truth data to train context models. This is achieved by directly training context models with the results of signal-driven detection. We demonstrated that this process is feasible for different qualities of signal-driven detection, and maintains the performance gains from context models. As it is by now widely accepted that signal-driven object detection can be significantly improved by context models, our method allows to train strongly improved detection systems without additional labor, and above all, cost.

## 6.4. Robot Multimodal Learning of Language and Action

### 6.4.1. Learning semantic components from sub-symbolic multi-modal perception

**Participants:** Olivier Mangin, Caio Tomazelli Da Silva Oliveira, Pierre-Yves Oudeyer.

Perceptual systems often include sensors from several modalities. However, existing robots do not yet sufficiently discover patterns that are spread over the flow of multimodal data they receive. In this work we establish a framework to learn multimodal components from perception. We use a nonnegative matrix factorization algorithm to learn a dictionary of components that represent meaningful elements present in the multimodal perception, without providing the system with a symbolic representation of the semantics. In [53] we illustrate this framework by showing how a learner discovers word-like components from observation of gestures made by a human together with spoken descriptions of the gestures, and how it captures the semantic association between the two. These experiments were further extended during the internship of Caio Tomazelli Da Silva Oliveira. Importantly these experiments provide an example of language grounding into perception, and feature global understanding of a linguistic task without requiring its compositional understanding. The code of the experiments from [53] as well as the motion dataset have been made publicly available to improve the reproducibility of the experiments.

### 6.4.2. Curiosity-driven exploration and interactive learning of visual objects with the ICub robot

**Participants:** Mai Nguyen, Natalia Lyubova, Damien Gerardeaux-Viret, David Filliat, Pierre-Yves Oudeyer.

We studied how various mechanisms for cognition and learning, such as curiosity, action selection, imitation, visual learning and interaction monitoring, can be integrated in a single embodied cognitive architecture. We have conducted an experiment with the iCub robot for active recognition of objects in 3D through curiosity-driven exploration, in which the robot can manipulate the robot or ask a human user to manipulate objects to gain information and recognise better objects (fig. 27). For this experiment carried out within the MACSi project, we address the problem of learning to recognise objects in a developmental robotics scenario. In a life-long learning perspective, a humanoid robot should be capable of improving its knowledge of objects with active perception. Our approach stems from the cognitive development of infants, exploiting active curiosity-driven manipulation to improve perceptual learning of objects. These functionalities are implemented as perception, control and active exploration modules as part of the Cognitive Architecture of the MACSi project. We integrated a bottom-up vision system based on swift feature points and motor-primitive based robot control with the SGIM-ACTS algorithm (Socially Guided Intrinsic Motivation with Active Choice of Task and Strategy) as the active exploration module. SGIM-ACTS is a strategic learner who actively chooses which task to concentrate on, and which strategy is better according to this task. It thus monitors the learning progress for each strategy on all kinds of tasks, and actively interacts with the human teacher. We obtained an active object recognition approach, which exploits curiosity to guide exploration and manipulation, such that the robot can improve its knowledge of objects in an autonomous and efficient way. Experimental results show the effectiveness of our approach: the humanoid iCub is now capable of deciding autonomously which actions must be performed on objects in order to improve its knowledge, requiring a minimal assistance from its caregiver. This work constitutes the base for forthcoming research in autonomous learning of affordances. This work has been published in a conference [57] and in a journal paper [28].

### 6.4.3. Imitation Learning and Language

**Participants:** Thomas Cederborg, Pierre-Yves Oudeyer.

We have studied how context-dependant imitation learning of new skills and language learning could be seen as special cases of the same mechanism. We argue that imitation learning of context-dependent skills implies complex inferences to solve what we call the "motor Gavagai problem", which can be viewed as a generalization of the so-called "language Gavagai problem". In a full symbolic framework where percepts and actions are continuous, this allows us to articulate that language may be acquired out of generic sensorimotor imitation learning mechanisms primarily dedicated at solving this motor Gavagai problem. Through the use

of a computational model, we illustrate how non-linguistic and linguistic skills can be learnt concurrently, seamlessly, and without the need for symbols. We also show that there is no need to actually represent the distinction between linguistic and non-linguistic tasks, which rather appears to be in the eye of the observer of the system. This computational model leverages advanced statistical methods for imitation learning, where closed-loop motor policies are learnt from human demonstrations of behaviours that are dynamical responses to a multimodal context. A novelty here is that the multimodal context, which defines what motor policy to achieve, includes, in addition to physical objects, a human interactant which can produce acoustic waves (speech) or hand gestures (sign language). This was published in [26].

#### **6.4.4. Learning to Interpret the Meaning of Teaching Signals in Socially Guided Robot Learning**

**Participants:** Manuel Lopes, Jonathan Grizou, Thomas Cederborg, Pierre-Yves Oudeyer.

We elaborated an algorithm to bootstrap shared understanding in a human-robot interaction scenario where the user teaches a robot a new task using teaching instructions yet unknown to it. In such cases, the robot needs to estimate simultaneously what the task is and the associated meaning of instructions received from the user. For this work, we consider a scenario where a human teacher uses initially unknown spoken words, whose associated unknown meaning is either a feedback (good/bad) or a guidance (go left, right, ...). We present computational results, within an inverse reinforcement learning framework, showing that a) it is possible to learn the meaning of unknown and noisy teaching instructions, as well as a new task at the same time, b) it is possible to reuse the acquired knowledge about instructions for learning new tasks, and c) even if the robot initially knows some of the instructions' meanings, the use of extra unknown teaching instructions improves learning efficiency. Published articles: [43], [45].

An extension to the use of brain signals has been made [44]. Do we need to explicitly calibrate Brain Machine Interfaces (BMIs)? Can we start controlling a device without telling this device how to interpret brain signals? Can we learn how to communicate with a human user through practical interaction? It sounds like an ill posed problem, how can we control a device if such device does not know what our signals mean? This paper argues and present empirical results showing that, under specific but realistic conditions, this problem can be solved. We show that a signal decoder can be learnt automatically and online by the system under the assumption that both, human and machine, share the same a priori on the possible signals' meanings and the possible tasks the user may want the device to achieve. We present results from online experiments on a Brain Computer Interface (BCI) and a Human Robot Interaction (HRI) scenario.

#### **6.4.5. Active Learning for Teaching a Robot Grounded Relational Symbols**

**Participants:** Johannes Kulick, Tobias Lang, Marc Toussaint, Manuel Lopes.

The present work investigates an interactive teaching scenario, where a human aims to teach the robot symbols that abstract geometric (relational) features of objects. There are multiple motivations for this scenario: First, state-of-the-art methods for relational Reinforcement Learning demonstrated that we can successfully learn abstracting and well-generalizing probabilistic relational models and use them for goal-directed object manipulation. However, these methods rely on given grounded action and state symbols and raise the classical question Where do the symbols come from? Second, existing research on learning from human-robot interaction has focused mostly on the motion level (e.g., imitation learning). However, if the goal of teaching is to enable the robot to autonomously solve sequential manipulation tasks in a goal-directed manner, the human should have the possibility to teach the relevant abstractions to describe the task and let the robot eventually leverage powerful relational RL methods (see Figure 33 ). We formalize human-robot teaching of grounded symbols as an Active Learning problem, where the robot actively generates geometric situations that maximize his information gain about the symbol to be learnt. We demonstrate that the learned symbols can be used in a relational RL framework for the robot to learn probabilistic relational rules and use them to solve object manipulation tasks in a goal-directed manner. [47].



Figure 33. Active learning of symbol descriptions on a real world robot.

## 6.5. Robot Design and Morphological Computation

### 6.5.1. The Poppy Humanoid Robot: Leg Design for Biped Locomotion

**Participants:** Matthieu Lapeyre, Pierre Rouanet, Pierre-Yves Oudeyer.

In this paper introduced for Poppy as a novel humanoid robotic platform designed to jointly address three central goals of humanoid robotics: 1) study the role of morphology in biped locomotion; 2) study full-body compliant physical human-robot interaction; 3) be robust while easy and fast to duplicate to facilitate experimentation. The taken approach relies on functional modeling of certain aspects of human morphology, optimizing materials and geometry, as well as on the use of 3D printing techniques. In this article, we have focused on the presentation of the design of specific morphological parts related to biped locomotion: the hip, the thigh, the limb mesh and the knee. We also presented an initial experiments showing properties of the robot when walking with the physical guidance of a human. [50].

### 6.5.2. Poppy Humanoid Platform: Experimental Evaluation of the Role of a Bio-inspired Thigh Shape

**Participants:** Matthieu Lapeyre, Pierre Rouanet, Pierre-Yves Oudeyer.

In this paper, we present an experimental evaluation of the role of the morphology in the Poppy humanoid platform. More precisely, we have investigated the impact of the bio-inspired thigh, bended of  $6^\circ$ , on the balance and biped locomotion. We compare this design with a more traditional straight thigh. We describe both the theoretical model and real experiments showing that the bio-inspired thigh allows the reduction of falling speed by almost 60% (single support phase) and the decrease of the lateral motion needed for the mass transfer from one foot to the other by 30% (double support phase). We also present an experiment where the robot walks on a treadmill thanks to the social and physical guidance of expert users and we show that the bended thigh reduces the upper body motion by about 45% indicating a more stable walk.[48].

### 6.5.3. Morphological computation and body intelligence

#### 6.5.3.1. Comparative Study of the Role of Trunk in Human and Robot Balance Control

**Participants:** Matthieu Lapeyre [correspondant], Christophe Halgand, Jean-Renã© Cazalet, Etienne Guillaud, Pierre-Yves Oudeyer.





Figure 34. Five pictures have been taken while Poppy was walking and were stacked to obtain a qualitative view of the difference in the walking behavior in function of the morphology of the thigh.

Numerous studies in the field of functional motor rehabilitation were devoted to understanding the functioning of members but few are interested in the coordination of the trunk muscles and the relationship between axial and appendicular motricity which is essential in maintaining balance during travel. Acquiring new knowledge on this subject is a prerequisite in the development of new therapeutic strategies to restore motor function to the overall development of robotic orthosis that would assist the movement. Many robotic orthosis using EMG signals were unfortunately using few joints [82] and a system for controlling a multi articulated spine has not yet been developed. We propose here to use a multidisciplinary approach to define the neuro-mechanical principles where an axial system is operating in synergy with human and robot limbs.

To bring us a theoretical framework, we chose to study the reactions of the Acroban humanoid robot. Including 5 joints in the trunk, Acroban can reproduce in part the fluid movements of the human body [101] and especially to test its behavior when its trunk is held fixed or his arms are no longer used for rebalance. To disrupt postural balance in humans and robots, we have developed a low cost mobile platform (see Figure 35). This platform is made up of a broad stable support ( $0.8 \times 5m$ ) mounted on a skateboard having a power of 800W. The substitution of the initial order of skate by an embedded microcontroller allows us to generate mono-axial perturbations precise intensity and duration to ensure repeatability of the disturbance. We capture movements (Optitrack 250Hz) and record the acceleration of the platform (accelerometer embedded 2kHz), the center of pressure (WiiBalanceBoard 60Hz), and electromyography (EMG).

The experimental device (mobile platform and synchronized recordings) is operational. Preliminary experiments have allowed us to refine the profiles of disturbance on the robot Acroban. The analysis of preliminary results is in progress. Following this study, we hope to improve the modeling of the motor system in humans and robotic simulation as a basis for the development of robotic orthosis axial system. Second, the results provide a basis for improved balancing of Acroban primitives but also the development of future humanoid robots.

## 6.6. Educational Technologie

### 6.6.1. KidLearn: Adaptive Personalization of Educational Content with Machine Learning

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



Figure 35. Experimental setup for comparative study of the role of the trunk in human and robot balance control

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. One overview can be found in [27]. 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 2013, we have run a first 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. A report is available at: <http://arxiv.org/pdf/1310.3174v1.pdf>.

## 6.7. Other applications

### 6.7.1. Real-time Reaction-Diffusion Simulation: a Machine Learning Technique

**Participants:** Thomas Degris, Nejib Zemzemi.

Carmen is an Inria team working on modeling the electrical activity of the human heart. Their models are mainly based on reaction-diffusion equations. These methods are expensive in terms of computational costs which limits their use in practice. More specifically, some recent surgical intervention techniques on the heart (atrial ablation) requires to identify the source of the electrical wave. Finding such sources requires an optimization procedure. Using classical methods, this procedure is very heavy computationally.

In this project, our goal is to reduce the computational cost using supervised learning techniques. The idea is to replace the incremental resolution of partial differential equations by more suitable data structures for real-time running. Starting from data generated by simulating different excitations scenarios on a human atria, this data is afterwards used as a training data set for machine learning algorithms. This approach will allow a faster optimization procedure.

This work is in collaboration with Nejib Zemzemi from the Inria Carmen team. This project is in preliminary steps.

### **6.7.2. Appearance-based segmentation of indoors/outdoors sequences of spherical views**

**Participant:** David Filliat.

In collaboration with Patrick Rives and Alexandre Chapoulie from the Arobas team at Inria Sophia-Antipolis, we developed a method for environment segmentation based on spherical views [41]. Navigating in large scale, complex and dynamic environments requires reliable representations able to capture metric, topological and semantic aspects of the scene for supporting path planning and real time motion control. In a previous work, we addressed metric and topological representations thanks to a multi-cameras system which allows building of dense visual maps of large scale 3D environments. The map is a set of locally accurate spherical panoramas related by 6dof poses graph. The work presented here is a further step toward a semantic representation. We aim at detecting the changes in the structural properties of the scene during navigation. Structural properties are estimated online using a global descriptor relying on spherical harmonics which are particularly well-fitted to capture properties in spherical views. A change-point detection algorithm based on a statistical Neyman-Pearson test allows us to find optimal transitions between topological places. Results are presented and discussed both for indoors and outdoors experiments.

### **6.7.3. Modelling Stop Intersection Approaches using Gaussian Processes**

**Participant:** David Filliat.

In collaboration with Javier-Ibanez Guzman and Alexandre Armand from Renault, we developed an approach toward the development of an electronic co-pilot adapted to the driver behavior [39]. Indeed, each driver reacts differently to the same traffic conditions, however, most Advanced Driving Assistant Systems (ADAS) assume that all drivers are the same. This work proposes a method to learn and to model the velocity profile that the driver follows as the vehicle decelerates towards a stop intersection. Gaussian Processes (GP), a machine learning method for non-linear regressions are used to model the velocity profiles. It is shown that GP are well adapted for such an application, using data recorded in real traffic conditions. GP allow the generation of a normally distributed speed, given a position on the road. By comparison with generic velocity profiles, benefits of using individual driver patterns for ADAS issues are presented.

## LAGADIC Project-Team

# 6. New Results

## 6.1. Visual tracking

### 6.1.1. 3D model-based tracking

**Participants:** Antoine Petit, 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 [53], [52], [68], [11].

### 6.1.2. Pose estimation through multi-planes tracking

**Participants:** Bertrand Delabarre, Eric Marchand.

This study dealt with dense visual tracking robust towards scene perturbations using 3D information to provide a space-time coherency. The proposed method is based on a piecewise-planar scenes visual tracking algorithm which aims at minimizing an error between an observed image and reference templates by estimating the parameters of a rigid 3D transformation taking into account the relative positions of the planes in the scene. Both the sum of conditional variance and mutual information have been considered [40] [67].

### 6.1.3. Pose estimation from spherical moments

**Participant:** François Chaumette.

This study has been realized in collaboration with Omar Tahri from ISR in Coimbra (Portugal) and Youcef Mezouar from Institut Pascal in Clermont-Ferrand. It was devoted to the classical PnP (Perspective-from-N-Points) problem whose goal is to estimate the pose between a camera and a set of known points from the image measurement of these points. We have developed a new method based on invariant properties of the spherical projection model, allowing us to decouple the pose estimation in two steps: the first one provides the translation by minimizing a criterium using an iterative Newton-like method, the second one directly provides the rotation by solving a Procrustes problem [65], [26].

### 6.1.4. 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 [64], we studied the problem of designing 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 developed in [64] has a general validity and can be applied to many different contexts: in [64] it is shown how to tailor the proposed machinery to two concrete SfM problems involving structure estimation for point features and for planar regions from measured image moments.

### 6.1.5. 3D reconstruction of transparent objects

**Participant:** Patrick Rives.

This work has been realized in collaboration with Nicolas Alt, Ph.D. student at the “Technische Universität München” (TUM).

Visual geometry reconstruction of unstructured domestic or industrial scenes is an important problem for applications in virtual reality, 3D video or robotics. With the advent of Kinect sensor, accurate and fast methods for 3D reconstruction have been proposed. However, transparent objects cannot be reconstructed with methods that assume a consistent appearance of the observed 3D structure for different viewpoints. We proposed an algorithm that searches the depth map acquired by a depth camera for inconsistency effects caused by transparent objects. Consistent scene parts are filtered out. The result of our method hence complements existing approaches for 3D reconstruction of Lambertian objects [30].

### 6.1.6. *Pseudo-semantic segmentation*

**Participants:** Rafik Sekkal, Marie Babel.

This study has been realized in collaboration with Ferran Marques from Image Processing Group of the Technical University of Catalonia (Barcelona). We designed a video segmentation framework based on contour projections. This 2D+t technique provides a joint hierarchical and multiresolution solution. Results obtained on state-of-the-art benchmarks have demonstrated the ability of our framework to insure the spatio-temporal consistency of the regions along the sequence.

### 6.1.7. *Augmented reality*

**Participants:** Pierre Martin, Eric Marchand.

Using Simultaneous Localization And Mapping (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 [46].

Dealing with AR, we have proposed a method named Depth-Assisted Rectification of Patches (DARP), which exploits depth information available in RGB-D consumer devices to improve keypoint matching of perspective distorted images [44].

## 6.2. Visual servoing

### 6.2.1. *Photometric moment-based visual servoing*

**Participants:** Manikandan Bakthavatchalam, François Chaumette.

This goal of this work is to use a set of photometric moments as visual features for visual servoing. We first determined the analytical form of the interaction matrix related to these moments. From the results obtained in the past from binary moments, we then selected a set of four features to control four degrees of freedom (dof) with excellent decoupling and stability properties [35]. More recently, thanks to a collaboration with Omar Tahri from ISR Coimbra in Portugal, these results have been extended to the full six dof case.

### 6.2.2. *Visual servoing of humanoid robot*

**Participant:** François Chaumette.

This study has been realized in collaboration with the Pal robotics company located in Barcelona, Spain. It was devoted to the control of the arm of a humanoid robot by visual servoing for manipulation tasks [29].

### 6.2.3. *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. This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6).

### 6.2.4. 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.

## 6.3. Visual navigation of mobile robots

### 6.3.1. 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 Figure 5 .a). This device allows acquiring such omnidirectional images at high frame rate (30 Hz). This approach has advantages over other alternatives used nowadays in terms of accuracy and real-time spherical image construction of indoor environments, which are of particular interest for mobile robotics. This device has important prospective applications, such as fast 3D-reconstruction or simultaneous localization and mapping (SLAM). A novel calibration method for such device has been developed. It does not require any specific calibration pattern, taking into account the planar structure of the scene to cope with the fact that there is no overlapping between sensors. A method to perform image registration and visual odometry has also 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 based on ICP.

### 6.3.2. Long term mapping

**Participants:** Tawsif Gokhool, Patrick Rives.

This work inscribes in the context of lifelong navigation and map building. The kind of representation that we focus on is made up of a topometric map consisting of a graph of spherical RGB-D views. Thanks to the use of a saliency map built from the photometric and geometric data, we are able to characterize the conditioning of the pose estimation algorithm and to keep as keyframes only a subset of the spherical RGB-D views acquired on the fly. Subsequently, a study on the spread of keyframes was made. The aim was to investigate ways of covering completely and optimally the explored environment in a pose graph representation. Again, over here, the benefits are twofold. Firstly, data acquisition at a throttle of 30 Hz induces many redundant information in the database, which may not necessarily contribute much to the registration phase. Therefore, intelligent selection of keyframes helped in the reduction of data redundancy. Furthermore, as pointed out in the literature, frame to keyframe alignment has the advantage of reducing trajectory drift since the propagation error is diminished as well (see Figure 5 .b)

### 6.3.3. Semantic mapping

**Participants:** Romain Drouilly, Patrick Rives.

Semantic mapping aims at building rich cognitive representations of the world in addition to classical topometric maps. A dense labeling has been achieved from high resolution outdoor images using an approach combining Random Forest (RF) and Conditional Random Field (CRF). A second development dealt with the use of semantic information for localization in indoor scenes. For this kind of scenes dense labeling is more difficult due to the large number of potential classes. Therefore algorithms developed for this task rely on a sparse representation of indoor environments called “pbmap”. It consists of a graph whose nodes are the planes present in a given scene. These planes are the only parts of the scene that are labeled. Very high labeling rates of planes has been reached (more than 90%) and it has been shown that these labeled planes could be useful for localization and navigation tasks.

### 6.3.4. Automous navigation of wheelchairs

**Participants:** Rafik Sekkal, François Pasteau, Marie Babel.

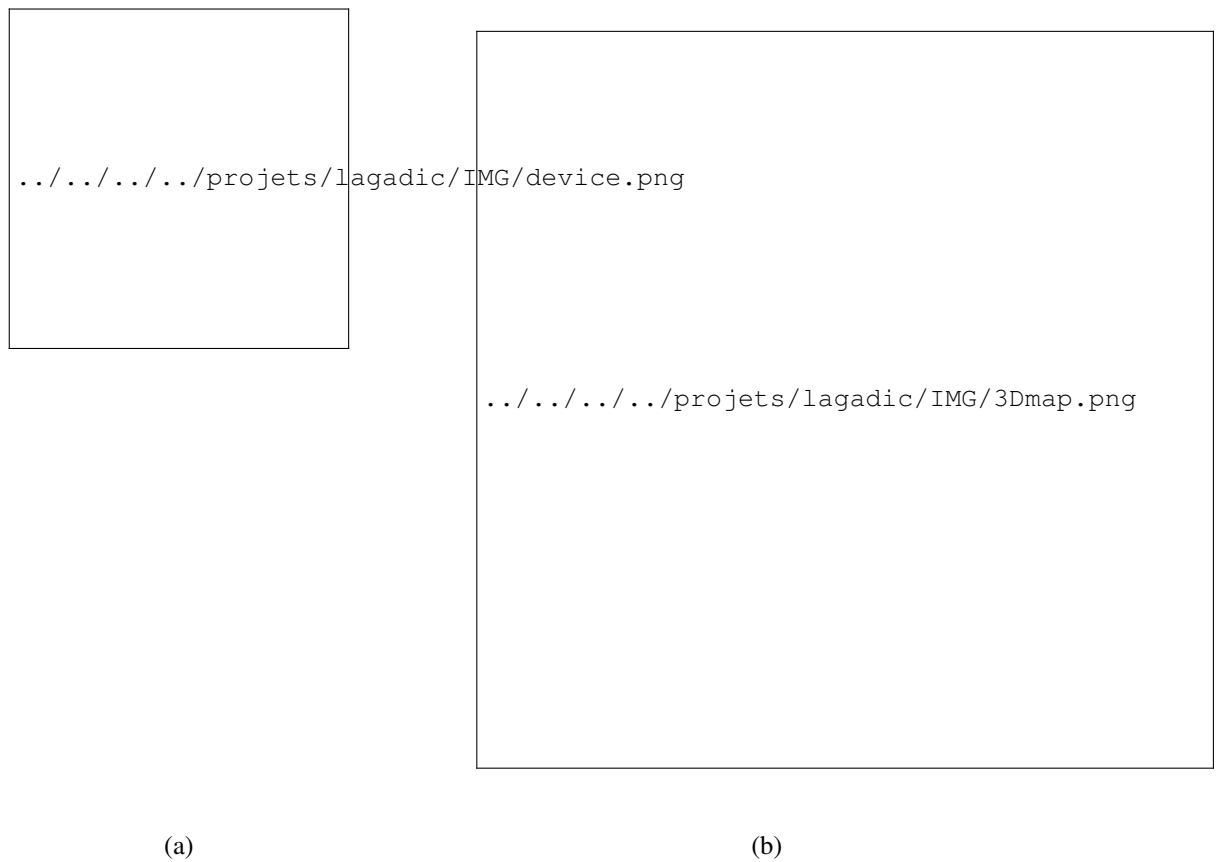


Figure 5. a) Omnidirectional RGB-D sensor, b) Top view of dense visual SLAM with fusion of intensity and depth

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 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 [60]. This overcomes the initialization issue typically raised in the literature. 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 [50]. 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 8.1.2).

### 6.3.5. *Semi-autonomous control of a wheelchair for navigation assistance along corridors*

**Participants:** Marie Babel, François Pasteau, Alexandre Krupa.

This study concerns a semi-autonomous control approach that we designed for safe wheelchair navigation along corridors. The control relies on the combination of a primary task of wall avoidance 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. Experimental results clearly show the ability of the approach to provide an efficient solution for wall avoiding purposes. 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 8.1.2).

### 6.3.6. *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. This study has been validated on our Pioneer robot (see Section 5.5).

### 6.3.7. *Obstacle avoidance*

**Participants:** 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 [9]. 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 [14]. Recent progresses have been obtained by considering moving obstacles [39].

## 6.4. Medical robotics

### 6.4.1. *Needle detection and tracking in 3D ultrasound*

**Participants:** Pierre Chatelain, Alexandre Krupa.

We developed an algorithm for detecting and tracking a flexible needle in a sequence of 3D ultrasound volumes when it is manually inserted, without any a priori information on the insertion direction. Our approach is based on the combination of a RANSAC algorithm with Kalman filtering in a closed loop fashion and allows real-time tracking of the needle. In addition, a pose-based visual servoing was developed for automatically moving a robotized 3D ultrasound probe in order to keep the needle tip centered in the volume and to align its main axis with the central plane of the volume. This needle detection algorithm and probe automatic guidance were experimentally validated during the insertion of a needle in a gelatin phantom [38].

### 6.4.2. *Non-rigid target tracking in ultrasound images*

**Participants:** Marie Babel, Alexandre Krupa.



In order to robustly track the motion of a tumour or cyst during needle insertion, we developed a new 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 new approach was validated from 2D ultrasound images of real human tissues undergoing rigid and non-rigid deformations.

#### 6.4.3. Adaptive arc-based path planning for robot-assisted needle 3D steering using duty-cycling control technique

**Participant:** Alexandre Krupa.

This study concerned the development of a method for three dimensional steering of a beveled-tip flexible needle that can be used in medical robotics for percutaneous assistance procedures. The proposed solution is the extension of an adaptive arc-based 2D planar approach. It combines the Rapidly-Exploring Random Tree (RRT) algorithm, the duty-cycling needle control technique and stop and turn phases to reorientate the needle in a new working plane each time it is necessary. Simulation results demonstrate the feasibility of this approach to reach a 3D target while avoiding obstacles and its robustness to needle kinematic model errors.

#### 6.4.4. Gait analysis

**Participants:** Cyril Joly, Patrick Rives.

Clinical evaluation of frailty in the elderly is the first step to decide the degree of assistance they require. Advances in robotics make it possible to turn a standard assistance device into an augmented device that may enrich the existing tests with new sets of daily measured criteria. We designed an augmented 4-wheeled rollator, equipped with a Kinect and odometers, for daily biomechanical gait analysis. It allows to estimate on line legs and feet configurations during the walk. Preliminary results [43] obtained on four healthy persons show that relevant data can be extracted for gait analysis (e.g. foot orientation and tibia-foot angle, feet position) during an assisted walk.

This work has been realized in collaboration with Claire Dune from the University of Toulon and in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6).

### 6.5. Control of single and multiple UAVs

#### 6.5.1. State estimation and flight control of quadrotor UAVs

**Participants:** Riccardo Spica, 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.

Key components for the successful deployment of such systems are (i) a reliable state estimation module able to deal with highly unstructured and/or GPS-denied indoor environments, and (ii) a robust flight control algorithm able to cope with model uncertainties and external disturbances (e.g., adverse atmospheric conditions). The difficulty of these estimation and control problems is also increased by the limited amount of sensing and processing capabilities onboard standard quadrotors: this clearly imposes additional strict requirements on the complexity of the employed algorithms.

In the context of robust flight control of standard quadrotors, the works [31], [32] addressed the theoretical developments and experimental validation of a novel nonlinear adaptive flight controller able to estimate online the UAV dynamic parameters (such as the position of the center of mass when carrying unmodeled payloads), and to compensate for external wind gusts. In parallel, we also developed in [63] a high performance and open-source hardware/software control architecture for flight control of quadrotor UAVs made available to the general public on a open repository. This was achieved by combining state-of-the-art filtering and control techniques with a careful customization and calibration of a commercially available and low-cost quadrotor platform. Finally, still in the context of flight control, the work [58] reported a successful experimental validation of several flight tests for a novel overactuated quadrotor design with tilting propellers behaving as a fully-actuated rigid body in 3D space (thus, able to control its position and orientation in a fully decoupled way).

As for state estimation, the work [41] introduces a novel nonlinear estimation filter meant to obtain a metric measurement of the body-frame linear velocity from optical flow decomposition (thus, visual input) and concurrent fusion of the accelerometer/gyro readings from the onboard IMU. The peculiarity of this filtering technique is the possibility to both explicitly characterize and impose the transient response of the estimation error (thus, the filter performance) by acting on the estimation gains and UAV motion (acceleration). This is in contrast with the consolidated use of EKF schemes which, because of their inherent linearization of the system dynamics, do not typically allow to draw any conclusions about the stability/transient response of the estimation error.

These works were realized in collaboration with the robotics groups at the University of Cassino, Italy, and at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

### 6.5.2. 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. In this respect, in [23] a fully decentralized strategy for continuous connectivity maintenance for a group of UAVs has been theoretically developed and experimentally validated on a team of 4 quadrotor UAVs. An extension for allowing an external planner (e.g., a human user) to vary online the minimum degree of connectivity of the group was also proposed in [59]. Finally, [48] dealt with the issue of coupling the purely reactive strategy for connectivity maintenance with an autonomous exploration algorithm in a cluttered 3D environment (still experimentally tested on a team of quadrotor UAVs). The complete software architecture developed for performing these and similar multi-UAV experiments was also published in [42].

These works were realized in collaboration with the robotics group at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

## AYIN Team

# 6. New Results

## 6.1. Markov Random Fields

### 6.1.1. Hierarchical multitemporal and multiresolution classification in remote sensing imagery

**Participants:** Ihsen Hedhli, Josiane Zerubia [contact].

*This activity was conducted in collaboration with Dr. Gabriele Moser and Prof. Sebastiano B. Serpico (Department of Electrical, Electronic, and Telecommunications Engineering and Naval Architecture, DITEN, University of Genoa) [<http://www.unige.it>] with partial financial support from CNES [<http://www.cnes.fr>].*

Markov random field (MRF), hierarchical classification, satellite image time series

The capability to monitor the Earth's surface, and especially urban and built-up areas, for environmental disasters such as floods or earthquakes, and to assess the ground impact and damage caused by such events, play important roles from multiple social, economic, and human viewpoints. Current and forthcoming satellite missions for Earth observation (EO; e.g., Pleiades, COSMO-SkyMed, TerraSAR-X, Sentinel) possess huge potential for such applications, as they allow a spatially distributed and temporally repetitive view of the monitored area at the desired spatial scales. In this framework, accurate and time-efficient classification methods using time series are especially important tools for supporting rapid and reliable assessment of the ground changes and damage induced by a disaster, in particular when an extensive area has been affected. Given the huge amount and variety of data available, the main difficulty is to find a classifier that takes into account multi-band, multi-resolution, multi-date, and possibly multi-sensor data.

This research addresses the problem of supervised classification at multiple spatial resolutions for multiple dates. The approach is based on the extension of recent methods proposed by DITEN and/or AYIN [4], [5], [6]. These methods focus on a supervised Bayesian classifier that combines joint class-conditional statistical modeling and a hierarchical Markov random field. The key idea of the proposed method is to combine the multiresolution modeling capabilities of this previous technique with a model for the temporal correlation among distinct images in a time series. For this purpose, a hierarchical spatio-temporal Markov random field model has been proposed that is aimed at fusing the pixel-wise, neighborhood, multiresolution, and temporal information associated with the input time series. Pixel-wise information is characterized through separate statistical modeling for each target class (e.g., vegetation, urban, etc.) by using a finite mixture model, estimated using a modified stochastic expectation maximization algorithm. Such a model is well suited to dealing with heterogeneous classes, and each mixture component may reflect the contribution of the different materials contained in a given class. At each considered resolution, the different input bands are statistically combined by using multivariate copulas, and the resulting statistical pixel-wise model is integrated in a hierarchical Markov random field based on a quad-tree structure. Among the different algorithms employed in the literature, we chose to use an exact estimator based on the marginal posterior mode (MPM). Specifically, a new formulation of MPM is developed to formalize, within the aforementioned hierarchical model, a 'cascade' multi-date decision rule. Such a classifier is sufficiently flexible to take into account different types of data (e.g., multispectral, panchromatic, synthetic aperture radar). The method is being experimentally validated with data acquired over a given area at different resolutions (e.g., multiresolution Pleiades images), directly integrated at the different levels of the cascade hierarchical model. An example of a classification result is illustrated in Fig. 1. Here, Pleiades multiresolution images (panchromatic resolution: 50 centimeters and multispectral resolution: 2 meters) acquired over Port-au-Prince quay (Haiti) on two different dates are considered. Spatially disjoint training areas were manually annotated. The classification has been performed with respect to 5 main classes: urban areas, natural landscape, sand, containers, and wet areas. A visual analysis of the resulting map suggests that the proposed approach achieves remarkable accuracy.



*Figure 1. Right: Initial optical image of Port-au-Prince (Haiti) (©CNES, 2013). Left: Classification map obtained with the proposed multi-temporal hierarchical method for the 5 classes (blue: wet areas; green: vegetation; red: urban areas; yellow: sand; purple: containers).*

### 6.1.2. A multi-layer Markov model for change detection in temporally separated aerial image pairs

**Participants:** Praveer Singh, 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/>].*

Multilayer Markov Random Fields (MRF), Histogram of Gradients (HOG), change detection, graph-cut optimization, aerial / satellite images.

In the proposed approach, 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).

### 6.1.3. Graph-cut model for spectral-spatial classification of hyperspectral images

**Participants:** Aakanksha Rana, Yuliya Tarabalka [contact].

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. 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.

## 6.2. Marked point processes

### 6.2.1. Marked point process models for boat extraction from high resolution remotely sensed optical images

**Participants:** Paula Craciun, Josiane Zerubia [contact].



*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).*



*Figure 3. Hyperspectral image of the University of Pavia. (a) Ground-truth (b) Support vector machines classification map. (c) Graph-cut classification map.*

This work was done in collaboration with Dr. Mathias Ortner (ASTRIUM EADS) [<http://www.astrium.eads.net>] and Prof. Pierre del Moral (ALEA team, Inria Bordeaux).

Stochastic geometry, Markov model, detection, parallel algorithm

Marked point process models have been successfully applied to object extraction problems in high resolution optical images, ranging from tree crown or road extraction to flamingo or crowd counting. We try to model the problem of boat detection and counting in harbors. The difficulty of this problem resides in the particular distribution of the objects. The model consists of two energy terms: a data term, which reflects the model's fidelity to the input image, and a prior term containing knowledge about the objects to be extracted. The model relies on a high number of parameters and is computationally intensive. The purpose of this research is to extend a previously developed marked point process model of ellipses and make it more computationally manageable. In particular, we add a preprocessing step in which we determine the global and local direction of the objects [8], [17]. Additionally, segmentation of land and water areas is implemented as a preprocessing step. Boat extraction results are shown in Figure 4. Finally, we implement an improved parallel sampler, thereby drastically improving computation times.



Figure 4. Boat extraction in a harbor using a marked point process model (a) harbor image ©CNES; (b) extraction results.



### 6.2.2. *Parameter estimation for automatic object detection in very high resolution optical images*

**Participants:** Aurélie Boisbunon, Josiane Zerubia [contact].

*This work was partially funded by the French Space Agency CNES [<http://www.cnes.fr>].*

Markov model, Monte Carlo method, evolutionary algorithm, optimization, image processing, detection

The main goal of this work is to study parameter estimation for several marked point processes. Currently, the parameters of such models are estimated by a Stochastic Expectation and Minimization (SEM) algorithm, which is computationally expensive. We will investigate and propose new parameter estimation techniques, based on Randomized Quasi-Likelihood and evolutionary algorithms, for the parameters of the probability density of a marked point process. The goal is to improve computation times with respect to SEM while maintaining similar accuracy. The first application envisioned is boat detection for harbor activity monitoring (see Figure 5).



*Figure 5. Harbor activity monitoring. ©CNES*

### 6.2.3. *Wrinkle detection using a marked point process*

**Participants:** Seong-Gyun Jeong, Yuliya Tarabalka, Josiane Zerubia [contact].

Skin image processing, wrinkle detection, line detection, marked point process, RJMCMC

We developed a novel wrinkle detection algorithm using a marked point process (MPP). Since wrinkles are the most important visual features of aging, automatic wrinkle detection algorithm can have many applications, such as the evaluation of cosmetic products, age estimation, and aging synthesis. In order to detect wrinkles of arbitrary shape, we represent wrinkles as a set of small line segments. Note that each line segment consists of a length and an orientation. A stochastic wrinkle model density exploits the local edge profile and constrains the spatial placement of adjacent lines. To maximize the model density, we employ a reversible jump Markov chain Monte Carlo (RJMCMC) sampler. A state of the Markov chain corresponds to a wrinkle configuration, and it is updated according to the acceptance ratio of sub-transition kernels: line segment births and deaths, and an affine transformation kernel. The transition kernels perturb the Markov chain by adding, removing, or modifying a wrinkle segment in the current configuration. In addition, an acceleration scheme has been developed for the RJMCMC sampler that enforces the connectivity of line segments. RJMCMC with acceleration reduces mixing time and improves detection accuracy as well.

Figure 6 compares wrinkle detection results simulated by random walk and the proposed acceleration scheme. The proposed algorithm faithfully detects wrinkles as smoothly connected lines. In addition, Figure 6 (d) plots the energy as a function of the number of iterations. It shows that the proposed acceleration method reaches a lower energy more rapidly than the random walk method.

## 6.3. Shapes and contours

### 6.3.1. Shape reconstruction from lidar data

**Participant:** Ian Jermyn [contact].

*This work is being done in collaboration with Dr. Stuart Jones, Dr. Jochen Einbeck, and PhD student Thomai Tsiftsi of Durham University, UK [<https://www.dur.ac.uk>].*

sand body, petroleum, shape, submanifold,

The cross-sectional shapes of ‘sand bodies’, ancient underground river channels filled with sediment, are of great interest in geology, and to the petroleum industry, because the shape is strongly correlated with the nature of the sediment, and in particular with its porosity, which in turn helps determine the volume fraction of crude oil contained in the sand body. The geological literature, however, only discusses simple characterizations of these shapes, and there is much room for improvement. This project aims to build probabilistic models of the cross-sectional shapes of sand bodies based on lidar point cloud data gathered from surface-projecting sand bodies by geologists in the field. Such models, when built, can be used to test the current geological classification of sand bodies, to generate new and geologically relevant classes, and to build functional models of the connection between sand body shape and oil yield.

Current work is focused on extracting reliable cross-sectional shapes from the lidar data (see Figure 7), a difficult task in itself since the sand bodies are frequently occluded or otherwise incomplete. Bayesian inference based on parameterized models of shape suggested by the current geological classification are used for this purpose. Since sand body shapes are concentrated near a low-dimensional submanifold of shape space, these models will later be extended using techniques such as mixtures built on principal curves, adapted to curved manifolds, in order to find and characterize this submanifold.

### 6.3.2. Riemannian metrics on spaces of curves and surfaces

**Participant:** Ian Jermyn [contact].

*This work is being done in collaboration with Prof. Anuj Srivastava of Florida State University [<http://www.fsu.edu>].*

Shape, Riemannian metric, elastic, curve, surface, functional data, alignment



*Figure 6. Comparison of wrinkle detection results using different simulation procedures: (b) random walk and (c) the proposed acceleration scheme. Energy as a function of the number of iterations is plotted in (d).*



*Figure 7. Left: a point cloud containing a sand body extracted from a larger cloud. Right: cross-sectional shape derived from the point cloud.*

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 defined on 'shape space', the quotient of spaces of sphere or disc embeddings by similarities or other geometric group, and the diffeomorphism group of the sphere or disc. These metrics are defined by Riemannian metrics on the space of embeddings on which the transformation groups act by isometries, and so attention is focused on understanding such metrics and their properties.

Current work is focused on two areas. The first is on classifying and describing the diffeomorphism-invariant metrics on function spaces (shapes in one dimension) that satisfy additional desiderata useful in different applications, with particular application to function alignment. The second is on generalizing to surfaces the elastic metric much used in the case of curves, and in finding surface representations that permit analytic results to be derived, or that simplify computations, in the same way that the square root velocity representation simplifies computations involving the elastic metric on curves (see Figure 8).



Figure 8. Top: interpolating surfaces based on a previous Riemannian metric. Bottom: interplooting shapes based on the generalized elastic metric.

### 6.3.3. Sampling methods for random field models of shape

**Participant:** Ian Jermyn [contact].

*Part of this work is being done in collaboration with Prof. Zoltan Kato and PhD student Csaba Molnar of the University of Szeged, Hungary [<http://www.inf.u-szeged.hu/~kato/>], and part in collaboration with PhD student Michael Racovitan of Durham University, UK [<https://www.dur.ac.uk>].*

Shape, long range interaction, Markov random field, phase field, contour, learning, wavelet

The detection and segmentation of objects from images is a problem with innumerable applications in many domains. Probabilistic models of shape, used as prior distributions in the inference process, are a necessity in solving any nontrivial instance of this problem. In many cases of importance, the shapes to be modelled cannot be treated efficiently, or at all, with current techniques, for example when multiple instances of an object must be segmented. The overall goal of this project is to develop a general shape modelling methodology capable of dealing with these difficult cases, as well as more traditional instances of the problem.

Recent algorithmic work has focused on developing efficient sampling methods for the models, for use in parameter and model learning. The models, whether expressed in terms of shape boundaries, phase fields, or binary fields, contain many long-range frustrated interactions, and hence are not amenable to standard techniques. Simplifications of the interaction structure using adapted wavelet bases, and re-expressions of the models using varieties of Hubbard-Stratanovich transformation are two directions being explored.



Figure 9. A typical result on an image of lipid cells.

### 6.3.4. Multiple-instance object detection via a third-order active contour shape model

**Participants:** Ikhlef Bechar, Ian Jermyn, Josiane Zerubia [contact].

*This work was funded by the EADS Foundation [<http://www.fondation.eads.com>].*

Object detection, multiple objects, shape, invariance, prior, higher-order active contour (HOAC), energy minimization

Recent modelling work has focused on generalizing the higher-active contour methodology to families of shapes whose members consist of an arbitrary number of object instances, each of which is similar to a given reference shape. This means finding energies on the space of regions that possess low-energy local minima corresponding to an arbitrary number of instances of the reference shape. To this end, we have studied a family of fourth-order energy functionals on regions based on a kernel given in closed form as a function of the reference region. The energy has, amongst its global minima, regions consisting of an arbitrary number of well-separated instances of the reference shape, each under an arbitrary Euclidean transformation, thereby eliminating the need to estimate group-valued 'pose' parameters. It may be combined with a likelihood energy, and the result minimized using gradient descent, speeded up by use of the Fourier domain. Although problems still remain, a series of experiments on both synthetic and real images has demonstrated the feasibility of the approach (see Figure 10).



Figure 10. Detection of a shape in a noisy infrared image (SNR = 5dB): (top row) without using prior shape knowledge, and (bottom row) using the proposed fourth-order prior shape model. First column: initialization; second column: intermediate contour; third column: final contour; fourth column: segmentation.

## 6.4. Shapes in time

### 6.4.1. Graph-based model for multitemporal segmentation of sea ice floes from satellite data

**Participants:** Claudio Price González, Yuliya Tarabalka [contact].

*This work has been done in collaboration with Dr. Ludovic Brucker (NASA GSFC, USA) [<http://www.nasa.gov>].*

Multitemporal segmentation, region growing, MODIS, sea ice floes

Automated segmentation of the evolution of sea ice from satellite images would allow scientists studying climate change to build accurate models of the sea ice meltdown process, which is a sensitive climate indicator. In this work, we proposed a new method which uses shape analysis and graph-based optimization to segment a multiyear ice floe from time series of satellite images [13]. The new approach combines data from two instruments onboard the NASA Aqua satellite, enabling several measurements per day over the Earth's polar regions: Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E); and Moderate-Resolution Imaging Spectroradiometer (MODIS). The method performs best merge region growing, followed by energy minimization on the image graph, where the energy consists of two terms describing the floe shape (shape term) and the gradient between the floe and the background (data term), respectively. We validated the performance of the proposed method for segmentation of a shrinking ice floe from a sequence of AMSR-E and MODIS images acquired in August–October 2008 (see Figure 11). The results obtained showed both the effectiveness of the proposed approach and its robustness to low-contrast data.

### 6.4.2. 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 (Computer Vision Laboratory at ETH Zurich and Asclepios team at 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 [15]. 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 [16]; 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 12 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.

## 6.5. Other detection approaches

### 6.5.1. Illumination modeling and chromophore identification in dermatological images for skin disease analysis

**Participants:** Zhao Liu, Josiane Zerubia [contact].





*Figure 11. Comparison of results for the MODIS image sequence acquired in August-October 2008. Manual segmentation of the ice floe contour is shown in green, hierarchical step-wise optimization result in red, and the new graph-based approach in blue.*



*Figure 12. 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.*

*This work is part of the LIRA Skin Care Project, which includes four key partners: Philips R&D [<http://www.research.philips.com>], CWI (Netherlands) [<http://www.cwi.nl>], Inria (France), and Fraunhofer Institutes (Germany) [<http://www.fraunhofer.de/en.html>].*

Chromophore identification, illumination modeling, skin disease analysis, dermatology

Skin color is an important characteristic for the accurate diagnosis and grading of cutaneous lesions by experienced dermatologists in clinical practice. However, the visual perception of skin color is not only a function of the major chromophores (melanin and hemoglobin) underneath the skin surface, but is also affected by external illumination and the spectral responses of imaging detectors. Skin color representation in a specific color space (e.g. RGB and its transformations) is not a genuine physical quantity. It sometimes fails to provide precise information about the concentrations of cutaneous chromophores, and is easily influenced by external imaging factors. As a result, conventional colorimetry may not properly describe the underlying histological content of skin, and hence tends to yield less trustworthy results when applied directly for skin disease analysis.

Building on a previous study that considered human skin as a diffuse reflectance surface, our work models human skin as having specular and diffuse reflectance, leading to a novel illumination correction method. Based on this method, we have developed a new scheme for chromophore identification from dermatological photographs. The algorithm has three steps. First, specular reflectance is separated from diffuse reflectance in the original skin images through specular pixel localization and image interpolation using a nonlinear weighted averaging process. Second, the resultant diffuse reflectance component is decomposed into a base layer and a detail layer. The base layer, representing low-frequency illumination and shading effects, is approximated by polynomial curve fitting using an initial illumination map using an adaptive bilateral filter as a prior. The detail layer, primarily containing high-frequency chromophore reflectance, can then be calculated by subtracting the base layer from the corresponding diffuse spectral band in logarithmic form. Finally, by incorporating knowledge of chromophore absorption characteristics, melanin and hemoglobin densities are identified using the detail layers from different spectral channels [11].

For algorithm evaluation, the method was applied to two skin disease analysis problems: computer-aided melanoma diagnosis [11] and automatic acne detection [12]. For melanoma diagnosis, 201 conventional RGB skin lesion images (62MMs, 139 benign nevi (BN)) were collected from free public databases (<http://www.dermquest.com/>, <http://www.dermis.net/>) to form an experimental data set. Figure 13 -(I) shows an example of a superficial spreading melanoma with obvious horizontal shading effects, and the corresponding experimental results. It is clear that the proposed algorithm successfully removed the imaging artifacts from the original skin lesion photographs.

For acne detection, a set of 50 challenging images were tested as a qualitative evaluation to demonstrate the usefulness of the proposed method. Automatic acne segmentation is performed using an MRF model based on chromophore descriptors. Figure 13 -(II) shows one acne example captured in an uncontrolled environment from a free public database (<http://www.dermnetz.org/>). The detected acne areas are highly consistent under visual inspection, and the inflammatory acne can be distinguished from hyperpigmentation by comparing the average values of the melanin and hemoglobin indices.



*Figure 13. Examples of different types of skin disease requiring shading removal and chromophore identification: (I-a) Original melanoma image; (I-b) Corrected melanoma image; (I-c) Melanin index map of image (I-a); (I-d) Hemoglobin index map of image (I-a); (I-e) Two-class segmentation results from Otsu's method on the original melanoma image (black line) and the corrected melanoma image (blue line), respectively; (II-a) Original acne image; (II-b) Melanin index map of image (II-a); (II-c) Hemoglobin index map of image (II-a); (II-d) Acne segmentation result using an MRF model, highlighting inflammatory acne (blue line) and hyperpigmentation (black line), respectively.*

## LEAR Project-Team

# 6. New Results

## 6.1. Visual recognition in images

### 6.1.1. Label-Embedding for Attribute-Based Classification

**Participants:** Zeynep Akata, Florent Perronnin, Zaid Harchaoui, Cordelia Schmid.

Attributes are an intermediate representation, which enables parameter sharing between classes, a must when training data is scarce. We propose in [13] to view attribute-based image classification as a label-embedding problem: each class is embedded in the space of attribute vectors. We introduce a function which measures the compatibility between an image and a label embedding, as shown in Figure 1. The parameters of this function are learned on a training set of labeled samples to ensure that, given an image, the correct classes rank higher than the incorrect ones. Results on the Animals With Attributes and Caltech-UCSD-Birds datasets show that the proposed framework outperforms the standard Direct Attribute Prediction baseline in a zero-shot learning scenario. The label embedding framework offers other advantages such as the ability to leverage alternative sources of information in addition to attributes (e.g. class hierarchies) or to transition smoothly from zero-shot learning to learning with large quantities of data.

### 6.1.2. Good Practice in Large-Scale Learning for Image Classification

**Participants:** Zeynep Akata, Florent Perronnin, Zaid Harchaoui, Cordelia Schmid.

In this paper [2], 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%.

### 6.1.3. Segmentation Driven Object Detection with Fisher Vectors

**Participants:** Ramazan Gokberk Cinbis, Jakob Verbeek, Cordelia Schmid.

In [18], we present an object detection system based on the Fisher vector (FV) image representation computed over SIFT and color descriptors. For computational and storage efficiency, we use a recent segmentation-based method to generate class-independent object detection hypotheses, in combination with data compression techniques. Our main contribution is a method to produce tentative object segmentation masks to suppress background clutter in the features. As illustrated in Figure 2, re-weighting the local image features based on these masks is shown to improve object detection significantly. We also exploit contextual features in the form of a full-image FV descriptor, and an inter-category rescoring mechanism. Our experiments on the VOC 2007 and 2010 datasets show that our detector improves over the current state-of-the-art detection results.

### 6.1.4. Image Classification with the Fisher Vector: Theory and Practice

**Participants:** Jorge Sánchez, Florent Perronnin, Thomas Mensink, Jakob Verbeek.



*Figure 1. Much work in computer vision has been devoted to image embedding (left): how to extract suitable features from an image? We focus on label embedding (right): how to embed class labels in a Euclidean space? We use attributes as side information for the label embedding and measure the “compatibility” between the embedded inputs and outputs with a function  $F$*



*Figure 2. The image on the left and the one on the right show the top detection without and with using our segmentation-driven descriptors, respectively.*

A standard approach to describe an image for classification and retrieval purposes is to extract a set of local patch descriptors, encode them into a high-dimensional vector and pool them into an image-level signature. The most common patch encoding strategy consists in quantizing the local descriptors into a finite set of prototypical elements. This leads to the popular Bag-of-Visual words (BOV) representation. In [10], we propose to use the Fisher Kernel framework as an alternative patch encoding strategy: we describe patches by their deviation from a “universal” generative Gaussian mixture model. This representation, which we call Fisher Vector (FV) has many advantages: it is efficient to compute, it leads to excellent results even with efficient linear classifiers, and it can be compressed with a minimal loss of accuracy using product quantization. We report experimental results on five standard datasets – PASCAL VOC 2007, Caltech 256, SUN 397, ILSVRC 2010 and ImageNet10K – with up to 9M images and 10K classes, showing that the FV framework is a state-of-the-art patch encoding technique. In figure 3 we show a representative benchmark performance comparison between BOV and FV representations.



Figure 3. Accuracy of the BOV and the FV as a function of the number of Gaussians (left) and feature dimensionality (right) on PASCAL VOC 2007 with SIFT descriptors only.

## 6.2. Learning and statistical models

### 6.2.1. Kernel-Based Methods for Hypothesis Testing: A Unified View

**Participants:** Zaid Harchaoui, Francis Bach, Olivier Cappe, Eric Moulines.

Kernel-based methods provide a rich and elegant framework for developing nonparametric detection procedures for signal processing. Several recently proposed procedures can be simply described using basic concepts of reproducing kernel Hilbert space embeddings of probability distributions, namely mean elements and covariance operators. In [5], we propose a unified view of these tools, and draw relationships with information divergences between distributions (see Figure 4).

### 6.2.2. Supervised Feature Selection in Graphs with Path Coding Penalties and Network Flows

**Participants:** Julien Mairal, Bin Yu.





*Figure 4. A schematic view of kernel embedding and mean element*

In this paper [6], we consider supervised learning problems where the features are embedded in a graph, such as gene expressions in a gene network. In this context, it is of much interest to automatically select a subgraph with few connected components; by exploiting prior knowledge, one can indeed improve the prediction performance or obtain results that are easier to interpret. Regularization or penalty functions for selecting features in graphs have recently been proposed, but they raise new algorithmic challenges. For example, they typically require solving a combinatorially hard selection problem among all connected subgraphs. In this paper, we propose computationally feasible strategies to select a sparse and well-connected subset of features sitting on a directed acyclic graph (DAG), see Figure 5. We introduce structured sparsity penalties over paths on a DAG called “path coding” penalties. Unlike existing regularization functions that model long-range interactions between features in a graph, path coding penalties are tractable. The penalties and their proximal operators involve path selection problems, which we efficiently solve by leveraging network flow optimization. We experimentally show on synthetic, image, and genomic data that our approach is scalable and leads to more connected subgraphs than other regularization functions for graphs.



Figure 5. Network Flow Model with Costs on Arcs for the Path Selection Problem

### 6.2.3. Structured Penalties for Log-linear Language Models

**Participants:** Anil Nelakanti, Cédric Archambeau, Julien Mairal, Francis Bach, Guillaume Bouchard.

Language models can be formalized as loglinear regression models where the input features represent previously observed contexts up to a certain length  $m$ . The complexity of existing algorithms to learn the parameters by maximum likelihood scale linearly in  $nd$ , where  $n$  is the length of the training corpus and  $d$  is the number of observed features. In this paper [26], we present a model that grows logarithmically in  $d$ , making it possible to efficiently leverage longer contexts (see Figure 6). We account for the sequential structure of natural language using tree-structured penalized objectives to avoid overfitting and achieve better generalization.

### 6.2.4. Optimization with First-Order Surrogate Functions

**Participant:** Julien Mairal.



Figure 6. The classical measure of performance for natural language models is the perplexity (lower is better). Our models are denoted by  $\ell_2^T$  and  $\ell_{\text{inf}}^T$ .

In this paper [23], we study optimization methods consisting of iteratively minimizing surrogates of an objective function, as illustrated in Figure 7. By proposing several algorithmic variants and simple convergence analyses, we make two main contributions. First, we provide a unified viewpoint for several first-order optimization techniques such as accelerated proximal gradient, block coordinate descent, or Frank-Wolfe algorithms. Second, 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 7. 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.

### 6.2.5. Stochastic Majorization-Minimization Algorithms for Large-Scale Optimization

**Participant:** Julien Mairal.

Majorization-minimization algorithms consist of iteratively minimizing a majorizing surrogate of an objective function. Because of its simplicity and its wide applicability, this principle has been very popular in statistics and in signal processing. In this paper [24], we intend to make this principle scalable. We introduce a stochastic majorization-minimization scheme which is able to deal with large-scale or possibly infinite data sets. When applied to convex optimization problems under suitable assumptions, we show that it achieves an expected convergence rate of  $O(1/\sqrt{n})$  after  $n$  iterations, and of  $O(1/n)$  for strongly convex functions. Equally important, our scheme almost surely converges to stationary points for a large class of non-convex problems. We develop several efficient algorithms based on our framework. First, we propose a new stochastic proximal gradient method, which experimentally matches state-of-the-art solvers for large-scale  $\ell_1$ -logistic regression.

Second, we develop an online DC programming algorithm for non-convex sparse estimation. Finally, we demonstrate the effectiveness of our approach for solving large-scale structured matrix factorization problems.

## 6.3. Recognition in video

### 6.3.1. Temporal Localization of Actions with Actoms

**Participants:** Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

In this paper [4], we address the problem of localizing actions, such as opening a door, in hours of challenging video data. We propose a model based on a sequence of atomic action units, termed "actoms", that are semantically meaningful and characteristic for the action. Our Actom Sequence Model (ASM) represents an action as a sequence of histograms of actom-anchored visual features, which can be seen as a temporally structured extension of the bag-of-features. Training requires the annotation of actoms for action examples. At test time, actoms are localized automatically based on a non-parametric model of the distribution of actoms, which also acts as a prior on an action's temporal structure. We present experimental results on two recent benchmarks for action localization "Coffee and Cigarettes" and the "DLSBP" dataset. We also adapt our approach to a classification-by-localization set-up, and demonstrate its applicability on the challenging "Hollywood 2" dataset. We show that our ASM method outperforms the current state of the art in temporal action localization, as well as baselines that localize actions with a sliding window method (see Figure 8).



Figure 8. Illustration of actoms-based decomposition of actions.

### 6.3.2. Activity representation with motion hierarchies

**Participants:** Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

Complex activities, e.g., pole vaulting, are composed of a variable number of sub-events connected by complex spatio-temporal relations, whereas simple actions can be represented as sequences of short temporal parts. In [3], we learn hierarchical representations of activity videos in an unsupervised manner. These hierarchies of mid-level motion components are data-driven decompositions specific to each video. We introduce a spectral divisive clustering algorithm to efficiently extract a hierarchy over a large number of tracklets (i.e., local trajectories). We use this structure to represent a video as an unordered binary tree. We model this tree using nested histograms of local motion features. We provide an efficient positive definite kernel that computes the structural and visual similarity of two hierarchical decompositions by relying on models of their parent-child relations. We present experimental results on four recent challenging benchmarks: the High Five dataset, the Olympics Sports dataset, the Hollywood 2 dataset, and the HMDB dataset. We show that per-video hierarchies provide additional information for activity recognition. Our approach improves over unstructured activity models, baselines using other motion decomposition algorithms, and the state of the art (see Figure 9).



Figure 9. Illustration of motion hierarchies for weight-lifting.

### 6.3.3. DeepFlow: Large displacement optical flow with deep matching

**Participants:** Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, Cordelia Schmid.

Optical flow computation is a key component in many computer vision systems designed for tasks such as action detection or activity recognition. However, despite several major advances over the last decade, handling large displacement in optical flow remains an open problem. Inspired by the large displacement optical flow of Brox and Malik, our approach, termed DeepFlow, blends a matching algorithm with a variational approach for optical flow. We propose in [31] a descriptor matching algorithm, tailored to the optical flow problem, that allows to boost performance on fast motions. The matching algorithm builds upon a multi-stage architecture with 6 layers, interleaving convolutions and max-pooling, a construction akin to deep convolutional nets. Figure 10 shows an outline of our approach. Using dense sampling, it allows to efficiently retrieve quasi-dense correspondences, and enjoys a built-in smoothing effect on descriptors matches, a valuable asset for integration into an energy minimization framework for optical flow estimation. DeepFlow efficiently handles large displacements occurring in realistic videos, and shows competitive performance on optical flow benchmarks. Furthermore, it sets a new state-of-the-art on the MPI-Sintel dataset.

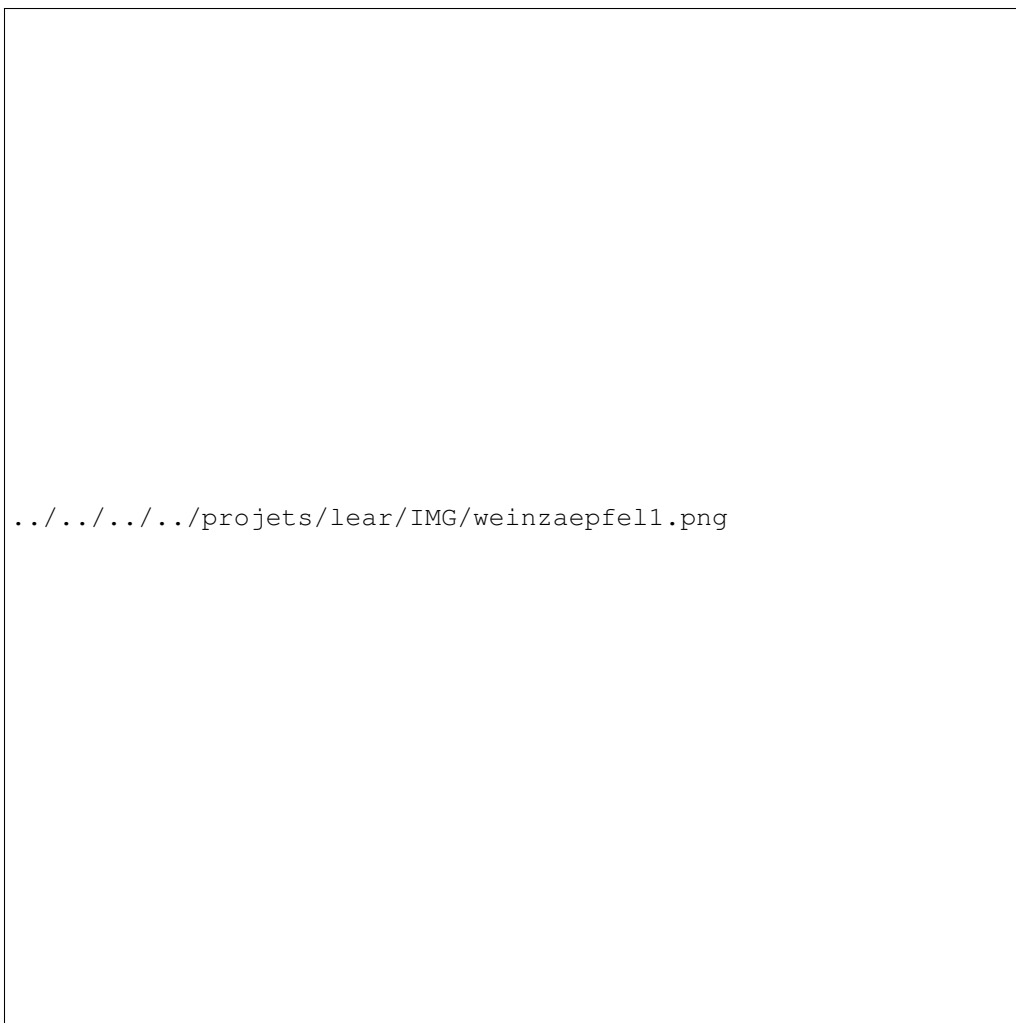


Figure 10. Outline of DeepFlow.

#### 6.3.4. Event retrieval in large video collections with circulant temporal encoding

**Participants:** Jerome Revaud, Matthijs Douze, Cordelia Schmid, Hervé Jégou.

This paper [28] presents an approach for large-scale event retrieval. Given a video clip of a specific event, e.g., the wedding of Prince William and Kate Middleton, the goal is to retrieve other videos representing the same event from a dataset of over 100k videos. Our approach encodes the frame descriptors of a video to jointly represent their appearance and temporal order. It exploits the properties of circulant matrices to compare the videos in the frequency domain. This offers a significant gain in complexity and accurately localizes the matching parts of videos, see Figure 11 . Furthermore, we extend product quantization to complex vectors in order to compress our descriptors, and to compare them in the compressed domain. Our method outperforms the state of the art both in search quality and query time on two large-scale video benchmarks for copy detection, Trecvid and CCweb. Finally, we introduce a challenging dataset for event retrieval, EVVE, and report the performance on this dataset.

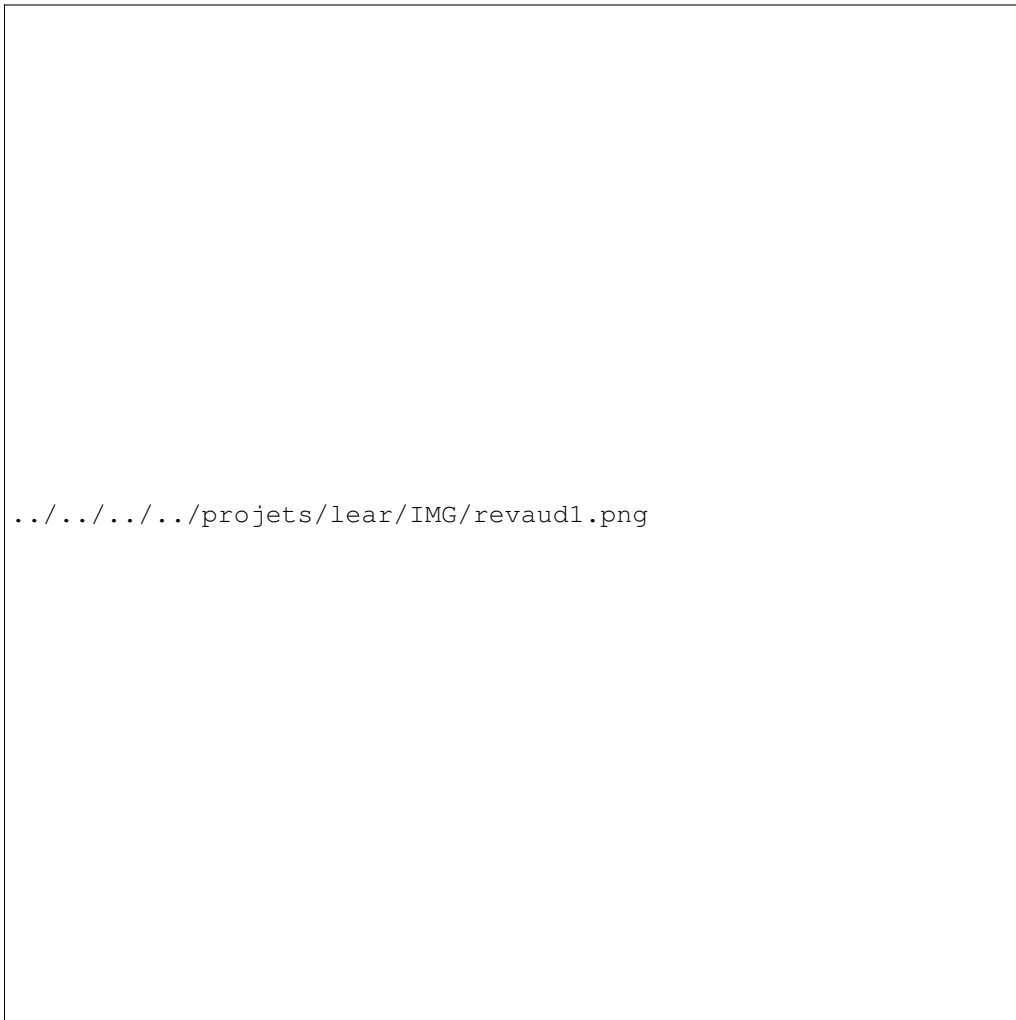


Figure 11. Example of correctly aligned videos. Each row is a different video, and each column corresponds to temporally aligned frames from the videos.



### 6.3.5. Dense trajectories and motion boundary descriptors for action recognition

**Participants:** Heng Wang, Alexander Kläser, Cordelia Schmid, Cheng-Lin Liu.



*Figure 12. Illustration of the information captured by HOG, HOF, and MBH descriptors. Gradient/flow orientation is indicated by color (hue) and magnitude by saturation. The optical flow (top, middle) shows constant motion in the background, which is due to the camera movements. The motion boundaries (right) encode the relative motion between the person and the background.*

This paper [11] introduces a video representation based on dense trajectories and motion boundary descriptors. Trajectories capture the local motion information of the video. A state-of-the-art optical flow algorithm enables a robust and efficient extraction of the dense trajectories. As descriptors we extract features aligned with the trajectories to characterize shape (point coordinates), appearance (histograms of oriented gradients) and motion (histograms of optical flow). Additionally, we introduce a descriptor based on motion boundary histograms (MBH) (see the visualization in Figure 12), which is shown to consistently outperform other state-of-the-art descriptors, in particular on real-world videos that contain a significant amount of camera motion.

We evaluate our video representation in the context of action classification on nine datasets, namely KTH, YouTube, Hollywood2, UCF sports, IXMAS, UIUC, Olympic Sports, UCF50 and HMDB51. On all datasets our approach outperforms current state-of-the-art results.

### 6.3.6. Action Recognition with Improved Trajectories

**Participants:** Heng Wang, Cordelia Schmid.



Figure 13. Visualization of human detection and inlier matches (top, left) as well as removed background trajectories, which are due to camera motion (top, right). The bottom row compares the original optical flow (bottom, left) and the warped version (bottom, right).

This paper [30] improves dense trajectories by taking into account camera motion to correct them. To estimate camera motion, we match feature points between frames using SURF descriptors and dense optical flow, which are shown to be complementary. These matches are, then, used to robustly estimate a homography with RANSAC. Human motion is in general different from camera motion and generates inconsistent matches. To improve the estimation, a human detector is employed to remove these matches. Given the estimated camera motion, we remove trajectories consistent with it. We also use this estimation to cancel out camera motion from the optical flow. This significantly improves motion-based descriptors, such as HOF and MBH (see Figure 13). Experimental results on four challenging action datasets (i.e., Hollywood2, HMDB51, Olympic Sports and UCF50) significantly outperform the current state of the art.

### 6.3.7. Action and event recognition with Fisher vectors on a compact feature set

**Participants:** Dan Oneață, Jakob Verbeek, Cordelia Schmid.

Action recognition in uncontrolled video is an important and challenging computer vision problem. Recent progress in this area is due to new local features and models that capture spatio-temporal structure between local features, or human-object interactions. Instead of working towards more complex models, we focus in this paper [27] on the low-level features and their encoding. We evaluate the use of Fisher vectors as an alternative to bag-of-word histograms to aggregate a small set of state-of-the-art low-level descriptors, in combination with linear classifiers. We present a large and varied set of evaluations, considering (i) classification of short actions in five datasets, (ii) localization of such actions in feature-length movies, and (iii) large-scale recognition of complex events. We find that for basic action recognition and localization MBH features alone are enough for state-of-the-art performance. For complex events we find that SIFT and MFCC features provide complementary cues. On all three problems we obtain state-of-the-art results, while using fewer features and less complex models.

### 6.3.8. *Stable hyper-pooling and query expansion for event detection*

**Participants:** Matthijs Douze, Jerome Revaud, Cordelia Schmid, Hervé Jégou.

This work [19] makes two complementary contributions to event retrieval in large collections of videos. First, we compare different ways of quantizing video frame descriptors in terms of temporal stability. Our best choices compare favorably with the standard pooling technique based on k-means quantization, see Figure 14. Second, we introduce a technique to improve the ranking. It can be interpreted either as a query expansion method or as a similarity adaptation based on the local context of the query video descriptor. Experiments on public benchmarks show that our methods are complementary and improve event retrieval results, without sacrificing efficiency.

### 6.3.9. *Finding Actors and Actions in Movies.*

**Participants:** Piotr Bojanowski, Francis Bach, Ivan Laptev, Jean Ponce, Cordelia Schmid, Josef Sivic.

This work [16] addresses the problem of learning a joint model of actors and actions in movies using weak supervision provided by scripts. Specifically, we extract actor/action pairs from the script and use them as constraints in a discriminative clustering framework. The corresponding optimization problem is formulated as a quadratic program under linear constraints. People in video are represented by automatically extracted and tracked faces together with corresponding motion features. First, we apply the proposed framework to the task of learning names of characters in movies and demonstrate significant improvements over previous methods used for this task. Second, we explore joint actor/action constraints and show their advantage for weakly supervised action learning. We validate our method in the challenging setting of localizing and recognizing characters and their actions in the feature length movies Casablanca and American Beauty. Figure 15 shows an example of our results.



*Figure 14. Several quantizations of video frame descriptors (left) to a color-coded index in  $\{0, \dots, 31\}$ . Leftmost column: standard  $k$ -means, right: the proposed SSC. Time runs vertically.*



*Figure 15. Automatic detection and annotation of characters and their actions in the movie Casablanca. The automatically resolved correspondence between video and script is color-coded.*

## MAGRIT Project-Team

# 6. New Results

## 6.1. Motion, Scene and Camera Reconstruction

**Participants:** Marie-Odile Berger, Srikrishna Bhat, Pierre Rolin, Gilles Simon, Frédéric Sur.

- **Metrological performance enhancement and resolution assessment for experimental solid mechanics**

This work is motivated by image processing problems from experimental solid mechanics. One of the problem in this field is to measure heterogeneous strains on the surface of specimens subjected to mechanical tests, through an imaging device. Among full-field measurement techniques, the grid method consists in transferring a regular grid on the surface of the specimen and in taking images of the grid before and after deformation. Windowed Fourier analysis then gives an estimation of the surface displacement and strain components. In a collaboration with Institut Pascal (Université Blaise Pascal, Clermont Ferrand), we have shown that the estimations obtained by this technique are a first-order approximation of the convolution of the actual values with the analysis window. We have also characterized how the noise in the grid image impairs the displacement and strain maps<sup>3</sup>. This study has allowed us to improve the metrological performance of the grid method with deconvolution algorithms. A numerical and experimental study can be found in [10], [16], [21]. As any contactless measurement method, the resolution of the grid method is limited by the noise impairing the sensor. We have also characterized this resolution within a Poisson-Gaussian noise model, which is known to be realistic for CCD or CMOS sensors<sup>4</sup>.

- **Matching in difficult conditions**

Visual vocabularies are emerging as a new tool for building point correspondences for pose estimation. Within S. Bhat's PhD thesis [9] we have proposed several methods for visual word construction dedicated to point matching, with structure from motion and pose estimation applications in view. The three dimensional geometry of a scene is first extracted with bundle adjustment techniques based on keypoint correspondences. These correspondences are obtained by grouping the set of all SIFT descriptors from the training images into visual words. We obtain a more accurate 3D geometry than with classical image-to-image point matching. In a second on-line step, these visual words serve as 3D point descriptors that are robust to viewpoint change, and are used for building 2D-3D correspondences on-line during application, yielding the pose of the camera by solving the PnP problem. Several visual word formation techniques have been compared with respect to robustness to viewpoint change between learning and the test images.

The PhD thesis of P. Rolin comes within the scope of camera pose estimation from an unstructured 3D point dataset, endowed with image descriptors. His work focuses on improving pose estimation with respect to strong viewpoint changes. 2D-3D correspondences are actually difficult to establish if there are too large viewpoint changes between the image whose pose is sought and the images that gave the 3D point dataset. P. Rolin currently assesses viewpoint simulation techniques in order to enhance the description of the 3D points with information from different viewpoints.

- **Acquisition of 3D calibrated data**

---

<sup>3</sup>F. Sur, M. Grédiac. Towards deconvolution to enhance the grid method for in-plane strain measurement. To appear in AIMS Inverse Problems and Imaging, American Institute of Mathematical Sciences, 2014.

<sup>4</sup>M. Grédiac, F. Sur. Effect of sensor noise on the resolution and spatial resolution of displacement and strain maps estimated with the grid method. To appear in Strain, Wiley, 2014.

In situ modeling is generating increasing interest in the community as it makes it possible to build AR applications in unprepared environments. In [19], we present a new method for interactive modeling of polygonal scenes, using a tablet PC, a laser rangefinder, an inertial measurement unit (IMU) and a camera. A well-founded calibration method is used to determine the orientation of the IMU and the origin and direction of the laser beam in the camera coordinate system. A new hybrid, driftless orientation tracking method is proposed, inspired by the tracking-by-synthesis algorithm adapted to 3-degree-of-freedom camera motions. Visual hints are provided during the tracking-and-modeling process in order to help the user get the best possible accuracy. These visual hints are based on a PCA analysis of the reconstructed laser point clouds and statistical measurements of the camera tracking accuracy.

## 6.2. Medical Imaging

**Participants:** René Anxionnat, Marie-Odile Berger, Nazim Haouchine, Erwan Kerrien, Matthieu Loosvelt, Pierre-Frédéric Villard, Brigitte Wrobel-Dautcourt, Ahmed Yureidini.

- **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. Most of all, and despite recent progress on the sensors, X-rays are bad for the patient's health and X-ray images are 2D projections deprived of any depth hint such as occlusions or shading. To quote a fellow physician: "it is rather uncanny to use 2D images to perform a gesture that is, by nature, 3D". Two of our long term aims in interventional neuroradiology are to reduce the operation time, and provide the interventional radiologists with a real-time visual feedback in 3D.

All our research activity in this field is led in collaboration with the Department of Interventional Neuroradiology from Nancy University Hospital. This year was pivotal in this activity where some projects ended and other new projects started.

We've been collaborating with Shacra Inria project-team (Lille-Nord Europe) in the context of the SOFA-InterMedS Inria Large-Scale Initiative for 4 years. Ahmed Yureidini is on the verge of defending his PhD thesis and the last step of his work consisted in validating the model he devised for the blood vasculature as a tree of local implicit surfaces [8]. Comparisons were made against simulations using triangular meshes against our implicit model and they showed a reduction by 2 orders of magnitude in computing time while numerical instabilities encountered with meshes (jaggy motions, unrealistic sticking of the catheter tip on the vessel surface, ...) were not observed with our implicit model. Publication of these results is under way.

We also collaborate with Shacra team within the ANR IDeaS project. Computer simulations are very sensitive to inaccuracies in the various mechanical parameters or geometrical boundary conditions. Such inaccuracies are ubiquitous when dealing with patient-based data. We aim at developing Image-Driven Simulation to add the live X-ray images as new constraints to make the simulated surgical tool virtual visualization fit their position seen in the actual images. This year, a sensor was designed and tested to capture the motion of the line-shaped micro-tools (catheters, guidewires, etc...) and progress was made to design Kalman-like filters compliant with Sofa simulation platform.

Our long-term collaboration with GE Healthcare took a new step this year with the arrival of Charlotte Delmas as a PhD student. She will work towards devising algorithms to reconstruct the micro-tools in 3D from fluoroscopy images.

- **Designing respiration models for patient based simulators**

Respiration models are useful in many ways. They can be used in: 1) pulmonary radiotherapy, where the tumor displacement should be accurately known to be targeted by ionizing radiation, 2) thoracic surgery simulators, where breathing motion increases the realism of virtual patients, 3) interventional radiology, where augmented medical imaging that incorporates breathing motion can be used during treatment.

However building and parameterizing a fast and accurate respiration model is still an open problem. We continue this year to work on evolutionary methods to estimate the parameters of a complex 15-D respiration model on 5 patients [23]. A compound fitness function has been designed to take into account various quantities that have to be minimized.

The optimized parameters have been applied to an interventional radiology simulator that takes into account the respiration [14]. It also includes: segmentation, physically based modeling, haptics rendering, pseudo-ultrasound generation and the concept of a physical mannequin. It is the result of a close collaboration between different universities (Liverpool, Manchester, Imperial College, Bangor, Leeds, Hull) involving computer scientists, clinicians, clinical engineers and occupational psychologists.

- **Realistic simulation of organ dissection** Whilst laparoscopic surgical simulators are becoming increasingly realistic they can not, as yet, fully replicate the experience of live surgery. In particular tissue dissection is one task that is particularly challenging to replicate. Limitation of current attempts to simulate tissue dissection include: poor visual rendering; over simplification of the task and; unrealistic tissue properties. In an effort to generate a more realistic model of tissue dissection in laparoscopic surgery we worked on a novel method based on task analysis. Initially we have chosen to model only the basic geometrics of this task rather than a whole laparoscopic procedure. This year preliminary work has led to the development of a real time simulator performing organ dissection with a haptic thread at 1000Hz. 2D soft-tissue models replicate the process of tissue cutting.

- **Physics-based augmented reality**

The development of AR systems for use in the medical field faces one major challenge: the correct superposition of pre-operative data onto intraoperative images. This task is especially difficult when laparoscopic surgery is considered since superposition must be achieved on deformable organs. Most existing AR systems only consider rigid registration between the pre and intraoperative data and the transformation is often computed interactively or from markers attached to the patient's body.

In cooperation with the Shacra team, we have proposed in [17], [18] a framework for real-time augmentation of the vascular network and tumors during minimally invasive liver surgery. Internal structures computed from pre-operative CT scans can be overlaid onto the laparoscopic view for surgery guidance. Compared to state-of-the-art methods, our method uses a real-time biomechanical model to compute a volumetric displacement field from partial three-dimensional liver surface motion.

The main contributions of this work are threefold: a) the use of a biomechanical model of liver deformation allows us to account for heterogeneity and anisotropy due to veins and arteries. In addition, the physical model is used as regularizer for the unreliable measurement of the visual tracking and as motion compensation in poorly textured areas; b) a real-time implementation of this virtual liver model has been proposed c) appropriate boundary conditions and external force have been defined which guide the biomechanical model using partial 3D motion estimated at the liver surface from a stereo video stream.

Thanks to this framework, we are able to estimate, in real-time, relevant positions of internal structures of the liver (vessels and tumors) taking into account liver deformations and tissue heterogeneity.



## MORPHEO Team

# 6. New Results

## 6.1. Robust human body shape and pose tracking

This work considers markerless human performance capture from multiple camera videos and, in particular, the recovery of both shape and parametric motion information, as often required in applications that produce and manipulate animated 3D contents using multiple videos. To this aim, an approach is proposed that jointly estimates skeleton joint positions and surface deformations by fitting a reference surface model to 3D point reconstructions. The approach is based on a probabilistic deformable surface registration framework coupled with a bone binding energy. The former makes soft assignments between the model and the observations while the latter guides the skeleton fitting. The main benefit of this strategy lies in its ability to handle outliers and erroneous observations frequently present in multi view data. For the same purpose, we also introduce a learning based method that partitions the point cloud observations into different rigid body parts that further discriminate input data into classes in addition to reducing the complexity of the association between the model and the observations. We argue that such combination of a learning based matching and of a probabilistic fitting framework efficiently handle unreliable observations with fake geometries or missing data and hence, it reduces the need for tedious manual interventions. The work was presented at the 3DV conference [7] where it received the best paper runner up award.



Figure 4. Human pose recovery with 3 different standard datasets.

## 6.2. Inverse dynamics on rock climbing with and without measurement of contact forces

Rock climbing involves complex interactions of the body with the environment (Figure 5). It represents an interesting problem in biomechanics as multiple contacts in the locomotion task make it an underconstrained problem. In this study we are interested in evaluating how a climber transfers weight through the holds. The motivation of this study is also technical as we are developing an inverse dynamics method that automatically estimates in 3D, not only the usual torques at joint angles, but also the wrenches at contacts [9].



Figure 5. Inverse dynamics on rock climbing with and without measurement of contact forces.

### 6.3. Video-based methodology for markerless human motion analysis

This study presents a video-based experiment for the study of markerless human motion. Silhouettes are extracted from a multi-camera video system to reconstruct a 3D mesh for each frame using a reconstruction method based on visual hull. For comparison with traditional motion analysis results, we set up an experiment integrating video recordings from 8 video cameras and a Vicon™ marker-based motion capture system (Figure 6). Our preliminary data provided distances between the 3D trajectories from the Vicon system and the 3D mesh extracted from the video cameras. In the long term, the main ambition of this method is to provide measurement of skeleton motion for human motion analyses while eliminating markers [8].

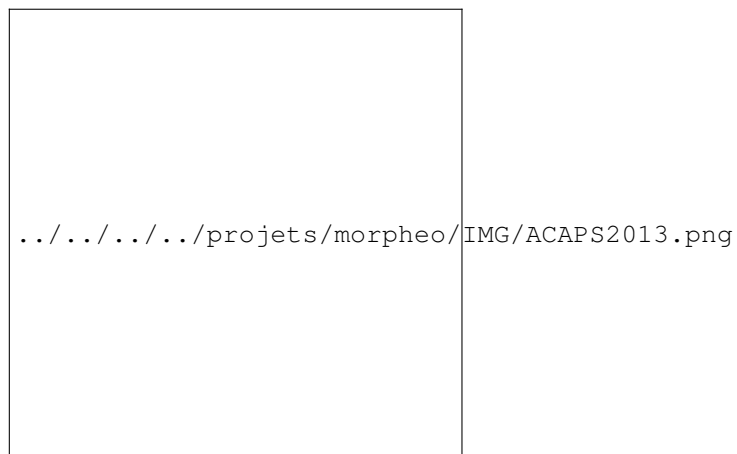


Figure 6. Video-based methodology for markerless human motion analysis.

## 6.4. 3D shape cropping

We introduce shape cropping as the segmentation of a bounding geometry of an object as observed by sensors with different modalities. Segmenting a bounding volume is a preliminary step in many multi-view vision applications that consider or require the recovery of 3D information, in particular in multi-camera environments. Recent vision systems used to acquire such information often combine sensors of different types, usually color and depth sensors. Given depth and color images we present an efficient geometric algorithm to compute a polyhedral bounding surface that delimits the region in space where the object lies. The resulting cropped geometry eliminates unwanted space regions and enables the initialization of further processes including surface refinements. Our approach exploits the fact that such a region can be defined as the intersection of 3D regions identified as non empty in color or depth images. To this purpose, we propose a novel polyhedron combination algorithm that overcomes computational and robustness issues exhibited by traditional intersection tools in our context. We show the correction and effectiveness of the approach on various combination of inputs. This work was presented at the Vision Modeling and Visualization workshop 2013 [6].

## 6.5. Multi-view object segmentation in space and time

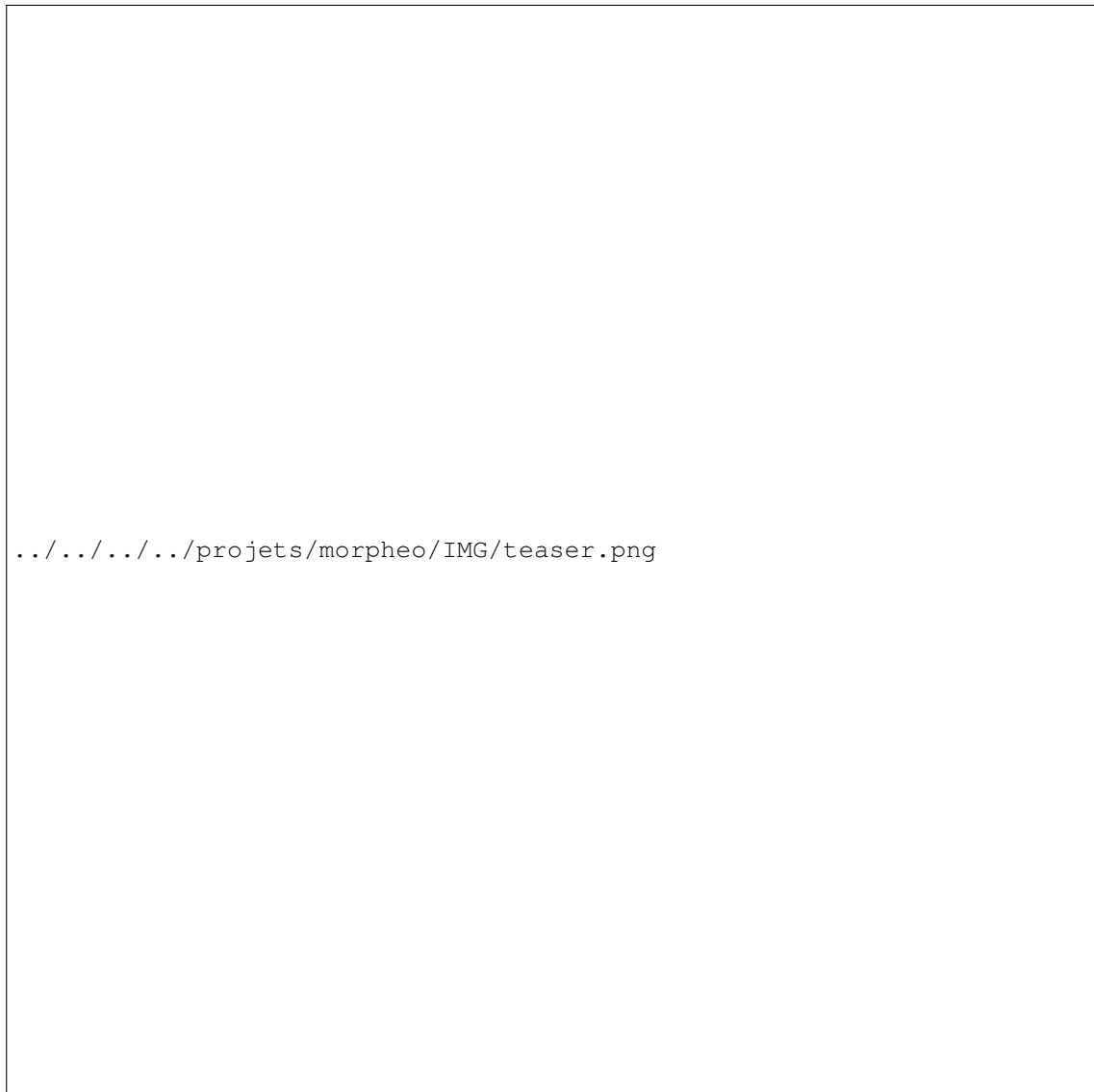
In this work, we address the problem of object segmentation in multiple views or videos when two or more viewpoints of the same scene are available. We propose a new approach that propagates segmentation coherence information in both space and time, hence allowing evidences in one image to be shared 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, as well as on videos. With static views, results compete with state of the art methods but they are achieved with significantly fewer viewpoints. With multiple videos, we report results that demonstrate the benefit of segmentation propagation through temporal cues, in ICCV 2013 [5].

## 6.6. Segmentation of temporal mesh sequences into rigidly moving components

This work considers the segmentation of meshes into rigid components given temporal sequences of deforming meshes (Figure 9). We have proposed a fully automatic approach that identifies model parts that consistently move rigidly over time. This approach can handle meshes independently reconstructed at each time instant. It allows therefore for sequences of meshes with varying connectivities as well as varying topology. It incrementally adapts, merges and splits segments along a sequence based on the coherence of motion information within each segment. In order to provide tools for the evaluation of the approach, we also introduce new criteria to quantify a mesh segmentation. Results on both synthetic and real data as well as comparisons are provided in the paper [1].

## 6.7. Segmentation of plant point cloud models into elementary units

High-resolution terrestrial Light Detection And Ranging (tLiDAR), a 3-D remote sensing technique, has recently been applied for measuring the 3-D characteristics of vegetation from grass to forest plant species. The resulting data are known as a point cloud which shows the 3-D position of all the hits by the laser beam giving a raw sketch of the spatial distribution of plant elements in 3-D, but without explicit information on their geometry and connectivity. In this study we propose a new approach based on a delineation algorithm that clusters a point cloud into elementary plant units. The algorithm creates a graph (points + edges) to recover plausible neighbouring relationships between the points and embed this graph in a spectral space in order to segment the point-cloud into meaningful elementary plant units. Our approach is robust to inherent geometric outliers and/or noisy points and only considers the x, y, z coordinate tLiDAR data as an input. It has been presented at the FSPM conference [4].



*Figure 7. Result of shape cropping using three input depth maps for polyhedral reconstruction.*



*Figure 8. Multi-view object segmentation using our method with the 3 wide-baseline views shown only, with no photo-consistency hypothesis and no user interaction.*



*Figure 9. Segmentation of temporal mesh sequences into rigidly moving components.*



*Figure 10. Segmentation of a plant point cloud model into elementary units.*

## PERCEPTION Team

### 6. New Results

#### 6.1. High-resolution depth maps based on TOF-stereo fusion

The combination of range sensors with color cameras can be very useful for a wide range of applications, e.g., robot navigation, semantic perception, manipulation, and telepresence. Several methods of combining range- and color-data have been investigated and successfully used in various robotic applications. Most of these systems suffer from the problems of noise in the range-data and resolution mismatch between the range sensor and the color cameras, since the resolution of current range sensors is much less than the resolution of color cameras. High-resolution depth maps can be obtained using stereo matching, but this often fails to construct accurate depth maps of weakly/repetitively textured scenes, or if the scene exhibits complex self-occlusions. Range sensors provide coarse depth information regardless of presence/absence of texture. The use of a calibrated system, composed of a time-of-flight (TOF) camera and of a stereoscopic camera pair, allows data fusion thus overcoming the weaknesses of both individual sensors. We propose a novel TOF-stereo fusion method based on an efficient seed-growing algorithm which uses the TOF data projected onto the stereo image pair as an initial set of correspondences. These initial “seeds” are then propagated based on a Bayesian model which combines an image similarity score with rough depth priors computed from the low-resolution range data. The overall result is a dense and accurate depth map at the resolution of the color cameras at hand. We show that the proposed algorithm outperforms 2D image-based stereo algorithms and that the results are of higher resolution than off-the-shelf color-range sensors, e.g., Kinect. Moreover, the algorithm potentially exhibits real-time performance on a single CPU. Two journal papers were submitted in 2013 and currently they are under review.

#### 6.2. 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 video alignment technique, dynamic *frame* warping (DFW), which performs isolated recognition based on a per-frame representation of videos and on aligning a test sequence with a model sequence. Next we devise two extensions which are able to perform action recognition and video segmentation in a concomitant manner, namely one-pass DFW and two-pass DFW. Both these algorithms have their roots in the continuous speech recognition domain but, to the best of our knowledge, their extension to visual recognition of actions and activities has been overlooked. We test and illustrate the proposed methods with several public-domain datasets and we compare both the isolated and continuous recognition algorithms with several recently published methods. One journal paper was submitted in 2013 and currently is under review.

#### 6.3. High-dimensional regression

We addressed the problem of approximating high-dimensional data with a low-dimensional representation. We make the following contributions. We propose an inverse regression method which exchanges the roles of input and response, such that the low-dimensional variable becomes the regressor, and which is tractable. We introduce a mixture of locally-linear probabilistic mapping model that starts with estimating the parameters of inverse regression, and follows with inferring closed-form solutions for the forward parameters of the high-dimensional regression problem of interest. Moreover, we introduce a partially-latent paradigm, 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. We devise expectation-maximization (EM) procedures based on a data augmentation strategy which facilitates the maximum-likelihood search over the model parameters. We propose two augmentation schemes and we

describe in detail the associated EM inference procedures that 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. We provide experimental evidence that our method outperforms several existing regression techniques. See [26], [12].

#### 6.4. Simultaneous sound-source separation and localization

Human-robot communication is often faced with the difficult problem of interpreting ambiguous auditory data. For example, the acoustic signals perceived by a humanoid with its on-board microphones contain a mix of sounds such as speech, music, electronic devices, all in the presence of attenuation and reverberations. We proposed a novel method, based on a generative probabilistic model and on active binaural hearing, allowing a robot to robustly perform sound-source separation and localization. We show how interaural spectral cues can be used within a constrained mixture model specifically designed to capture the richness of the data gathered with two microphones mounted onto a human-like artificial head. We describe in detail a novel expectation-maximization (EM) algorithm that alternates between separation and localization, we analyze its initialization, speed of convergence and complexity, and we assess its performance with both simulated and real data. Subsequently, we studied the *binaural manifold*, i.e., the low-dimensional space of sound-source locations embedded in the high-dimensional space of perceived interaural spectral features, and we provided a method for mapping interaural cues onto source locations. See [21], [12]. A journal paper was submitted in 2013 and accepted with minor revisions.

#### 6.5. The geometry of non-coplanar microphone arrays

We addressed 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. 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 clearly 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. See [18], [25], [11].

#### 6.6. Audiovisual calibration and alignment

We addressed the problem of aligning visual (V) and auditory (A) data using a sensor that is composed of a camera-pair and a microphone-pair. The original contribution of the paper is a method for AV data aligning through estimation of the 3D positions of the microphones in the visual-centred coordinate frame defined by the stereo camera-pair. We exploit the fact that these two distinct data sets are conditioned by a common set of parameters, namely the (unknown) 3D trajectory of an AV object, and derive an EM-like algorithm that alternates between the estimation of the microphone-pair position and the estimation of the AV object trajectory. The proposed algorithm has a number of built-in features: it can deal with A and V observations that are misaligned in time, it estimates the reliability of the data, it is robust to outliers in both modalities, and it has proven theoretical convergence. We report experiments with both simulated and real data. See [24] (this work received the best paper award).



## **6.7. Audiovisual fusion for human-robot interaction**

Natural human-robot interaction in complex and unpredictable environments is one of the main research lines in robotics. In typical real-world scenarios, humans are at some distance from the robot and the acquired signals are strongly impaired by noise, reverberations and other interfering sources. In this context, the detection and localisation of speakers plays a key role since it is the pillar on which several tasks (e.g.: speech recognition and speaker tracking) rely. We address the problem of how to detect and localize people that are both seen and heard by a humanoid robot. We introduce a hybrid deterministic/probabilistic model. Indeed, the deterministic component allows us to map the visual information into the auditory space. By means of the probabilistic component, the visual features guide the grouping of the auditory features in order to form AV objects. The proposed model and the associated algorithm are implemented in real-time (17 FPS) using a stereoscopic camera pair and two microphones embedded into the head of the humanoid robot NAO. We performed experiments on (i) synthetic data, (ii) a publicly available data set and (iii) data acquired using the robot. The results we obtained validate the approach and encourage us to further investigate how vision can help robot hearing. See [19], [20], [27], [11], [13]

## Prima Project-Team

### 5. New Results

#### 5.1. Attention-Based Navigation

**Participants:** Adrian Bourgaud, Carlos Di Pietro, Thierry Fraichard, Rémi Paulin, Patrick Reignier, Andre Van Den Berg.

Assistant robots and robot companions are designed to share the human living space, to navigate among and interact with human beings. From the mobility point of view, roboticists have recently striven to develop navigation scheme geared towards achieving so-called “socially acceptable motions”. To that end, various concepts borrowed from environmental psychology and anthropology have been used, the “personal space” concept from Proxemics being perhaps the most widely used. The purpose of our work here is to further the research in this area by taking into account other factors such as human activities, interaction configurations and intentions. An attentional model derived from cognitive psychology is used to dynamically determine the “focus of attention” of the persons involved in a given task. Depending on the task at hand, the robot uses the attention information in order to decide its future course of action so as, for instance, to attract one person’s attention or, on the contrary, to minimize the disturbance caused. In 2013, a paper describing the first results obtained was presented during the Israeli Conf. on Robotics [14].

#### 5.2. 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 Platform datasets.

#### 5.3. iKio, a sociable kiosk

**Participants:** Rémi Barraquand [correspondant], Jiří Pytela, Johan Girod.

In the Personal Assisted Living project we investigate the design of iKio: a sociable kiosk. A simple sketch of the iKio is illustrated in figure 6 . The general idea is to enhance the interaction ability of tablet and smartphone. What motivates the choice of this type of devices is the observation that people have come to treat these gadgets as their own body appendage. As pointed out by the recent study conducted by the Pew Research Center <sup>2</sup>, people are starting to use their phones and tablets for more sensitive activities that were almost considered taboo in the past, also these devices are becoming substitute for other traditional devices like photo and video cameras.

---

<sup>2</sup><http://pewinternet.org/>

The design of iKio is therefore influenced by this emerging form of symbiosis and aims to enhance both user-experience and human-technologies interaction. As follow, iKio does not have a fixed body per se, instead it is embodied in a tablet which can be carried along with people in their daily activities but which can also be docked into any mechanical structure that will provide it with enhanced abilities. Using such mechanical structure iKio can express emotion and interact more easily in the physical space of people. The core of iKio is specifically designed to handle and to support ostensive-inferential communication which is characteristic of human communication in contrast with the code model of communication argued to be the main reason of unadapted and autistic interaction between technologies and human. An early prototype of iKio is illustrated in figure 6 . It was constructed using the Bioloid <sup>3</sup> construction kit.

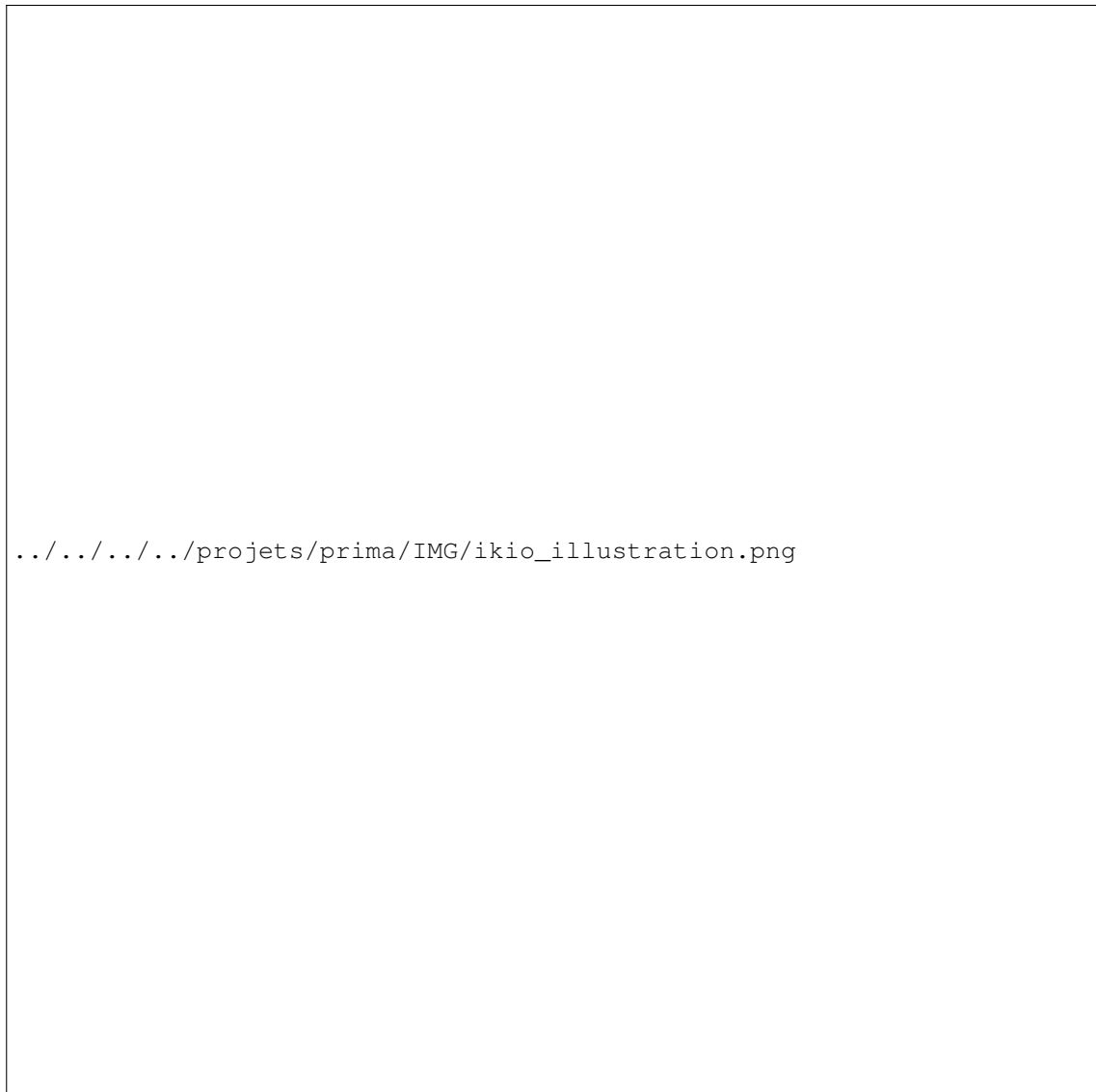
#### **5.4. Limits and performances of embedded RGBD sensors on mobile robots for social interaction**

**Participants:** Amaury Nègre, Dominique Vaufreydaz [correspondant].

While working on sociably acceptable companion robots, we highlighted some problems of embedding RGBD sensors on mobile robots. Performances of our algorithms can be severely decreased by intrinsic parameters of the robot: linear and angle speeds, height and angle of view of the mounted RGBD sensor, etc. We are currently conducting experiments on influence of these parameters on our perception of humans within a home-like environment. As an extra expected results, we will provide to the research community a corpus that can be used as benchmark for several tasks in mobility: 2D and 3D face detection, body and skeleton detection, fall detection and engagement detection.

---

<sup>3</sup>[http://www.robotis.com/x/bioloid\\_en](http://www.robotis.com/x/bioloid_en)



*Figure 6. Preliminary sketch of the iKio together with an early prototype, both the 3d model and its realization using the Bioloid Kit*



*Figure 7. Robotic platform within the home-like environment for mobile RGBD experiments.*

## 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, a new method to detect salient objects has been proposed [32], [18]. The principle relies upon low-level visual features and super-pixel segmentation. First, the original image is simplified by performing super-pixel segmentation and adaptive color quantization. On the basis of super-pixel representation, inter-super-pixel similarity measures are then calculated based on difference of histograms and spatial distance between each pair of super-pixels. For each super-pixel, its global contrast measure and spatial sparsity measure are evaluated, and refined with the integration of inter super-pixel similarity measures to finally generate the super-pixel-level saliency map. Experimental results on a dataset containing 1,000 test images with ground truths demonstrate that the proposed saliency model outperforms state-of-the-art saliency models. Figure 1 illustrates some results.

### 6.1.2. Image Memorability

**Participant:** Olivier Le Meur.

This work has been carried out in collaboration with Mattei Mancas (researcher of the University of Mons) during his visit of the team. The image memorability consists in the faculty of an image to be recalled after a period of time. Recently, the memorability of an image database was measured and some factors responsible for this memorability were highlighted. In [34] we proposed to improve an existing method by using attention-based visual features. To determine whether the visual attention plays a role in the memorability mechanism, eye tracking experiment has been performed by using a set of images of different memorability scores. Two important results have been observed. First the fixation duration is longer for the most memorable images (especially for the very first fixations) which shows a higher cognitive activity for memorable images. Second the observers congruency (agreement between observers) is significantly higher for the most memorable images. This shows that when there are areas with high attraction on all viewers, this induces higher memorability.

Following these first two observations, attention-based visual features were used to predict image memorability scores. A new set of features was then defined and used to train a model. Compared to an existing approach, we improve on the quality of the prediction of 2% while reducing the number of parameters by 14%. More specifically we replace the 512 features related to the GIST by 17 features which are directly related to visual attention.

### 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. The goal is also to study the differences in 2D visual attention in comparison with 3D visual attention.



*Figure 1. Illustration of the proposed approach: first row: original image; second row: saliency map; third row: extraction of the salient object.*

Several subjective experiments have been carried out in order to study visual attention in different viewing conditions. The goal of the first experiment, involving 135 observers, was to study visual attention in three different conditions (2D, 3D comfortable and 3D uncomfortable), to eventually establish whether depth influences visual attention and whether there is a link between comfort and visual attention. The use of an eye-tracker allowed to record and to track observer's gaze. By analyzing the results, we found out that visual strategy to observe 2D images and 3D images with uncrossed disparity is very similar; there was no significant influence of discomfort on visual attention.

The second question which has then been addressed is how visual attention is influenced by objects with crossed disparity. A second test has been designed to answer this question, involving 51 observers. Considering scenes with crossed disparity it was revealed that objects located in front of the display plane are the most salient, even if observers experience discomfort. In the third experiment, we extended the study using scenes with crossed and uncrossed disparities. We verified the hypothesis that texture and contrast are more influential in guiding our gaze than the amount of depth. The features influencing the saliency of the objects in stereoscopic conditions were also evaluated with low-level visual stimuli. It was discovered that texture is the most salient feature in comparison to depth. Crossed disparity significantly influences the process of selecting the objects, while uncrossed disparity is less important, the process of selection being in this latter case similar to 2D conditions.

#### **6.1.4. Epitome-based video representation**

**Participants:** Martin Alain, Christine Guillemot.

This work is carried out in collaboration with Technicolor (D. Thoreau, Ph. Guillotel) and aims at studying novel spatio-temporal representations for videos based on epitomes. An epitome is a condensed representation of an image (or a video) signal containing the essence of the textural properties of this image. Different forms of epitomes have been proposed in the literature, such as a patch-based probability model learned either from still image patches or from space-time texture cubes taken from the input video. These probability models together with appropriate inference algorithms, are useful for content analysis inpainting or super-resolution. Another family of approaches makes use of computer vision techniques, like the KLT tracking algorithm, in order to recover self similarities within and across images. In parallel, another type of approach consists in extracting epitome-like signatures from images using sparse coding and dictionary learning.

We have in the past (in the context of the PhD thesis of S. Cherigui) developed a method for constructing epitomes for representing still images. The algorithm tracks self-similarities within the image using a block matching (BM) algorithm. The epitome is constructed from disjoint pieces of texture ("epitome charts") taken from the original image and a transform map which contains translational parameters (see Fig.2). Those parameters keep track of the correspondences between each block of the input image and a block of the epitome. An Intra image compression scheme based on the epitome has been developed showing significant rate savings on some images, including the rate cost of the epitome texture and of the transform map. The entire image can be reconstructed from the epitome texture with the help of the transform map. The method is currently being extended to construct epitome representations of video segments rather than simple images. Such spatio-temporal epitome should pave the way for novel video coding architectures and open perspectives for other video processing problems which we have started to address such as denoising and super-resolution.

## **6.2. Rendering, inpainting and super-resolution**

image-based rendering, inpainting, view synthesis, super-resolution

### **6.2.1. Image and video inpainting**

**Participants:** Mounira Ebdelli, Christine Guillemot, Olivier Le Meur.





Figure 2. Original image and corresponding epitome.

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.

Novel methods have been developed investigating two complementary directions first for image inpainting. The first direction which has been explored is the estimation of the unknown pixel with different neighbor embedding methods, i.e. Locally Linear embedding (LLE), LLE with a low-dimensional neighborhood representation (LLE-LDNR), Non-Negative Matrix Factorization (NMF) with various solvers [16]. The second method developed uses a two-steps hierarchical (or coarse to fine) approach to reduce the execution time [17]. In this hierarchical approach, a low resolution version of the input image is first inpainted, this first step being followed by a second one which recovers the high frequency details of the inpainted regions, using a single-image super-resolution method. To be less sensitive to the parameters setting of the inpainting, the low-resolution input picture is inpainted several times with different settings. Results are then efficiently combined with a loopy belief propagation. Experimental results in a context of image editing, texture synthesis and 3D view synthesis demonstrate the effectiveness of the proposed method.

The problem of video inpainting has also been considered. A first video inpainting algorithm has been developed in 2012, using a spatio-temporal exemplar-based method. The algorithm proceeds in three steps. The first one inpaints missing pixels in moving objects using motion information. Then the static background is inpainted exploiting similarity between neighboring frames. The last step fills in the remaining holes in the current frame using spatial inpainting. This approach works well with static cameras but not so well when the video has been captured by free-moving cameras.

In 2013, we have therefore addressed the problem of video inpainting 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. The proposed approach, although less complex than state-of-the-art methods, provides more natural results (see Fig.3).



Figure 3. Mask of the image to be inpainted; Results with the proposed video inpainting algorithm.

### 6.2.2. Image priors for inpainting

**Participants:** Raul Martinez Noriega, Aline Roumy.

Image inpainting is an ill-posed inverse problem which has no well-defined unique solution. To make this problem more "well-defined" it is necessary to introduce image priors. We consider here the problem of extracting such priors to help restoring the connection of long edges across the missing region. The prior is defined as a binary image that contains the locations of salient edge points located at the boundary of the missing region as well as the linear edges that join these points across the missing region. A method has been developed to extract such priors. It first detect edges which are then successively pruned in order to keep only informative edges, i.e., which have coherent gradients and are either part of a salient structure, or at the border between two different textures. Edges which are quasi-perpendicular to the boundary of the missing region are finally retained. Directions of the retained edges are computed and pairs of edges with similar directions are then connected with straight lines. These lines are used to segment the image into different regions and to define the processing order of the patches to be inpainted. Only patches from the known part and belonging

to the same region as the input patch are used. This avoids bringing details of one texture into another one, as well as the unconnected edge problem [35].

### 6.2.3. Image and video super-resolution

**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 (see, for example, Fig. 4 for a comparison between bicubic interpolation and SR). We mostly focused on the single-image problem, where only a single LR image is available.

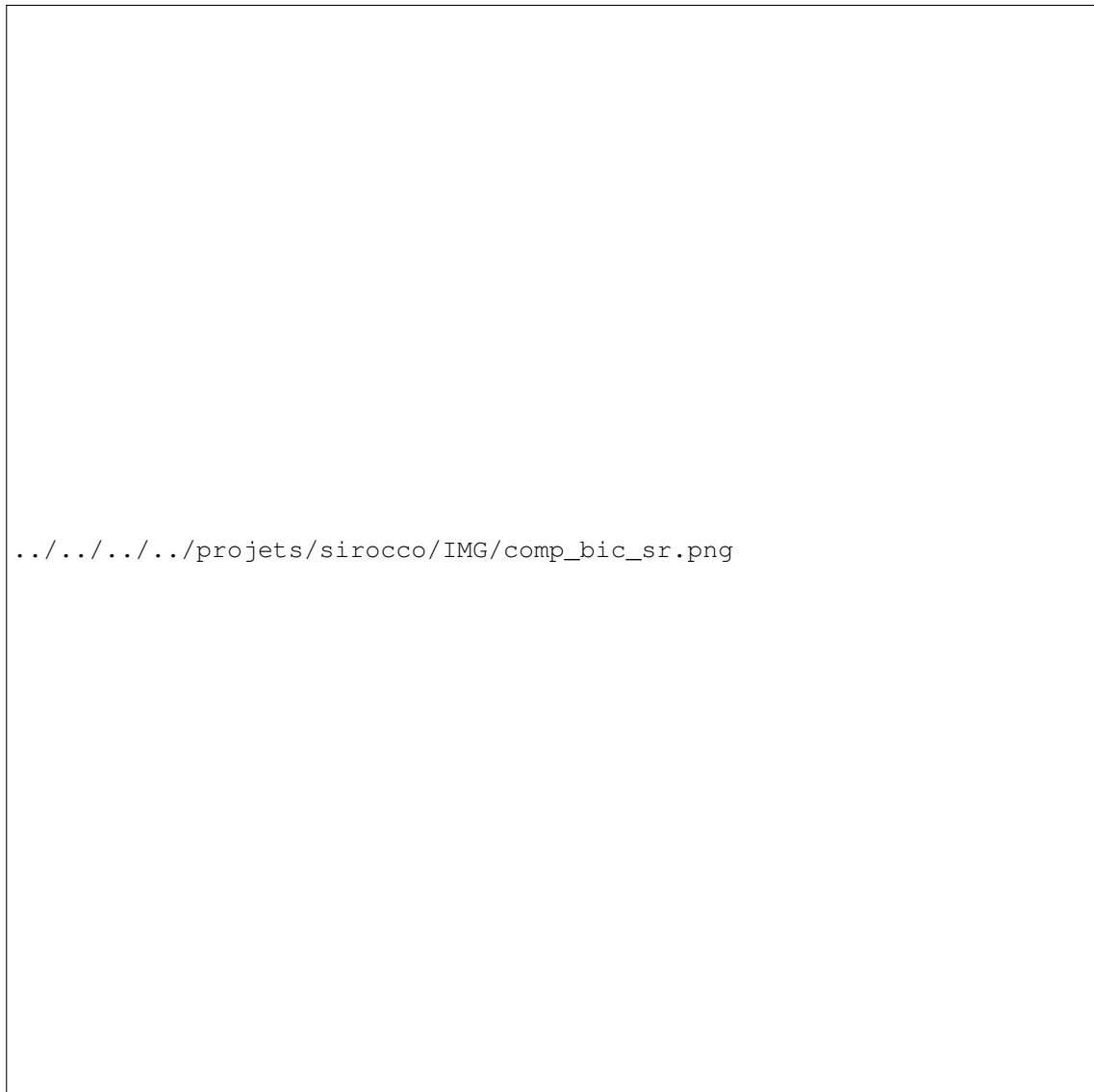
We have adopted the example-based framework, where the relation between the LR and HR image spaces is modeled with the help of pairs of small “examples”, i.e. texture patches. Each example pair consists of a LR patch and its HR version that also includes high-frequency details; the pairs of patches form 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 super-resolved 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 learnt 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.

To overcome this problem, in [23] we proposed a novel method to build a compact external dictionary. The method consists in first jointly clustering LR and HR patches. The aim of this procedure, which we called JKC (Jointly K-means Clustering), is to prune the dictionary of the “bad” pairs of patches, i.e. those ones for which the cluster assignments of the related LR and HR patches do not correspond. Once the dictionary is clustered, it is summarized, by sampling some prototype patches, and applying on them simple geometrical transformations, in order to enrich the dictionary. The so constructed compact dictionary is shown to give equivalent or even better performance than the initial large dictionary with any input image.

The dictionary construction method described in [23] has been used as a basis for designing a full single-image SR algorithm. The new algorithm, presented in [25], follows the traditional scheme of example-based SR with an external dictionary, where a new way to generate the training patches is introduced. Given a HR training image  $H$ , the corresponding LR image  $L$  is generated; but instead of directly sampling patches from  $H$  and  $L$ , as usually done, the training images are further processed. An enhanced interpolation of  $L$ , using an iterated back projection, is used as a source of LR patches, and a high-frequency residual image, given by the difference between  $H$  and the interpolated LR image, is used for extracting HR patches. The JKC procedure is then applied to get the final compact dictionary. A special example-based SR algorithm has been designed, where the final HR output patches are constructed by combining selected HR residual patches from the dictionary with nonnegative weights. In the context of this study, we have also introduced a novel non-negative dictionary learning method [24]. The proposed method consists of two steps which are alternatively iterated: a sparse coding and a dictionary update stage. As for the dictionary update, an original method has been proposed, which we called K-WEB, as it involves the computation of  $k$  WEighted Barycenters.

Besides SR for still images, a preliminary work on video sequences has been also conducted [26]. In particular, we have considered the case of a LR video sequence with periodic high-resolution (HR) key frames. Given this scenario, a specific SR procedure has been designed to upscale each intermediate frame, by using the internal dictionary constructed from the two neighbor key frames.



*Figure 4. Comparison between bicubic interpolation and SR when upscaling the same LR image (by a factor of 3).*

### 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. *Multi-view plus depth video compression*

**Participants:** Christine Guillemot, Laurent Guillo.

Multi-view plus depth video content represent very large volumes of input data which need to be compressed for storage and transmission to the rendering device. The huge amount of data contained in multi-view sequences indeed motivates the design of efficient representation and compression algorithms. The team has worked on motion vector prediction in the context of HEVC-compatible Multi-view plus depth (MVD) video compression. The HEVC compatible MVD compression solution implements a 6 candidate vector list for merge and skip modes. When a merge or a skip mode is selected, a merge index is written in the bitstream. This index is first binarized using a unary code, then encoded with the CABAC. A CABAC context is dedicated to the first bin of the unary coded index while the remaining bins are considered as equiprobable. This strategy is efficient as long as the candidate list is ordered by decreasing index occurrence probability. We have improved the construction of the candidate list by proposing two new candidates derived from disparity motion vectors in order to exploit inter-view correlation. This work has led to a joint proposal with Qualcomm and Mediatek which has been adopted in the HEVC-3DV standard in July 2013.

#### 6.3.2. *Spatio-temporal video prediction with neighbor embedding*

**Participants:** Martin Alain, Christine Guillemot.

The problem of texture prediction can be regarded as a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the block to be predicted. We have in 2012 developed texture prediction methods as well as inpainting algorithms using sparse representation as with learned dictionaries [19], or using neighbor embedding techniques [11], [30]. The methods which we have more particularly considered are Locally Linear Embedding (LLE), LLE with Low-dimensional neighborhood representation (LDNR), and Non-negative Matrix Factorization (NMF) using various solvers. In 2013, we have addressed the problem of temporal prediction for inter frame coding of video sequences using locally linear embedding (LLE). LLE-based prediction computes the predictor as a linear combination of  $K$  nearest neighbors (K-NN) searched within one or several reference frames. We have explored different  $K$ -NN search strategies in the context of temporal prediction, leading to several temporal predictor variants using or not motion information [22]. A parallel was also drawn between such multi-patch based prediction and the adaptive interpolation filtering (AIF) method. The LLE-based inter prediction techniques, when used as extra modes for inter prediction in an H.264 codec, are shown to bring significant Rate-Distortion (RD) performance gains compared to H.264 (up to 21.76 % bit-rate saving) and with respect to the use of AIF.

#### 6.3.3. *Dictionary learning for sparse coding 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, therefore the rate without compression is very high (about 70 Gbits/sec). The goal is to achieve a rate after compression of 600 Mbits/sec, i.e., 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.

Sparse representation of a signal consists in representing a signal  $y \in \mathfrak{R}^n$  as a linear combination of columns, known as atoms, from a dictionary matrix. The dictionary  $D \in \mathfrak{R}^{n \times K}$  is generally overcomplete and contains  $K$  atoms. The approximation of the signal can thus be written  $y \approx Dx$  and is sparse because a small number of atoms of  $D$  are used in the representation, meaning that the vector  $x$  has only a few non-zero coefficients. Sparsity of the representation depends on how the dictionary is representative of the data at hand, hence the need to learn appropriate dictionaries.

We have developed methods for learning adaptive tree-structured dictionaries, called Tree K-SVD [20]. 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. We recently developed a new sparse coding method adapted to this tree-structure to improve the results [20]. The tree-structured dictionary has been further improved by studying different branch pruning strategies. The use of these dictionaries in an HEVC-based intra coder is under study. The dictionaries are also considered for scene classification and for detecting the MTF (Modulation Transfer Function) of the optical capturing system.

#### 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 are generally used to represent the pixel data. Floating point video formats are widely used in the visual effects industry. Moreover, the development of a new standardized workflow ACES intends to generalize the use of such formats to the whole cinema production pipeline. The increasing use of floating point representations, however, comes with a technical issue concerning the storage space required for those videos with higher precision than the current 8 or 10 bit standards.

In collaboration with Technicolor (D. Thoreau), we worked on floating point video compression. Different approaches exist in the literature. Several methods consist in compressing directly the floating point data using its internal representation (i.e. sign, exponent and mantissa bits). These methods are generally limited to lossless compression schemes. Another type of approach makes use of the existing compression standards such as H264/AVC or HEVC to encode a floating point sequence of images previously converted to lower bit depth integers. In this approach, the conversion is designed to be reversible with minimal loss. However the converted integer images are not intended for being displayed directly. Finally a last family of approach aims at keeping backward compatibility with an existing compression standard. The original image sequence is first tone mapped and encoded to obtain a low dynamic range (LDR) version that can be visualized on a standard LDR display. In parallel, a residual information needed to reconstruct the HDR image from the LDR version is also encoded.

In our study, a floating point to integer conversion method was developed to be applied before HEVC compression. The original floating point RGB values are converted to high bit depth integers with an approximate logarithmic encoding that is reversible without loss. The RGB values are then converted to a YUV color space. The bit depth must also be reduced to be supported by the compression standard. This bit depth reduction is performed adaptively depending on the minimum and maximum values (i.e. darkest and brightest points respectively) which characterize the real dynamic of the data. In the best case, the difference between the extreme values is sufficiently low to perform this operation without loss.

Three variants of the method have been compared. The conversion can be performed either by Groups of Pictures (GOP), or independently on each frame of the sequence, or even more locally, by blocks of pixels. The GOP-wise approach combined with spatial and temporal predictions in the encoder gives the best results for low bit rate compression. The block-wise approach can reduce the bit depth with less data loss but breaks the continuity between the blocks, which degrades the Rate Distortion (RD) performance especially at low bit rates. However, we have shown that this approach gives the best results in the context of near lossless compression. The frame-wise version is intermediate between the global (GOP-wise) and local (block-wise) versions. It is adapted to high quality compression. This method was also compared to another frame-wise

conversion method in the recent literature called adaptive LogLuv transform, and a 50% rate saving was obtained at high bitrates.

### 6.3.5. HEVC coding optimization

**Participants:** Nicolas Dholand, Christine Guillemot, Bihong Huang, Olivier Le Meur.

The team has two collaborations in the area of HEVC-based video coding optimization. The first research activity is carried out in collaboration with Orange labs (Felix Henry) and UPC (Philippe Salembier) in Barcelona. The objective is to design novel methods for predicting the residues resulting from spatio-temporal prediction. We have indeed observed that the redundancy in residual signals (hence the potential rate saving) is high. In 2013, different methods have been investigated to remove this redundancy, such as generalized lifting and different types of predictors. The generalized lifting is an extension of the lifting scheme of classical wavelet transforms which permits the creation of nonlinear and signal probability density function (pdf) dependent and adaptive transforms.

The second collaboration is with Thomson Video Networks and aims at designing an innovative architecture for effective real-time broadcast encoders of Ultra High Definition (UHD) contents. Currently, the only way to transmit acceptable UHD contents around 10 – 20 Mbits/sec is the new compression standard HEVC (finalized in January 2013). Yet, UHD requires at minimum 8 times more computation than the actual HDTV formats, and HEVC has a computing complexity which is already from 2 to 10 times that of MPEG4-AVC. To reduce the encoding complexity on UHD content, a pre-analysis with a lower resolution version (HD) of the input content has been considered to infer some decisions and coding parameters on the UHD video. A speed-up of a factor 3 has already been achieved for a small rate loss of 4 – 5%.

## 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. Loss concealment based on video inpainting

**Participants:** Mounira Ebdelli, Christine Guillemot, Ronan Le Boulch, Olivier Le Meur.

We have developed a loss concealment scheme based on a new hierarchical video exemplar-based inpainting algorithm. The problem of loss concealment is to estimate unknown pixels after decoding when the corresponding transport packets have been lost on the transmission network. Before proceeding to the video texture inpainting, the motion vectors of the lost blocks must first be estimated from the motion vectors of the received blocks in the spatial neighborhood. The Motion vectors (MV) of damaged blocks are estimated using a Bilinear Motion Field Interpolation (BMFI) technique.

The algorithm follows a coarse to fine approach and first inpaints a low resolution version of the damaged video. Moving objects, detected thanks to the estimated motion vectors, are processed first. The most similar patches (similar to the known pixels of the patch to be completed) is searched within a motion-compensated window in adjacent frames, and used as an estimate of the pixels to be filled in. Then the static background is inpainted using known co-located pixels of neighboring frames. The remaining holes are filled-in using spatial inpainting.

In a second step, the high frequency details of the inpainted areas are recovered using a super-resolution technique, in the same vein as described in Section 6.2.1 for still images. The inpainted low resolution video is first interpolated using a simple lanczos interpolation. The idea is then to search for the nearest neighbor (the best match) of the interpolated version of each inpainted block, within the known part of the current image of the impaired video at the native resolution. The found correspondences form a so-called nearest neighbor field (NNF) which connects inpainted and interpolated patches of the low resolution video to high resolution patches of known parts of the high resolution (HR) video. The found NN patch is then copied to replace the low resolution inpainted patch. The two-step approach allows significantly reducing the execution time of the video inpainting process, while preserving a satisfactory quality.



#### 6.4.2. Universal distributed 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 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 correlation between the successive frames is not directly used at the encoder, and are therefore seen as distributed. Traditional approaches (from an information theoretical but also practical point of view) assume that the correlation channel between 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 correlation channel is unknown and belongs to a class parametrized by some unknown parameter vector. We proposed four uncertainty models that depend on the partial knowledge we have on the correlation channel and derived the information theoretical bounds [28]. A complete coding scheme has also been proposed that works well for any distribution in the class [27]. At the encoder, the proposed scheme encompasses the determination of the coding rate and the design of the encoding process. Both contributions result from the information-theoretical compression bounds of universal lossless source coding with side information. Then a novel decoder is proposed that takes into account the available information regarding the class. The proposed scheme avoids the use of a feedback channel or the transmission of a learning sequence, which both would result in a rate increase at finite length.



## STARS Project-Team

# 6. New Results

## 6.1. 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.

### 6.1.1. Perception for Activity Recognition

**Participants:** Julien Badie, Slawomir Bak, Vasanth Bathrinarayanan, Piotr Bilinski, François Brémond, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Carolina Garate, Vaibhav Katiyar, Ratnesh Kumar, Srinidhi Mukanahallipatna, Marco San Biago, Silviu Serban, Malik Souled, Kartick Subramanian, Anh Tuan Nghiem, Monique Thonnat, Sofia Zaidenberg.

This year Stars has extended an algorithm for tuning automatically the parameters of the people tracking algorithm. We have evaluated the algorithm for re-identification of people through a camera network while taking into account a large variety of potential features together with practical constraints. We have designed several original algorithms for the recognition of short actions and validated its performance on several benchmarking databases (e.g. ADL). We have also worked on video segmentation and representation, with different approaches and applications.

More precisely, the new results for perception for activity recognition concern:

- Background Subtraction and People Detection in Videos (6.2)
- Tracking and Video Representation (6.3)
- Video segmentation with shape constraint (6.4)
- Articulating motion (6.5)
- Lossless image compression (6.6)
- People detection using RGB-D cameras (6.7)
- Online Tracking Parameter Adaptation based on Evaluation (6.8)
- People Detection, Tracking and Re-identification Through a Video Camera Network (6.9)
- People Retrieval in a Network of Cameras (6.10)
- Global Tracker : an Online Evaluation Framework to Improve Tracking Quality (6.11)
- Human Action Recognition in Videos (6.12)
- 3D Trajectories for Action Recognition Using Depth Sensors (6.13)
- Unsupervised Sudden Group Movement Discovery for Video Surveillance (6.14)
- Group Behavior Understanding (6.15)

### 6.1.2. Semantic Activity Recognition

**Participants:** Guillaume Charpiat, Serhan Cosar, Carlos -Fernando Crispim Junior, Hervé Falciani, Baptiste Fosty, Qiao Ma, Rim Romdhane.

During this period, we have thoroughly evaluated the generic event recognition algorithm using both sensors (RGB and RGBD video cameras). This algorithm has been tested on more than 70 videos of older adults performing 15 min of physical exercises and cognitive tasks. In Paris subway, we have been able to demonstrate the recognition in live of group behaviours. We have also been able to store the meta-data (e.g. people trajectories) generated from the processing of 8 video cameras, each of them lasting 2 or 3 days. From these meta-data, we have automatically discovered few hundreds of rare events, such as loitering, collapsing, ... to display on the screen of subway security operators.

Concerning semantic activity recognition, the contributions are :

- Evaluation of an Activity Monitoring System for Older People Using Fixed Cameras (6.16 )
- A Framework for Activity Detection of Older People Using Multiple Sensors (6.17 )
- Walking Speed Detection on a Treadmill using an RGB-D Camera (6.18 )
- Serious Game for older adults with dementia (6.19 )
- Unsupervised Activity Learning and Recognition (6.20 )
- Extracting Statistical Information from Videos with Data Mining (6.21 )

### 6.1.3. Software Engineering for Activity Recognition

**Participants:** François Brémond, Daniel Gaffé, Julien Gueytat, Sabine Moisan, Anh Tuan Nghiem, Annie Ressouche, Jean-Paul Rigault, Luis-Emiliano Sanchez.

This year Stars has continued the development of the SUP platform. This latter is the backbone of the team experiments to implement the new algorithms. We continue to improve our meta-modelling approach to support the development of video surveillance applications based on SUP. This year we have focused on metrics to drive dynamic architecture changes and on component management. We continue the development of a scenario analysis module (SAM) relying on formal methods to support activity recognition in SUP platform. We improve the CLEM toolkit and we rely on it to build SAM. Finally, we are improving the way we perform adaptation in the definition of a multiple services for device adaptive platform for scenario recognition.

The contributions for this research axis are:

- SUP (6.22 )
- Model-Driven Engineering for Activity Recognition (6.23 )
- Scenario Analysis Module (6.24 )
- The Clem Workflow (6.25 )
- Multiple Services for Device Adaptive Platform for Scenario Recognition (6.26 )

## 6.2. Background Subtraction and People Detection in Videos

**Participants:** Vasanth Bathrinathan, Srinidhi Mukanahallipatna, Silviu Serban, François Brémond.

**Keywords:** Background Subtraction, People detection, Automatic parameter selection for algorithm

**Background Subtraction** Background subtraction is a vital real time low-level algorithm, which differentiates foreground and background objects in a video. We have thoroughly evaluated our Extended Gaussian Mixture model containing a shadows-removal algorithm, which performs better than other state of the art methods. Figure 10 shows the comparison of 13 background subtraction algorithms results on a challenging railway station monitoring video dataset from Project CENTAUR, which includes illumination change, shadows, occlusion and moving trains. Our algorithms performs the best in terms of result and with good processing speed too. Figure 11 is an example of our background subtraction algorithm's output on an indoor sequence of a surveillance footage from the Project SUPPORT.

Ongoing research include automatic parameter selection for this algorithm based on some learnt context. Since tuning the parameters is a daunting task for a non-experienced person, we try to learn some context information in a video like occlusion, contrast variation, density of foreground, texture etc. and map them to appropriate parameters of segmentation algorithm. Thus designing a controller to automatically adapt parameters of a algorithm as the scene context changes over time.

### People Detection

A new robust real-time person detection system was proposed [45], which aims to serve as solid foundation for developing solutions at an elevated level of reliability. Our belief is that clever handling of input data correlated with efficacious training algorithms are key for obtaining top performance. A comprehensive training method on very large training database and based on random sampling that compiles optimal classifiers with minimal bias and overfit rate is used. Building upon recent advances in multi-scale feature computations, our approach attains state-of-the-art accuracy while running at high frame rate.



Figure 10. Background Subtraction result on confidential dataset from CENTAUR project



Figure 11. Background Subtraction result on a video to count number of people walking through the door after using their badge inside the terminal area (Project SUPPORT) - Autonomous Monitoring for Securing European Ports

Our method combines detection techniques that greatly reduce computational time without compromising accuracy. We use efficient LBP and MCT features which we compute on integral images for optimal retrieval of rectangular region intensity and nominal scaling error. AdaBoost is used to create cascading classifiers with significantly reduced detection time. We further refine detection speed by using the soft cascades approach and by transferring all important computation from the detection stage to the training stage. Figure 12 shows some output samples from various datasets which it was tested on.



Figure 12. Detection output from PETS (left-top), VANAHEIM (right-top), Hospital (left-bottom) and ETISEO (right-bottom)

### 6.3. Tracking and Video Representation

**Participants:** Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

**keywords:** Fibers, Graph Partitioning, Message Passing, Iterative Conditional Modes, Video Segmentation, Video Inpainting

**Multiple Object Tracking** The objective is to find trajectories of objects (belonging to a particular category) in a video. To find possible occupancy locations, an object detector is applied to all frames of a video, yielding bounding boxes. Detectors are not perfect and may provide false detections; they may also miss objects sometimes. We build a graph of all detections, and aim at partitioning the graph into object trajectories. Edges in the graph encode factors between detections, based on the following :

- Number of common point tracks between bounding boxes (the tracks are obtained from an optical-flow-based point tracker)
- Global appearance similarity (based on the pixel colors inside the bounding boxes)
- Trajectory straightness : for three bounding boxes at different frames, we compute the Laplacian (centered at the middle frame) of the centroids of the boxes.
- Repulsive constraint : Two detections in a same frame cannot belong to the same trajectory.

We compute the partitions by using sequential tree re-weighted message passing (TRW-S). To avoid local minima, we use a label flipper motivated from the Iterative Conditional Modes algorithm.

We apply our approach to typical surveillance videos where object of interest are humans. Comparative quantitative results can be seen in Tables 1 and 2 for two videos. The evaluation metrics considered are : Recall, Precision, Average False Alarms Per Frame (FAF), Number of Groundtruth Trajectories (GT), Number of Mostly Tracked Trajectories, Number of Fragments (Frag), Number of Identity Switches (IDS), Multiple Object Tracking Accuracy (MOTA) and Multiple Object Tracking Precision (MOTP).

This work has been submitted to CVPR' 14.

Table 1. Towncenter Video Output

Method	MOTA	MOTP	Detector
[59] (450-750)	56.8	79.6	HOG
Ours (450-750)	53.5	69.1	HOG

Table 2. Comparison with recent proposed approaches on PETS S2L1 Video

Method	Recall	Precision	FAF	GT	MT	Frag	IDS
[77]	96.9	94.1	0.36	19	18	15	22
Ours	95.4	93.4	<b>0.28</b>	19	18	42	<b>13</b>

**Video Representation** We continued our work from the previous year on Fiber-Based Video Representation. During this year we focused on obtaining competitive results with the state-of-the-art (Figure 13).

The usefulness of our novel representation is demonstrated by a simple video inpainting task. Here a user input of only 7 clicks is required to remove the dancing girl disturbing the news reporter (Figure 14).

This work has been accepted for publication next year [41].

### 6.4. Video segmentation with shape constraint

**Participant:** Guillaume Charpiat.

**keywords:** video segmentation, graph-cut, shape growth, shape statistics, shape prior, dynamic time warping

#### 6.4.1. Video segmentation with growth constraint

This is joint work with Yuliya Tarabalka (Ayin Inria team) and Björn Menze (ETH Zurich, also MIT and collaborator of Asclepios Inria team).



Figure 13. **Top Row**: Left image displays a sequence as a volumetric display. Right image displays all fibers found, clustered at a particular hierarchy. **Bottom Row** : Left Image displays the highest level of the hierarchical clustering, with fiber extension. Right Image shows the result obtained from [71]. **Our result demonstrates better long term temporal coherency.**



Figure 14. Inpainting task. **Left** : Original video (top) and  $xt$  slice (bottom) showing trajectories. **Right** : Our result. Clusters of fibers were computed and selected with only 7 mouse clicks to distinguish the disturbing girl from the reporter and background. The girl was removed and the hole was filled by extending the background fibers in time.

**Context** : One of the important challenges in computer vision is the automatic segmentation of objects in videos. This task becomes more difficult when image sequences are subject to low signal-to-noise ratio or low contrast between intensities of neighboring structures in the image scene. Such challenging data is acquired routinely, for example in medical imaging or in satellite remote sensing. While individual frames could be analyzed independently, temporal coherence in image sequences provides crucial information to make the problem easier. In this work, we focus on segmenting shapes in image sequences which only grow or shrink in time, and on making use of this knowledge as a constraint to help the segmentation process.

**Approach and applications** : We had proposed last year an approach based on graph-cut (see Figure 15 ), able to obtain efficiently (linear time in the number of pixels in practice), for any given video, its globally-optimal segmentation satisfying the growth constraint. This year we applied this method to three different applications :

- forest fires in satellite images,
- organ development in medical imaging (brain tumor, in multimodal MRI 3D volumes),
- sea ice melting in satellite observation, with a shrinking constraint instead of growth (see Figure 16 ).

The results on the first application were published in IGARSS (International Geoscience and Remote Sensing Symposium) [48], while the last two applications and the theory were published in BMCV [47]. A journal paper is also currently under review. A science popularization article was also published [53]. Not related but also with the Ayin Inria team was published the last of a series of articles about optimizers for point process models [40], introducing graph-cuts in the multiple birth and death approach in order to detect numerous objects that should not overlap.

#### 6.4.2. Video segmentation with statistical shape prior

This is joint work with Maximiliano Suster (leader of the Neural Circuits and Behaviour Group at Bergen University, Norway).



*Figure 15. The approach : segmenting all frames together with a single graph-cut, with growth or shrinkage constraint, instead of segmenting independently each frame.*





*Figure 16. Example of a noisy, challenging sequence of melting sea ice, from satellite observations (first column). The frames were aligned beforehand, using inside/outside histograms of pixel intensities. The following columns show the results of varied approaches, ranging from frame-by-frame segmentation to techniques ensuring shape smoothness in time. Our approach (last column) has the right suitable prior (shape shrinkage) and thus performs the best.*

**Context :** The zebrafish larva is a model organism widely used in biology to study genetics. Therefore, analyzing its behavior in video sequences is particularly important for this research field. For this, there is a need to segment the animal in the video, in order to estimate its speed, and also more precisely to extract its shape, in order to express for instance how much it is bent, how fast it bends, etc. However, as the animal is stimulated by the experimenter with a probe, the full zebrafish larva is not always visible because of occlusion.

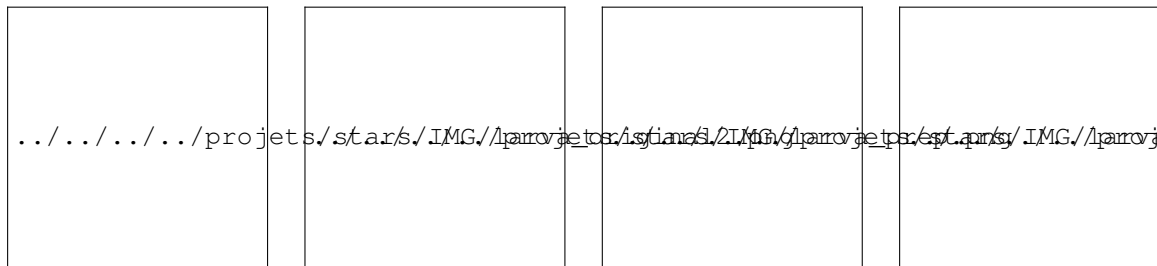


Figure 17. Example of a segmentation : initial image, processed image (based on video coherency), initialization of the active contour evolution, result. The total time spent per frame on average is reasonable for practical applications (magnitude order of 1 second).

**Approach :** We build a shape prior based on a training set of examples of non-occluded shapes, and use it to segment new images where the animal is occluded. This is however not straightforward.

- **Building a training set of shape deformations :** Given a set of training images containing non-occluded animals, we extract their contours via multiple robust thresholdings and morphomathematical operations. For each contour, we then estimate automatically the location of the tip of the tail. We then compute point-to-point correspondences between all contours, using a modified version of Dynamic Time Warping, as well as the approximate tip location information. This is done in a translation- and rotation-invariant way.
- **Building the shape prior :** Based on these matchings, the mean shape is computed, as well as modes of deformation with PCA.
- **Segmenting occluded images :** Images with occluded shapes are pre-processed in a similar way to non-occluded ones; however, the resulted segmentation does not contain only the parts of the larva but also the probe, which has potentially similar colors and location, and is moving. To identify the probe, whose shape depends on the video sequence, we make use of its rigidity and of temporal coherency. Then a segmentation criterion is designed to push an active contour towards the zones of interest (in a way that is robust to initialization), while keeping a shape which is feasible according to the shape prior.

Examples of data and results for a preliminary algorithm are shown in Figure 17 , with the associated shape prior shown in Figure 18 .

## 6.5. Articulating motion

**Participant:** Guillaume Charpiat.

**keywords:** shape evolution, metrics, gradient descent, Finsler gradient, Banach space, piecewise-rigidity, piecewise-similarity

This is joint work with Giacomo Nardi, Gabriel Peyré and François-Xavier Vialard (Ceremade, Paris-Dauphine University).



*Figure 18. First deformation modes of the shape prior used in the segmentation above.*

**Context in optimization :** A fact which is often ignored when optimizing a criterion with a gradient descent is that the gradient of a quantity depends on the metric chosen. In many domains, people choose by default the underlying  $L^2$  metric, while it is not always relevant. Here we extend the set of metrics that can be considered, by building gradients for metrics that do not derive from inner products, with examples of metrics involving the  $L^1$  norm, possibly of a derivative.

**Mathematical foundations :** This work introduces a novel steepest descent flow in Banach spaces. This extends previous works on generalized gradient descent, notably the work of Charpiat et al. [6], to the setting of Finsler metrics. Such a generalized gradient allows one to take into account a prior on deformations (e.g., piecewise rigid) in order to favor some specific evolutions. We define a Finsler gradient descent method to minimize a functional defined on a Banach space and we prove a convergence theorem for such a method. In particular, we show that the use of non-Hilbertian norms on Banach spaces is useful to study non-convex optimization problems where the geometry of the space might play a crucial role to avoid poor local minima.

**Application to shape evolution :** We performed some applications to the curve matching problem. In particular, we characterized piecewise-rigid deformations on the space of curves and we studied several models to perform piecewise-rigid evolutions (see Figure 19 ). We also studied piecewise-similar evolutions. Piecewise-rigidity intuitively corresponds to articulated motions, while piecewise-similarity further allows the elastic stretching of each articulated part independently. One practical consequence of our work is that any deformation to be applied to a shape can be easily and optimally transformed into an articulated deformation with few articulations, the number and location of the articulations being not known in advance. Surprisingly, this problem is actually convex.



Figure 19. Example of use of the Finsler gradient for the piecewise-rigid evolution of curves. Given an initial shape  $S$  and a target shape  $T$ , as well as a shape dissimilarity measure  $E(S) = Dissim(S, T)$ , any classical gradient descent on  $E(S)$  would draw the evolving shape  $S$  towards the target  $T$ . However the metric considered to compute the gradient changes the path followed. The top row is the evolution obtained with a Sobolev gradient  $H^1$ , which has the property of smoothing spatially the flow along the curve, to avoid irregular deformations. This is however not sufficient. The bottom row makes use of the Finsler gradient instead, with a metric favoring piecewise-rigid deformations.

An article was submitted to the journal Interfaces and Free Boundaries [52].

## 6.6. Lossless image compression

**Participant:** Guillaume Charpiat.

**keywords:** image compression, entropy coding, graph-cut

This is joint work with Yann Ollivier and Jamal Atif from the TAO Inria team.

**Context :** Understanding, modelling, predicting and compressing images are tightly linked, in that any good predictor can be turned into a good compressor via entropy coding (such as Huffman coding or arithmetic coding). Indeed, with such techniques, the more predictable an event  $E$  is, i.e. the higher its probability  $p(E)$ , the easier to compress it will be, with coding cost  $-\log(p(E))$ . Therefore we are interested in image compression, in order to build better models of images.

**MDL approach :** The state-of-the-art sequential prediction of time series based on the advice of various experts combines the different expert predictions, with weights depending on their individual past performance (cf. Gilles Stoltz and Peter Grünwald's work). This approach originates from the Minimum Description Length principle (MDL). This work was however designed for 1D data such as time series, and is not directly applicable to 2D data such as images. Consequently, our aim has been to adapt such an approach to the case of image compression, where time series are replaced with 2D series of pixel colors, and where experts are predictors of the color of a pixel given the colors of neighbors.

**New method and results :** This year, we have focused on lossless greyscale image compression, and proposed to encode any image with two maps, one storing the choice of the expert made for each pixel, and one storing the encoding of the intensity of each pixel according to its expert. In order to compress efficiently the first map, we ask the choices of experts to be coherent in space, and then encode the boundaries of the experts' areas. To find a suitable expert map, we optimize the total encoding cost explicitly, set as an energy minimization problem, solved with graph-cuts. An example of expert map obtained is shown in Figure 20. Preliminary results with a hierarchical ordering scheme already compete with standard techniques in lossless compression (PNG, lossless JPEG2000, JPEG-LS).

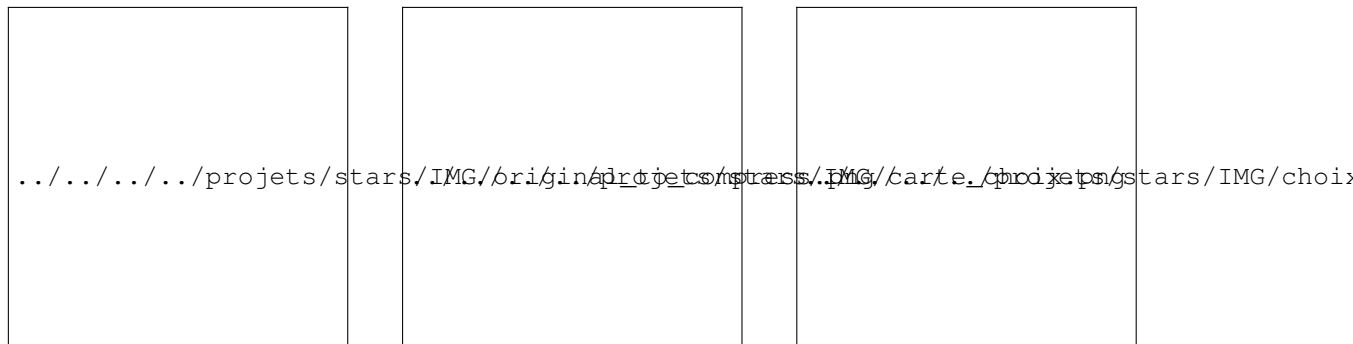


Figure 20. An image to encode; the map of the best experts, chosen independently for each pixel [each expert is represented by one color]; the expert map obtained with our approach enforcing spatial coherence in the expert choice.

## 6.7. People detection using RGB-D cameras

**Participants:** Anh-Tuan Nghiem, François Brémont.

**keywords:** people detection, HOG, RGB-D cameras

With the introduction of low cost RGB-D cameras like Kinect of Microsoft, video monitoring systems have another option for indoor monitoring beside conventional RGB cameras. Comparing with conventional RGB camera, reliable depth information from RGB-D cameras makes people detection easier. Besides that, constructors of RGB-D cameras also provide various libraries for people detection, skeleton detection or hand detection etc. However, perhaps due to high variance of depth measurement when objects are too far from the camera, these libraries only work when people are in the range of 0.5 to around 4.5 m from the cameras. Therefore, for our own video monitoring system, we construct our own people detection framework consisting of a background subtraction, a people classifier, a tracker and a noise removal component as illustrated in figure 21 .

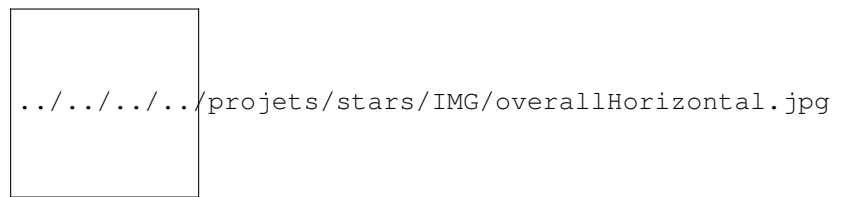


Figure 21. The people detection framework

In this system, the background subtraction algorithm is designed specifically for depth data. Particularly, the algorithm employs temporal filters to detect noise related to imperfect depth measurement on some special surface.

The people classification part is the extension of the work in [79]. From the foreground region provided by the background subtraction algorithm, the classification first searches for people head and then extracts HOG like features (Histogram of Oriented Gradient on binary image) above the head and the shoulder. Finally, these features are classified by a SVM classifier to recognise people.

The tracker links detected foreground regions in the current frame with the ones from previous frames. By linking objects in different frames, the tracker provides useful history information to remove noise as well as to improve the sensitivity of the people classifier.

Finally, the noise removal algorithm uses the object history constructed by the tracker to remove two types of noise: noise detected by temporal filter at the background subtraction algorithm and noise from high variance of depth measurement on objects far from the camera. Figure 22 illustrates the performance of noise removal on the detection results.

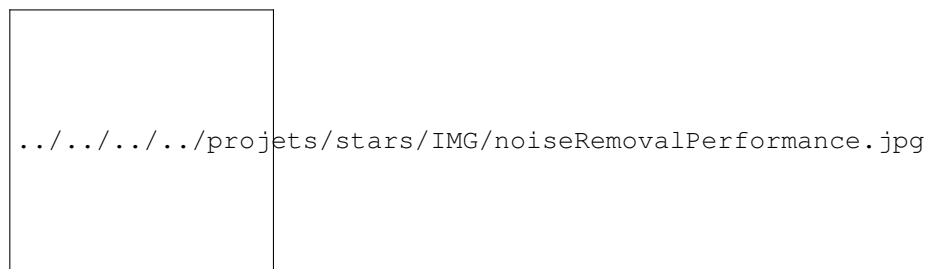


Figure 22. The people detection framework

The overall performance of our people detection framework is comparable to the one provided by Primesense, the constructor of RGB-D camera Microsoft Kinect.

Currently, we are doing extensive evaluation of the framework and the results will be submitted to a conference in the near future.

## 6.8. Online Tracking Parameter Adaptation based on Evaluation

**Participants:** Duc Phu Chau, Julien Badie, Kartick Subramanian, François Brémond, Monique Thonnat.

**Keywords:** Object tracking, parameter tuning, online evaluation, machine learning

Several studies have been proposed for tracking mobile objects in videos [50]. For example we have proposed recently a new tracker which is based on co-inertia analysis (COIA) of object features [44]. However the parameter tuning is still a common issue for many trackers. In order to solve this problem, we propose an online parameter tuning process to adapt a tracking algorithm to various scene contexts. The proposed approach brings two contributions: (1) an online tracking evaluation, and (2) a method to adapt online tracking parameters to scene contexts.

In an offline training phase, this approach learns how to tune the tracker parameters to cope with different contexts. Different learning schemes (e.g. neural network-based) are proposed. A context database is created at the end of this phase to support the control process of the considered tracking algorithm. This database contains satisfactory parameter values of this tracker for various contexts.

In the online control phase, once the tracking quality is evaluated as not good enough, the proposed approach computes the current context and tunes the tracking parameters using the learned values.

The experimental results show that the proposed approach improves the performance of the tracking algorithm and outperforms recent state of the art trackers. Figure 23 shows the correct tracking results of four people while occlusions happen. Table 3 presents the tracking results of the proposed approach and of some recent trackers from the state of the art. The proposed controller increases significantly the performance of an appearance-based tracker [63]. We obtain the best *MT* value (i.e. mostly tracked trajectories) compared to state of the art trackers.

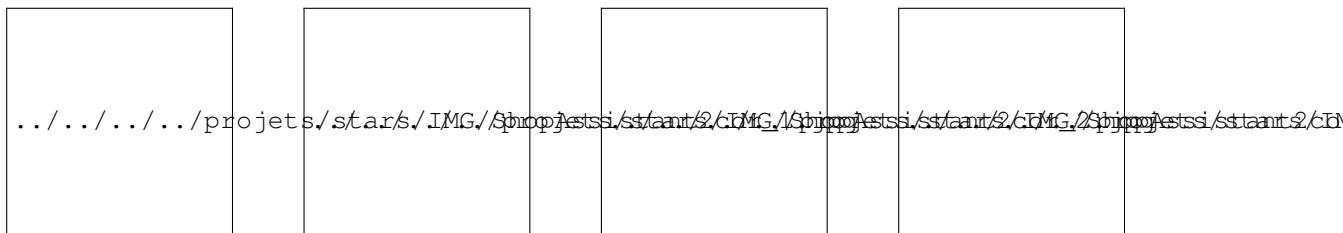


Figure 23. Tracking results of four people in the sequence *ShopAssistant2cor* (Caviar dataset) are correct, even when occlusions happen.

This work has been published in [33], [34].

## 6.9. People Detection, Tracking and Re-identification Through a Video Camera Network

**Participants:** Malik Souded, François Brémond.

**keywords:** People detection, Object tracking, People re-identification, Region covariance descriptors, SIFT descriptor, LogitBoost, Particle filters.

Table 3. Tracking results for the Caviar dataset. The proposed controller improves significantly the tracking performance. MT: Mostly tracked trajectories, higher is better. PT: Partially tracked trajectories. ML: Most lost trajectories, lower is better. The best values are printed in bold.

Approaches	MT (%)	PT (%)	ML (%)
Xing et al. [92]	84.3	12.1	3.6
Li et al. [76]	84.6	14.0	1.4
Kuo et al. [74]	84.6	14.7	<b>0.7</b>
D.P Chau et al. [63] without the proposed approach	78.3	16.0	5.7
<b>D.P Chau et al. [63] with the proposed approach</b>	<b>85.5</b>	9.2	5.3

This work aims at proposing a whole framework for people detection, tracking and re-identification through camera networks. Three main constraints have guided this work: high performances, real-time processing and genericity of the proposed methods (minimal human interaction/parametrization). This work is divided into three separate but dependent tasks:

### 6.9.1. People detection:

The proposed approach optimizes state-of-the-art methods [89], [93] which are based on training cascades of classifiers using the LogitBoost algorithm on region covariance descriptors. The optimization consists in clustering negative data before the training step, and speeds up both the training and detection processes while improving the detection performance. This approach has been published this year in [46]. The evaluation results and examples of detection are shown in Figures 24 and 25.

### 6.9.2. Object tracking:

The proposed object tracker uses a state-of-the-art background subtraction algorithm to initialize objects to track, with a collaboration of the proposed people detector in the case of people tracking. The object modelling is performed using SIFT features, detected and selected in a particular manner. The tracking process is performed at two levels: SIFT features are tracked using a specific particle filter, then object tracking is deduced from the tracked SIFT features using the proposed data association framework. A fast occlusion management is also proposed to achieve the object tracking process. The evaluation results are shown in Figure 26.

### 6.9.3. People re-identification:

A state-of-the-art method for people re-identification [67] is used as a baseline and its performance has been improved. A fast method for image alignment for multiple-shot case is proposed first. Then, texture information is added to the computed visual signatures. A method for people visible side classification is also proposed. Camera calibration information is used to filter candidate people who do not match spatio-temporal constraints. Finally, an adaptive feature weighting method according to visible side classification concludes the improvement contributions. The evaluation results are shown in Figure 27.

This work has been published in [28].

## 6.10. People Retrieval in a Network of Cameras

**Participants:** Sławomir Bąk, Marco San Biago, Ratnesh Kumar, Vasanth Bathrinarayanan, François Brémond.

**keywords:** Brownian statistics, re-identification, retrieval



```
../../../../projets/stars/IMG/People_Detector_Results.png
```

Figure 24. People detector evaluation and comparison on Inria, DaimlerChrysler, Caltech and CAVIAR datasets.

```
../../../../projets/stars/IMG/People_Detection_Examples.png
```

Figure 25. Some examples of detection using the proposed people detector.

```
../../../../projets/stars/IMG/Object_Tracking_Results.png
```

Figure 26. Object tracking evaluation on: (a) CAVIAR dataset using MT, PT and ML metrics. (b) ETI-VS1-BE-18-C4 sequence from ETISEO dataset, using ETISEO metrics.

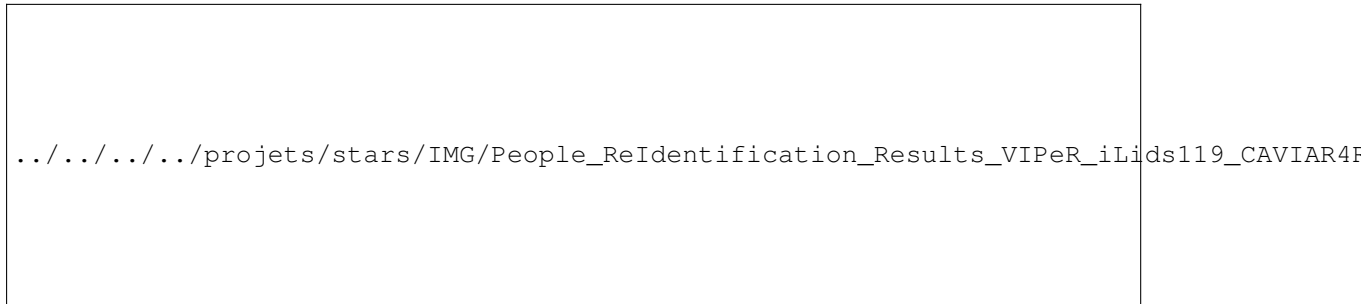


Figure 27. People re-identification evaluation on VIPeR (left), iLids-119 (middle) and CAVIAR4REID (right) datasets.

**Task.** Person re-identification (also known as multi-camera tracking) is defined as the process of determining whether a given individual has already appeared over a network of cameras. In most video surveillance scenarios, features such as face or iris are not available due to video low-resolution. Therefore a robust modeling of the global appearance of an individual (clothing) is necessary for re-identification. This problem is particularly hard due to significant appearance changes caused by variations in view angle, lighting conditions and different person pose. In this year, we focused on the two following aspects: new image descriptors and a design of a retrieval tool.

**New image region descriptors.** We have evaluated different image descriptors *w.r.t.* their recognition accuracy. As the covariance descriptor achieved the best results, we have employed this descriptor using different learning strategies to achieve the most accurate model for representing a human appearance [51]. We have also proposed a new descriptor based on recent advances in mathematical statistics related to Brownian motion [31]. This new descriptor outperforms the classical covariance in terms of matching accuracy and efficiency. We show that the proposed descriptor can capture richer characteristics than covariance, especially when fusing nonlinearly dependent features, which is often the case for images. The effectiveness of the approach is validated on three challenging vision tasks: object tracking & person re-identification [31] and pedestrian classification (the paper submitted to conference CVPR 2014). In all our experiments, we demonstrate competitive results while in person re-identification and tracking we significantly outperform the state-of-the-art.

**New design of retrieval tool for a large network of cameras.** Owing to the complexity of the re-identification problem, current state of the art approaches have relatively low retrieval accuracy, thus a fully automated system is still unattainable. However, we propose a retrieval tool [30], [29] that helps a human operator to solve the re-identification task (see Figure 28). This tool allows a human operator to browse images of people extracted from a network of cameras: to detect a person on one camera and to re-detect the same person few minutes later on another camera. The main stream is displayed on the left of the screen, while retrieval results are shown on the right. The results show lists of the most similar signatures extracted from each camera (green boxes indicate the correctly retrieved person). Below the main stream window a topology of the camera network is displayed. Detection and single camera tracking (see the main stream) are fully automatic. The human operator only needs to select a person of interest, thus producing retrieval results (right screen). The operator can easily see a preview of the retrieval results and can go directly to the original video content.

**Perspectives.** Currently, we are working not only on invariant image descriptors, which provide high recognition accuracy, but also on improving the alignment of the person pose, while matching appearance from cameras with significant difference in viewpoint. In addition to re-identification technology, we also work on designing an intuitive graphical interface, an important tool for the human operator analyzing retrieval results. Displaying retrieval results from a large camera network is still an issue, even after applying time-space constraints (the usage of topology of cameras).



(a)

*Figure 28. Re-identification tool*

### Acknowledgements

This work has been supported by PANORAMA and CENTAUR European projects.

## 6.11. 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, improving tracking results

This work addresses the problem of estimating the quality of a tracking algorithm during runtime and correcting the anomalies found. Evaluating and tuning a tracking algorithm generally requires multiple runs and a ground truth. The proposed online evaluation framework, called global tracker, overcome these limitations by proposing a three-steps algorithm to improve tracking results in any kind of situations (mono-camera, camera network, 3D camera, ...).

The first step aims at correcting small tracking errors (when detections in consecutive frames are missing from an object trajectory) by interpolating the detected object data.

The second steps aims at detecting and correcting any anomaly found in the output of the tracking algorithm. For each detected object on each frame, we compute three different sets of features : the features that are computed using only data from the object (eg. appearance, size, ...), the features that measure the level of interaction between two objects (eg. occlusion level, density) and the features that measure the level of interaction between the object and the environment (eg. occlusion with background element, entering or leaving zones, ...). By studying the evolution of the coefficients of variation of each feature, some anomalies can be detected. Then, by minimizing an energy function based on the object-only features, we correct the output of the tracking algorithm.

The final step uses re-acquisition and re-identification methods to merge detected objects representing the same real object. This algorithm aims at correcting ID change errors when an object leaves the scene and reappears in another camera or when the object reappears after a long-term occlusion. The method used here is a constrained clustering algorithm that create groups of detections representing the same real object.

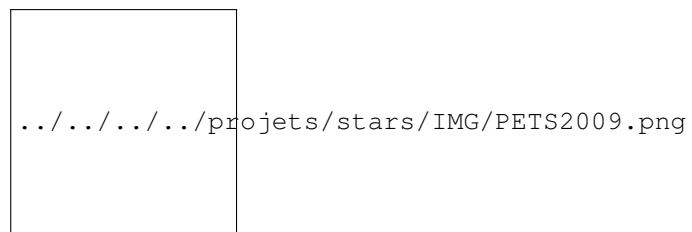


Figure 29. An example of the global tracker goal : offline learning of occlusion zones to manage online occlusion risks and optimize object trajectory.

This approach has been tested on several datasets (PETS 2009, Caviar, TUD-Stadtmitte). The results show that the global tracker, even associated with a tracking algorithm that does not have good results, can perform nearly as well as the state of the art and even exactly as well when a good tracker is used. On the Caviar dataset, the global tracker is even able to slightly outperform the result of the state of the art.

A part of this approach is described in an article published in AVSS 2013 [33]. This article shows how a tracking algorithm can use the global tracker results to tune its parameters and improve its results. This work was also used to improve the tracking results in 2 papers [38] [54].

Table 4. Comparison of the tracking results using CLEAR metrics on the sequence S2.L1 of the PETS 2009 dataset with and without the global tracker

Method	MOTA	MOTP	$\bar{M}$
Berclaz et al. [60]	0.80	0.58	0.69
Shitrit et al. [58]	0.81	0.58	0.70
Henriques et al. [72]	0.85	0.69	0.77
Chau et al. [33] without global tracker	0.62	0.63	0.63
Chau et al. [33] with global tracker	0.85	0.71	0.78

## 6.12. Human Action Recognition in Videos

**Participants:** Piotr Bilinski, Etienne Corvée, Slawomir Bak, François Brémond.

**keywords:** action recognition, tracklets, head detection, relative tracklets, bag-of-words.

In this work we address the problem of recognizing human actions in video sequences for home care applications.

Recent studies have shown that approaches which use a bag-of-words representation reach high action recognition accuracy. Unfortunately, these approaches have problems to discriminate similar actions, ignoring spatial information of features.

We propose a feature representation for action recognition based on dense point tracklets, head position estimation, and a dynamic coordinate system. Our main idea is that action recognition ought to be performed using a dynamic coordinate system corresponding to an object of interest. Therefore, we introduce a relative tracklet descriptor based on relative positions of a tracklet according to the central point of our dynamic coordinate system. As a center of our dynamic coordinate system, we choose the head position, providing description invariant to camera viewpoint changes. We use the bag-of-words approach to represent a video sequence and we capture global distribution of tracklets and relative tracklet descriptors over a video sequence. The proposed descriptors introduce spatial information to the bag-of-words model and help to distinguish similar features detected at different positions (*e.g.* to distinguish similar features appearing on hands and feet). Then we apply the Support Vector Machines with exponential chi-squared kernel to classify videos and recognize actions.

We report experimental results on three action recognition datasets (publicly available KTH and ADL datasets, and our locally collected dataset). Our locally collected dataset has been created in cooperation with the CHU Nice Hospital. It contains people performing daily living activities such as: standing up, sitting down, walking, reading a magazine, *etc.* Consistently, experiments show that our representation enhances the discriminative power of the tracklet descriptors and the bag-of-words model, and improves action recognition performance.

Sample video frames with extracted tracklets and estimated head positions are presented in Figure 30 .

This work has been published in [32].

### 6.12.1. Acknowledgments

This work was supported by the Région Provence-Alpes-Côte d’Azur. However, the views and opinions expressed herein do not necessarily reflect those of the financing institution

## 6.13. 3D Trajectories for Action Recognition Using Depth Sensors

**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], [62], [32], [90]) for Human Action Recognition to take advantage of the depth information from 3D sensors.



Figure 30. Sample video frames with extracted tracklets and estimated head positions for the KTH (first row), ADL (second row) and our locally collected dataset (third row).

We propose to add depth information to trajectory based algorithms ([32], [90]). 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 information map. Such computed 3D trajectories, combined with other appearance features (HOG, HOF), are subject to a Bag of Words model and SVM classifier.



*Figure 31. Visualization of MSR Daily Activity 3D data set. Left : video input frame; Middle : frame with detected trajectories (red = static points, green = detected trajectories); Right : corresponding depth map.*

The evaluation of our method was conducted on the "Microsoft Daily Activity3D" data set [91] which consist of 16 actions (drink, eat, read book, call cellphone, write on a paper, use laptop etc.) performed by 10 subjects.

The experiments showed that adding depth information to Dense Trajectories descriptor [90] gave gain in efficiency 57.72% to 64.12%. The mentioned work is going to be submitted in December 2013.

## 6.14. Unsupervised Sudden Group Movement Discovery for Video Surveillance

**Participants:** Sofia Zaidenberg, Piotr Bilinski, François Brémond.

**keywords:** Sudden Group Movement Discovery, Video Surveillance.

In this work we address the problem of discovering “sudden” movements in video surveillance videos. We propose an unsupervised approach which 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 provides a semantic interpretation of the scene. We have tested our approach on a dataset of nearly eight hours of videos recorded from two cameras in the Parisian subway for a European Project. For evaluation we annotated one hour of sequences containing 50 sudden movements. Our system, if parametrized to a high sensitivity, detects 100% of what the annotator considered as sudden potentially dangerous events, with a false positive rate of 21.2%. Setting the sensitivity to lower values we decrease the false positive rate to only 5.3% but we also decrease the success rate to 76%. An example of an unusual sudden movement annotated by a human and detected by our approach is presented in Figure 32. This work has been published in [49].

### 6.14.1. Acknowledgments

This work was supported by the Région Provence-Alpes-Côte d’Azur and by the European Community’s Seventh Framework Programme FP7/2007-2013 - Challenge 2 - Cognitive Systems, Interaction, Robotics - under grant agreement number: 248907-VANAHEIM. However, the views and opinions expressed herein do not necessarily reflect those of the financing institution.

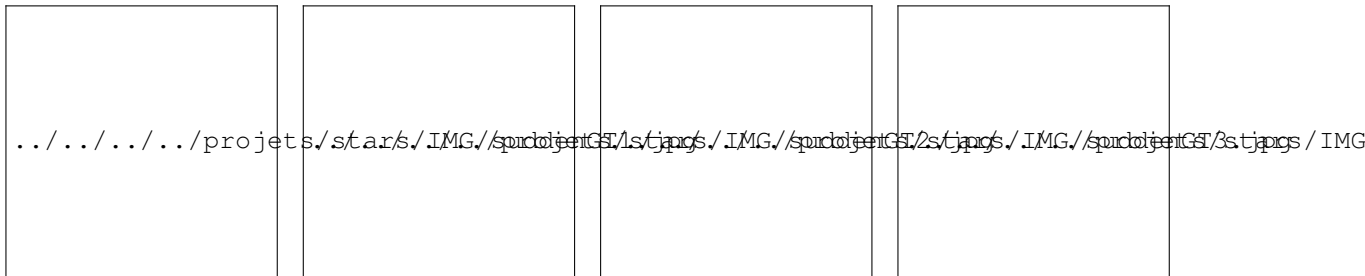


Figure 32. Example of an unusual sudden movement detected by our approach.

## 6.15. Group Behavior Understanding

**Participants:** Carolina Gárate, Sofia Zaidenberg, Julien Badie, François Brémond.

The goal is to recognize group behavior from videos. Dangerous and criminal behaviors are mostly observed within groups of people. The idea is to detect potentially dangerous situations while they are happening in the context of underground railway station security.

**keywords:** group tracking, scene understanding, group behavior recognition, video surveillance, event detection.



This research work considers a process consisting of 5 consecutive steps for video processing. The steps are: 1) segmentation, 2) blob detection, 3) physical objects tracking, 4) group tracking and 5) behavior recognition. Here, we are focussing on the last two phases: group tracking and behavior recognition.

The group tracking approach characterizes a group through three features: the average of the intra-object distance, the average standard deviations of speed and direction. The input for this algorithm is a set of trajectories for the physical objects (output of the stage 3: physical objects tracking) tracked by the algorithm described in [64]. The trajectories are processed using Mean-Shift clustering to create more reliable groups, see Figure 33 .

The behavior recognition approach identifies 2 steps: knowledge modeling and the event recognition algorithm. The ontology is implemented with the ScReK declarative language [94]. The grammar describes the objects and events using the extended BNF (Backus Naur Form) representation.

We process large amounts of long video surveillance data from Paris and Turin underground railway station to perform statistical analysis. This analysis automatically brings forward data about the usage of the station and the various behaviors of groups for different hours of the day. We present the results and interpretation of one month of processed data from a video surveillance camera in Turin subway.

One of the measures obtained in the experimentation is the agitation level which is represented by the variation of the size of the bounding box of a group. We consider 3 categories from *no agitation* ("Calm\_Group", having a bounding box with stable size) to *little agitation* ("Active\_Group") to *high agitation* ("Lively\_Group", the bounding box's size varies a lot, meaning that group members move around more often). Figure 34 shows that most of the time, this middle category predominates. Groups are neither too calm, nor too agitated. Moreover, it is more common for a group to be lively rather than calm. The rest of the results obtained were presented in [54].

## 6.16. Evaluation of an Activity Monitoring System for Older People Using Fixed Cameras

**Participants:** Carlos F. Crispim-Junior, Baptiste Fosty, Vasanth Bathrinarayanan, Salma Zouaoui-Elloumi, Monique Thonnat, François Brémond.

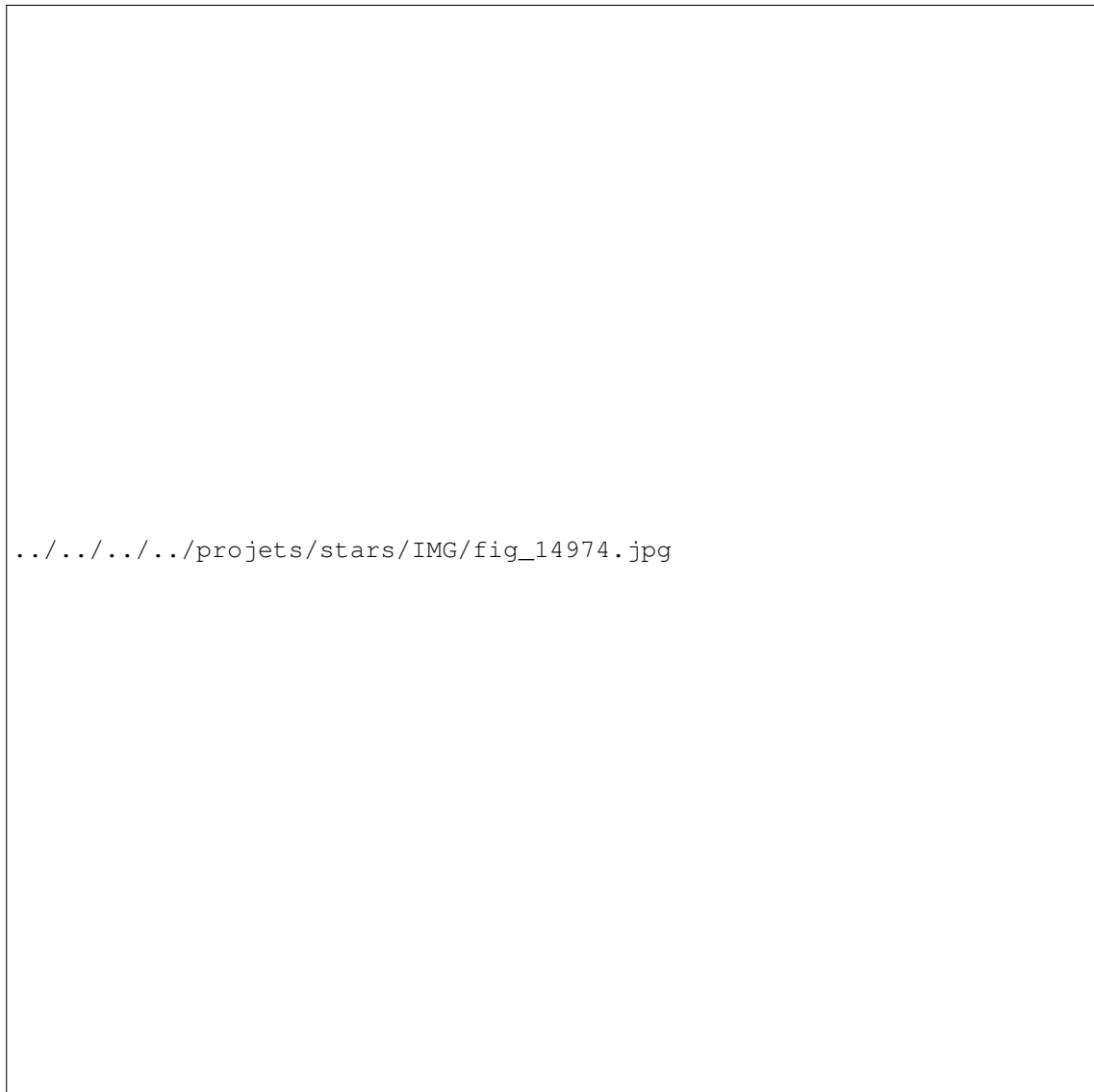
**keywords:** 2D-RGB cameras, RGB-D cameras, model-based activity recognition, older people

We have continued the evaluation of our model-based algorithm for complex activity recognition, now extending it to a larger dataset containing 38 older people participants undertaking instrumental activities of daily living (IADL) during 15 minutes (570 min. in total). The recordings have taken place in the observation room of the Memory Center of Nice hospital. Figure 35 presents the algorithm performance based on data obtained from a 2D-RGB video camera. A summary of the recognized activities (e.g., duration, frequency) is produced at the end of the event recognition task to be provided to doctors as a basis for the assessment of patient performance on IADL. This approach description and the evaluation results are published in 2013 AVSS Conference (see details in [36]). Figure 36 illustrates an example of a patient being monitored. Blue dots illustrates previous positions of the person in the scene.

The proposed approach has been also evaluated using a RGB-D camera, as this camera increases the robustness of the monitoring system against environment illumination changes and also eases the deployment of the system by providing real 3-D information on the scene. The evaluation of the RGB-D-based activity monitoring system has been published in [38]. A live demonstration of this system has also been presented and applied in the scope of Dem@care project (a FP7 project devoted to multi-sensor older people monitoring) in the exhibition held in November 2013 in conjunction with the 2013 edition of the ICT (Information Communication Technologies) Conference in Vilnius - Lithuania.

## 6.17. A Framework for Activity Detection of Older People Using Multiple Sensors

**Participants:** Carlos F. Crispim-Junior, Qiao Ma, Baptiste Fosty, François Brémond, Monique Thonnat.



*Figure 33. Illustration of a still group: in the right entrance zone a group of subway security staff is standing still.*



*Figure 34. Comparison of speed of groups and variation of group's bounding boxes during the day.*



Figure 35. Evaluation of Algorithm performance on 38 patient video recordings using a RGB camera

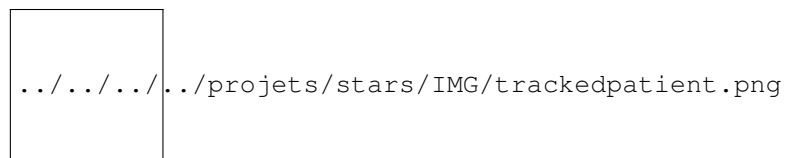


Figure 36. Example of a patient been monitored by the described system

**keywords:** model-based activity recognition, multi-sensor, Dempster-Shafer, Evidence Theory, older people.

We have extended our framework for multi-sensor activity detection by proposing a probabilistic approach for mutually exclusive activity conflict scenario. The proposed approach pre-learned a coefficient of reliability of each sensor with respect to each activity. The combination of the activities detected by multiple sensors is performed using the Dempster-Shafer evidence theory with an adapted combination rule based on runtime data from the sensor and the pre-computed coefficients of reliability. The combination of activities detected by multiple sensors can be performed in an iterative fashion taking into account several sensor contributions (see Fig.37 ). Tab. 5 presents the early results of the proposed probabilistic method at combining activities detected using RGB and RGB-D cameras. Results are presented individually for each camera and for the proposed approach.



*Figure 37. Probabilistic Approach Combination Scheme*

The proposed fusion scheme performs better than the camera individual process in most of cases even in the presence of noise (see the RGB-D individual result of sensitivity for standing posture detection). The complete evaluation of the proposed approach is published in [43]. The developed probabilistic approach is now integrated into our previously developed framework for multi-sensor activity modeling. The new framework version allows experts to precisely define the sensors which will be used to detect each activity, or to automatically (and then probabilistic) combine multiple instances of (conflicting) activities detected by different sensors. The new framework proposal for multi-sensor activity detection has been published in [37].

## **6.18. Walking Speed Detection on a Treadmill using an RGB-D Camera**

**Participants:** Baptiste Fosty, François Brémond.

Table 5. Performance of the proposed probabilistic approach on posture detection

Posture	Standing		Sitting	
	Precision	Sensitivity	Precision	Sensitivity
Sensor				
RGB	84.29	69.41	79.82	91.58
RGB-D	100.00	36.47	86.92	97.89
Fusion	82.35	91.30	91.04	95.31

**keywords:** RGB-D camera analysis, older people, serious games

Within the context of the Az@Game project, we have studied the potential of the RGB-D camera (Red Green Blue + Depth) for the control of a serious game dedicated to older people suffering from Alzheimer disease. Within this game, the patient is invited to perform some physical and cognitive tasks (walking on a treadmill at different speeds, performing gestures to control his/her character in the game, managing money) in order to assess the evolution of the disease, to stimulate them and improve their abilities. In this context, one of our goal is the computation of the walking speed of a person on a treadmill.

The proposed solution is divided into three distinct steps :

- people detection and tracking using a background subtraction algorithm.
- feet detection : based on the cloud of 3D points of the person, and more particularly on the lower body part, the axis of each leg is computed. The corresponding foot is then defined as the projection of the lowest point of the leg on the leg axis (see Fig. 38 , right picture).
- speed computation : computed from the successive positions of the feet, more precisely from the distances between the feet (see Fig. 38 , left graph representing the distance between the feet). Each time this distance reaches a local maximum (corresponding to each step), the current speed is computed as the maximum over the time since the last step. The speed is then averaged with the previous computed speed to smooth the values (see Fig. 38 , middle graph representing the speed in function of time).

Concerning the results, the first experimentation of the algorithm shows that, qualitatively, the computed walking speed is proportional to the real speed. Future work will focus on experimenting the proposed system on a larger scale (different people, location, etc.) in order to validate the approach. We will also focus on trying to detect arm gestures to have more control on the serious game.

About the Dem@Care project and the previous work on the activity recognition system developed to extract automatically and objectively evidences of early symptoms of Alzheimer's disease for older people, this contribution has been published at ASROB 2013, Workshop on Assistance and Service Robotics in a Human Environment (see [38]).

## 6.19. Serious Game for older adults with dementia

**Participants:** Minh Khue Phan Tran, François Brémond.

**keywords:** human-machine interaction, serious games, Alzheimer, activity recognition

Serious Games is carried out within the framework of the *Az@GAME project*. This project is to create games offering patient-oriented scenarios so as to measure their health progress, improve their physical fitness, stimulate their cognitive abilities, and help maintain their social skills. The main objective is to design a system interacting with older adults suffering from Alzheimer's or Alzheimer's related diseases. The three challenges in designing the system are:

- **perception's precision** : how does the system choose the "*best moments*" to interact with a patient ?
- **attractive-visualization** : how does the system make the patients comfortable?
- **ease of interaction** : how can it optimize the interaction with the patients ? In what way ?



*Figure 38. Walking speed computation on a treadmill. The left graph is the representation of the distance between the feet as a function of time. The middle graph is the representation of the speed of the person as a function of time. The right picture is the RGB-D camera view with the people detection and current speed. The two yellow circles show the positions of the detected feet.*

The first prototype is under development. The system consists of two parts: Recognition and Interaction . Each component requires a 3D camera (Microsoft Kinect for the recognition component and Asus Xtion Pro Live Camera for the interaction component). The recognition part consists in observing the scene and deciding the best time to interact with a patient via the Asus camera. It uses the SUP framework. Afterwards, the interactive system tries to engage the patient via an interface and through Microsoft Kinect, the patient can interact with the interface using voice or gesture. The interface is designed with the Unity 3D game engine.

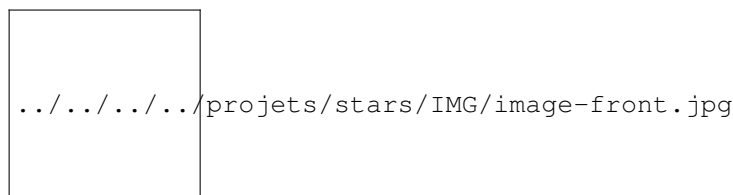


Figure 39. Example of a game and its gesture interface.

The first experiment will be conducted in the coffee area. The aim is to test the functionality of the system and measure its accuracy and effectiveness. The system will observe the scene and invite people who are getting coffee or taking a break to play the game. Depending on the interaction with the person, the system will offer different scenarios. Videos will be recorded, with the consent of the subject, in order to evaluate the effectiveness of system. The recorded videos and meta-data provided by SUP will be evaluated to determine the accuracy of the system.

## 6.20. Unsupervised Activity Learning and Recognition

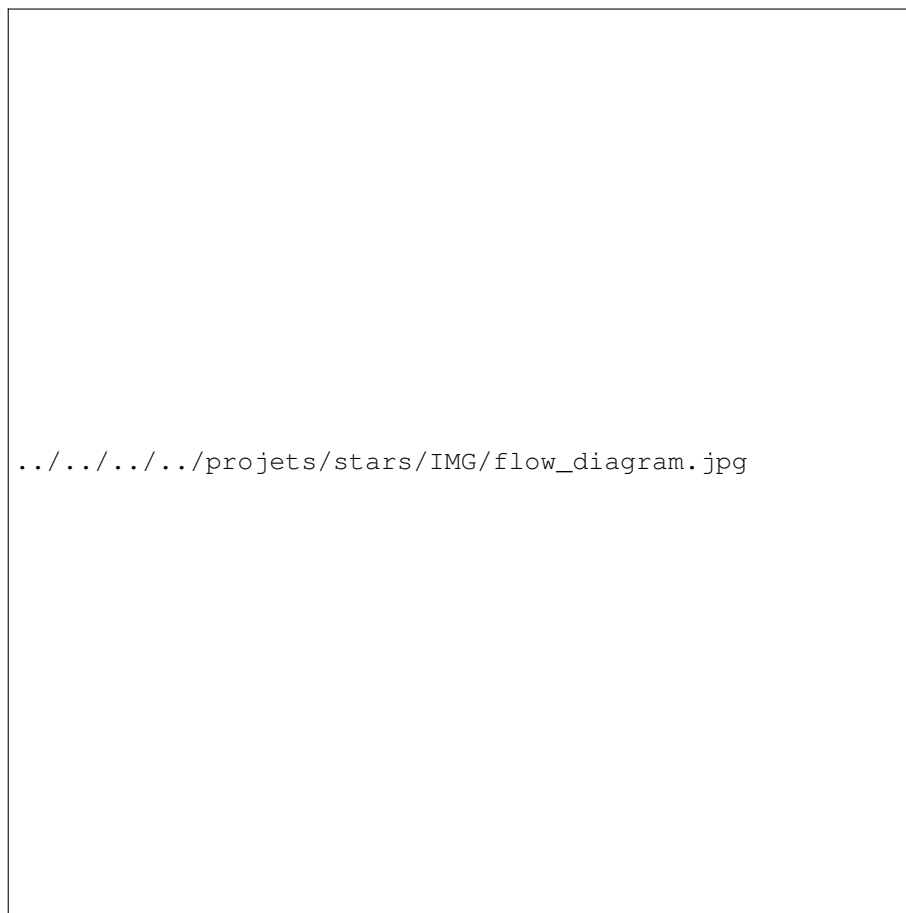
**Participants:** Serhan Cosar, Salma Zouaoui-Elloumi, François Brémond.

**keywords:** Unsupervised activity learning, hierarchical activity models, monitoring older people activities

The aim of this work is to monitor older people activities at hospital or at home environment in an unsupervised manner. We have extended the work in [81] that was initially based on user interface to label activities and proposed a new strongly unsupervised framework. It enables the discovery, modeling, and recognition of activities without user interaction. One advantage of this approach is that the framework learns individual behavioral patterns in unstructured scenes without restraining people to act based on a manually pre-defined model. The Figure 40 -(a) presents the off-line learning steps of this framework. It takes as input a set of videos pre-processed to obtain trajectory information of people in the scene. Using the trajectory information (global position and pixel tracklets of body parts) of each person, zones of interest, where the person performs an activity, are learned. As in [81], we obtain three levels of zones using  $k$ -means clustering for different  $k$  values. The obtained zones are used to create different levels of events from the coarser to the finer ones. Based on the three levels of events, a hierarchical model of activities is learned to represent each action (Figure 40 -(a)). For each new video, an on-line recognition process is performed by using the previously learned zones and models of activities (Figure 40 -(b)).

We have evaluated the performance of the unsupervised algorithm for RGB-D and 2D camera using 8 videos and 10 videos, respectively. Half of the videos are used for learning zones and models of activities. Videos are recorded in CHU Nice hospital while older people are visiting their doctors and include the following actions: “talking on the phone”, “preparing drugs”, “sitting at the table”, “preparing tea”, “looking at the bus map”, “watching TV” and “paying bill”. The trajectory information for 2D camera is obtained using the method in [81]. For RGB-D camera, we have used the person detection algorithm in [79] and tracking algorithm in [33]. The results obtained for both cameras are presented in Table 6 and Table 7, respectively. We have used the following metrics to evaluate the framework: TP: True positive, FP: False positive, FN: False Negative,





*Figure 40. The flow diagram of the unsupervised activity recognition framework: (a) off-line learning phase and (b) on-line recognition phase*

Sensitivity and Precision. According to the trajectory information, sometimes  $k$ -means clustering produces zones that are actually union of more than one zones. For such cases, we have combined the actions and presented as one single action.

Table 6. The recognition results obtained by using the 2D camera.

Actions	Instances	TP	FP	FN	Sensitivity(%)	Precision (%)
Paying bill	13	5	0	8	38.46	100
Preparing drugs	7	5	5	2	71.42	50
Looking at bus map+Watching TV	21	6	3	15	28.57	66.66
Sitting at the table	18	6	10	12	33.33	37.5
Talking on the phone	23	17	1	6	73.91	94.44
Preparing tea	23	11	3	12	47.82	78.57

Table 7. The recognition results obtained by using the RGB-D camera.

Actions	Instances	TP	FP	FN	Sensitivity(%)	Precision (%)
Paying bill + Watching TV	13	12	8	1	92.3	60
Preparing drugs	5	5	0	0	100	100
Looking at bus map	9	9	10	0	100	47.36
Sitting at the table	8	4	34	4	50	10.52
Talking on the phone	14	13	1	1	92.85	92.85
Preparing tea	16	9	5	7	56.25	64.28

As it can be seen in the tables, we obtain higher recognition rates by using the information coming from RGB-D camera.

Table 6 shows that for “talking on the phone” and “preparing drugs” actions occurring in two distant zones, using 2D camera gives high recognition rates (higher than 70 %). However, the actions “looking at bus map”, “watching TV” and “sitting at the table” are misclassified (low TP and high FP). Since the zones of these actions are very close to each other, the actions occurring in the borders are not well recognized. The reason of high FN is due to the problems in detection and tracking with 2D video cameras. The process of trajectory extraction described in [81] sometimes fails to track people. Because of the inadequate trajectory information, we have many FNs. Therefore, a better detection can considerably enhance the recognized actions.

By using the information coming from RGB-D camera, except for “sitting at the table” and “preparing tea” actions, we achieve high level of recognition rates (Table 7). However, similar to 2D camera, the recognition of “sitting at the table”, “paying bill” and “watching TV” actions fails because the learned zones in the scene are very close to each other. Hence, we have many false positives (FP) and false negatives (FN) for “sitting at the table” and “preparing tea” actions.

In the light of the preliminary experimental results, we can say that this unsupervised algorithm has a potential to be used for automated learning of behavioral patterns in unstructured scenes, for instance in home care environment for monitoring older people. Since the current framework does not require the user interaction to label activities, an evaluation process on big datasets will be easily performed. The proposed framework gives one action at each zone in an unsupervised way. We are currently focusing on refining the actions for each

zone by using the pixel tracklets of the person's body parts. This will be achieved by performing clustering among activity models. As an example, the action of "sitting at the table" will be decomposed to "reading newspaper while sitting at the table" and "distributing cards while sitting at the table".

## 6.21. Extracting Statistical Information from Videos with Data Mining

**Participants:** Giuseppe Donatiello, Hervé Falciani, Duc Phu Chau, François Brémond.

**keywords:** video data mining, activity recognition, clustering techniques

### Objective

Manual video observation is becoming less practical due to growing size of data. To tackle this problem, we have built a system to retrieve videos of interest thanks to an index based on activities recognized in an automated manner. We automatically detect activities in videos by combining data mining and computer vision to synthesize, analyze and extract valuable information from video data.

### Approach

Our research introduces a new method for extracting statistical information from a video. Specifically, we focus on context modeling by developing an algorithm that automatically learns the zones in a scene where most activities occur by taking as input the trajectories of detected mobiles. Using K-means clustering, we define activity zones characterizing the scene dynamics, we can extract then people activities by relating their trajectories to the learned zones.

### Results

To evaluate our system we have extended the OpenJUMP framework, an open source for Geographic Information System (GIS). The end user can have an overview of all activities of a large video, with the possibility of extracting and visualizing activities classified as usual or unusual. We have tested our approach on several videos recorded in subways in Turin (Italy) and Paris, as shown below, some examples of unusual activities (Figures 41, 42 and 43). The system has been showed in a live demonstration at RATP company in Paris for the European project Vanaheim (<http://www.vanaheim-project.eu/>).

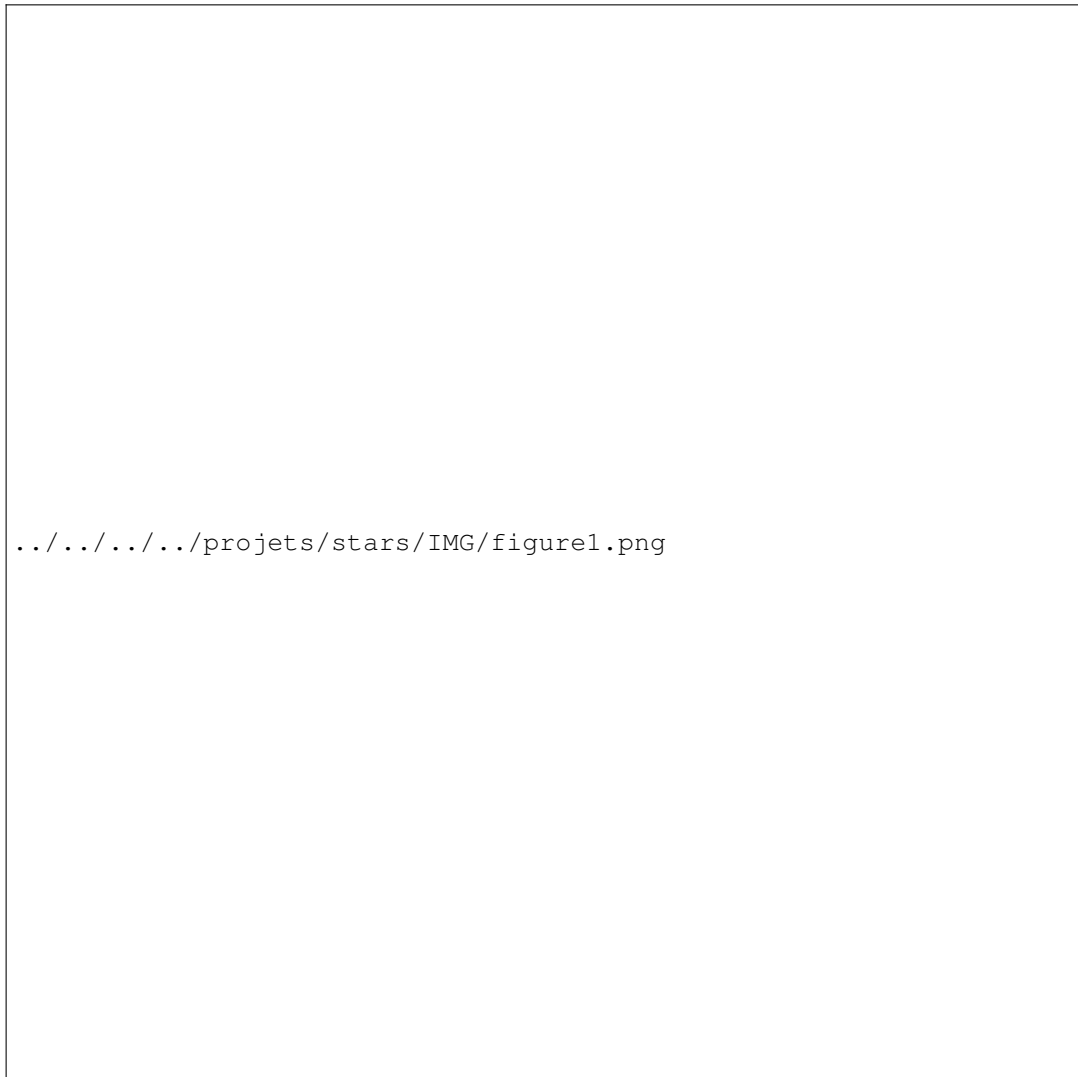
## 6.22. SUP

**Participants:** Julien Gueytat, François Brémond.

**keywords:** SUP, Software, Video Processing

### Presentation

SUP is a Scene Understanding Software Platform writtent in C++ designed for analyzing video content. (see Figure 44 ) SUP is splitting the workflow into several modules, such as acquisition, segmentation, etc., up to activity recognition. Each module has a specific interface, and different plugins implementing these interfaces can be used for each step of the video processing.



*Figure 41. Results : three categories of discovered abnormal behaviors.*



Figure 42. Trajectories and zones from camera M3114 in Paris (the yellow line is the trajectory selected by the user; the numbers represent the weight belonging to each segment : the higher, the more the people go through it).



Figure 43. Left : Person sitting. Middle : Person standing for a long time. Right : Unusual path.



*Figure 44. SUP workflow*

The plugins cover the following research topics:

- algorithms : 2D/3D mobile object detection, camera calibration, reference image updating, 2D/3D mobile object classification, sensor fusion, 3D mobile object classification into physical objects (individual, group of individuals, crowd), posture detection, frame to frame tracking, long-term tracking of individuals, groups of people or crowd, global tacking, basic event detection (for example entering a zone, falling...), human behaviour recognition (for example vandalism, fighting,...) and event fusion;
- languages : scenario description, empty 3D scene model description, video processing and understanding operator description;
- knowledge bases : scenario models and empty 3D scene models;
- algorithms of 2D & 3D visualisation of simulated temporal scenes and of real scene interpretation results;
- algorithms for evaluation of object detection, tracking and event recognition;
- learning techniques for event detection and human behaviour recognition;
- algorithms for image acquisition (RGB and RGBD cameras) and storage;
- algorithms for video processing supervision;
- algorithms for data mining and knowledge discovery;
- algorithms for image/video indexation and retrieval.

The software is already widely disseminated among researchers, universities, and companies:

- PAL Inria partners using ROS PAL Gate as middleware
- Nice University (Informatique Signaux et Systèmes de Sophia), University of Paris Est Créteil (UPEC - LISSI-EA 3956)
- European partners: Lulea University of Technology, Dublin City University,...
- Industrial partners: Toyota, LinkCareServices, Digital Barriers

And new sites are coming: EHPAD Valrose, Institut Claude Pompidou, Delvalle and Biot.

### Improvements

Our team focuses on developing a Scene Understanding Platform (SUP). This platform has been designed for analyzing video content. SUP is able to recognize events such as 'falling', 'walking' of a person. We can easily build new analyzing systems thanks to a set of algorithms also called plugins. The order of those plugins and their parameters can be changed at run time and the result visualized on a dedicated GUI. This platform has many more advantages such as easy serialization to save and replay a scene, portability to Mac, Windows or Linux, and easy deployment to quickly setup an experimentation anywhere. All those advantages are available since we are working together with the Inria software developer team SED. Many Inria teams are pushing together to improve a common Inria development toolkit DTK. Our SUP framework is one of the DTK-like framework developed at Inria.

Currently, the OpenCV library is fully integrated with SUP. OpenCV provides standardized datatypes, a lot of video analysis algorithms and an easy access to OpenNI sensors such as the Kinect or the ASUS Xtion PRO LIVE.

Updates and presentations of our framework can be found on our team website <https://team.inria.fr/stars/software>. Detailed tips for users are given on our Wiki website <http://wiki.inria.fr/stars> and sources are hosted thanks to the software developer team SED.

## 6.23. Model-Driven Engineering for Activity Recognition

**Participants:** Sabine Moisan, Jean-Paul Rigault, Luis Emiliano Sanchez.

**keywords:** Feature Model Optimization, Software Metrics, Requirement specification, Component-based system, Dynamic Adaptive Systems, Model-Driven Engineering, Heuristic Search

The domain of video surveillance (VS) offers an ideal training ground for Software Engineering studies, because of the huge variability in both the surveillance tasks, the video analysis algorithms and the context. Such systems require run time adaptation of their architecture to react to external events and changing conditions in their context of execution.

The feature model formalism is widely used to capture variability, commonalities and configuration rules of software systems. We thus use feature modeling to represent the variability of both the specification and component views of video surveillance systems. We also include cross-tree constraints that formalize extra feature dependencies.

Based on this feature model, we can both select an initial system configuration at deployment time and dynamically adapt the current configuration at run time. This year we focused on runtime adaptation, from feature model to running components.

### 6.23.1. Configuration Adaptation at Run Time

In the continuation of our work on metrics on feature models, we have integrated a configuration selection algorithm in our feature model manager. Context changes or user interactions imply to dynamically reconfigure the model (selecting or deselecting features). Following *model at run time* techniques, we are able to determine the set of valid configurations to apply in a new execution context. Since only one configuration can be applied at a given time, the role of the selection algorithm is to select the “best” one.

To this end we enriched our feature representation with a set of *quality attributes* that correspond to a monotonic quantification of interesting aspects of the system quality. Examples are response time, accuracy, availability, performance, component switching time, etc. The configuration selection algorithm optimises a cost function, a linear weighted combination of the quality attributes. Thus we can rank the possible valid configurations and choose an optimal one. Our algorithm is a variant of the Best-First Search algorithm, a heuristic graph search technique. It starts with the set of valid configurations, which is a feature model where some features are *unselected*. Then it performs a systematic search in a graph where nodes are configurations and edges are selections or deselections of unselected features. The goal is to obtain a full configuration (one without unselected features) optimizing the cost function. The algorithm is parameterized with different strategies and associated heuristics with different optimality and efficiency characteristics.

Search strategies decide which node to visit next. We choose two well-known informed strategies that rely on heuristic functions as choice criteria. First we used a variant of the A\* algorithm, BF\*, but with a node-cost function instead of a path-cost one; it favors optimality over efficiency. Second, we implemented a Greedy Best-First Search (GBFS) strategy, where the next visited node is the best successor of the current one; it favors efficiency over optimality.

Computing the exact value of the cost function for a partial configuration is too expensive. We thus use heuristics to obtain a quick *estimate*. We have tested two sorts of heuristics. The simplest one,  $H_A$ , ignores the differences between the various sorts of groups (AND, OR, XOR) in the feature model and does not consider cross-tree constraints; it is fast but not very accurate. The second one,  $H_B$ , just drops the cross-tree constraints; it is thus more accurate, yet at an higher cost.

We have run experiments using large (randomly generated) feature models and compared completeness, optimality and efficiency of the selection algorithm, with different combinations of strategies and heuristics [42]. From our experiments, the GBFS strategy with heuristics  $H_B$  appears as the ideal option for real time systems that have to adapt in bounded time. This strategy ensures polynomial time complexity and guarantees optimality over 90%, which is good enough for our purpose (see figure 45). On the other hand, BF\* strategy with heuristics  $H_B$  is ideal for offline decisions, such as defining the initial configuration of a system. Although this search strategy takes a significant time to compute, this is acceptable at deployment time to obtain the optimal configuration.

### 6.23.2. Run Time Components

When a configuration has been chosen, we must implement it with real components. We consider a configuration of a video-surveillance processing chain as a set of running components, that can be tuned, removed,





Figure 45. Optimality comparison of heuristics using GBFS strategy with respect to the number of attributes in the cost function

added, or replaced dynamically, in response to events. To apply such configuration changes, we need a way to represent and dynamically manipulate the components themselves.

In a first attempt, we used an OSGi-like C++ framework (SOF, Service Oriented Framework). However, SOF did not really fulfill our needs. First, SOF is the only C++ OSGi framework that we could find and its C++ implementation deserves some improvement. Moreover, like OSGi, it relies on the notion of “service”, as can be found in Web applications, but which does not really fit our real time requirements. This notion of service is not our concern and makes programming more complicated than necessary.

Thus, we decided to define our own component module and to integrate it in a multi-threaded layer, easy to use for our end-users who are video system developers. Each component runs by default in its own thread and communicates with other components through standardized communication channels. Our goal is to provide end-users with simple patterns to package their video codes into components. Thus we hide as much as possible the technical details such as threading synchronization, data exchange, and mechanisms for component management (replacement, tuning...) ensuring a continuous process.

We are currently setting up this framework on a simple video detection pipeline with OpenCV-based components. Then we shall integrate it within our Model at Run Time architecture.

## 6.24. Scenario Analysis Module

**Participants:** Annie Ressouche, Daniel Gaffé, Narjes Ghrairi, Sabine Moisan, Jean-Paul Rigault.

**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.

Since last year, we relied on CLEM (see section 6.25) synchronous language to express the automata semantics of scenario models as Boolean equation systems. This year, we continue our research in this direction and we are studying a specific semantics of SAM language operators that translates any SAM program into Boolean equation system. Therefore, we will benefit from CLEM compilation technique to generate recognizer for each scenario model.

This year we focus on the definition of an execution machine able to transform asynchronous events coming from SUP or other devices into synchronous significant events feeding recognition engines generated by SAM. The execution machine can listen three types of asynchronous events: SUP events, Boolean sensors, sampled sensors and pulse train sensors. According to the sampling period of each sensor, the execution machine builds the significant events defining the synchronous logical instants which trigger the reaction of the scenario recognition engine. Thanks to the synchronous approach, scenario recognition engines are able to dynamically express the expected synchronous events of the next step; the execution machine takes into account of this information to filter relevant events. We perform several tests with real SUP data sets and the execution machine has a convincing behaviour (see [55]). To complement this work, we will integrate a notion of incompatible events which will make the execution machine more efficient and robust.

## 6.25. The Clem Workflow

**Participants:** Annie Ressousche, Daniel Gaffé, Joel Wanza Weloli.

**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 9) 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 mainly work on the implementation of new theoretical results concerning the foundation of LE semantics. We also design 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). In [69], we chosen an algebra which is a bilattice and we show that it is well suited to solve our problem. It is a new application of general bilattice theory [70]. This year, this approach has been improved, validated in CLEM compiler and published in [39]. Our compilation technique needs to represent Boolean equation systems with Binary Decision Diagrams (BDD) and we study and design a specific BDD library well suited to ours needs. From a practical point of view, we integrate new operators in LE language (sustain until, no emit, strong abort). We focus on automata extension which can consider now three types of transition: weak transition, strong transition and normal termination transition.

Second, 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 design the CLES simulator which interprets LEC. The actual version don't integrate the data part of the language and we plan to do this integration.

## 6.26. Multiple Services for Device Adaptive Platform for Scenario Recognition

**Participants:** Annie Ressousche, Daniel Gaffé, Mohammed Cherif Bergheul, 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 [88], [87], [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 behavioural 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.

In [84], [83], 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 supply GALAXY automata editor to express Mealy machines and we extend AUTOM2CIRCUIT compiler to generate the internal code of WComp (C#). AUTOM2CIRCUIT is a tool developed by D. Gaffé since several years which compiles an explicit representation of automata into Boolean Mealy machine and generate a large and opened set of targets. This work is a preliminary study to integrate this generation of C# into CLEM.

## TEXMEX Project-Team

# 6. New Results

## 6.1. Description of multimedia content

### 6.1.1. Multiscale image representations with component trees

**Participants:** Petra Bosilj, Ewa Kijak.

*Joint work with Sébastien Lefevre, IRISA/SEASIDE, France.*

The goal of this work is to study deeply the use of component trees, which aim at representing an image by the regions it contains at various scales through a tree-based structure, and their ability in the context of content-based image indexing and retrieval. Their invariance properties and their robustness to noise have motivated recent work in image indexing [83], [97], [98], but their usage in this field stays limited. The first part of this work was mainly dedicated to the study of various existing hierarchical representations. This leads to the presentation of a technique that arranges the elements of hierarchical representations of images according to a coarseness attribute [24]. The transformation is similar to filtering a hierarchy with a non-increasing attribute, and includes the results of multiple simple filterings with an increasing attribute. The transformed hierarchy can be then used for search space reduction prior to the image analysis process because it allows for direct access to the hierarchy elements at the same scale or a narrow range of scales.

### 6.1.2. Image representation

**Participants:** Rachid Benmokhtar, Jonathan Delhumeau, Guillaume Gravier, Philippe-Henri Gosselin, Hervé Jégou, Wanlei Zhao.

*Partially in collaboration with Patrick Pérez, Technicolor, France.*

Recent work on image retrieval have proposed to index images by compact representations encoding powerful local descriptors, such as the closely related vector of aggregated local descriptors (VLAD) and Fisher vector (FV). By combining them with a suitable coding technique, it is possible to encode an image in a few dozen bytes while achieving excellent retrieval results. We have pursued the research on this line of research by proposing two complementary contributions.

In [30], we revisited some assumptions proposed in this context regarding the handling of "visual burstiness", and shows that ad-hoc choices are implicitly done which are not desirable. Focusing on VLAD without loss of generality, we propose to modify several steps of the original design. Albeit simple, these modifications significantly improve VLAD and make it compare favorably against the state of the art.

In [65], we proposed a pooling strategy for local descriptors to produce a vector representation that is orientation-invariant yet implicitly incorporates the relative angles between features measured by their dominant orientation. This pooling is associated with a similarity metric that ensures that all the features have undergone a comparable rotation. This approach is especially effective when combined with dense oriented features, in contrast to existing methods that either rely on oriented features extracted on key points or on non-oriented dense features. The interest of our approach in a retrieval scenario is demonstrated on popular benchmarks comprising up to 1 million database images.

In [22], we propose to reduce the dimensionality of visual features for image categorization. We iteratively select sets of projections from an external dataset, using Bagging and feature selection thanks to SVM normals. Features are selected using weights of SVM normals in orthogonalized sets of projections. The bagging strategy is employed to improve the results and provide more stable selection. The overall algorithm linearly scales with the size of features, and is thus able to process large state-of-the-art image representations. Given Spatial Fisher Vectors as input, our method consistently improves the classification accuracy for smaller vector dimensionality, as demonstrated by our results on the popular and challenging PASCAL VOC 2007 benchmark.

### 6.1.3. Video classification

**Participants:** Kleber Jacques Ferreira de Souza, Guillaume Gravier, Philippe-Henri Gosselin.

*In collaboration with Silvio Jamil F. Guimarães, PUC Minas, Brazil.*

Most current motion descriptors for video classification are based on simple video segments, such as rectangular space-time blocks, or more recently rectangular space blocks that follow local trajectories. The aim of this study is to consider more complex video segments that better fit space-time elements of videos, thanks to recent methods for video segmentation proposed by S. Guimarães et al. These methods combine at the same time a fast extraction and stable regions, two essential properties for video indexing. The computation of local motion descriptors on these video segments lead to better video classification for human action recognition, when compared to current video indexing techniques.

### 6.1.4. Geo-localization of videos with multi-modality

**Participants:** Jonathan Delhumeau, Guillaume Gravier, Hervé Jégou.

*Joint work with Michele Trevisiol, Yahoo! Labs, Spain, who visited the team in 2012.*

Geotagging is the process of automatically adding geographical identification metadata to media objects, in particular to images and videos. In [63], we present a strategy to identify the geographic location of videos. First, it relies on a multi-modal cascade pipeline that exploits the available sources of information, namely the user upload history, his social network and a visual-based matching technique. Second, we present a novel divide & conquer strategy to better exploit the tags associated with the input video. It pre-selects one or several geographic area of interest of higher expected relevance and performs a deeper analysis inside the selected area(s) to return the coordinates most likely to be related to the input tags. The experiments were conducted as part of the MediaEval 2012 Placing Task, where we obtained the best results among the competitors when using no external information, i.e. not using any gazetteers nor any other kind of external information.

### 6.1.5. Violent key sound detection with audio words and Bayesian networks

**Participants:** Guillaume Gravier, Patrick Gros, Cédric Penet.

*Joint work with Claire-Hélène Demarty, Technicolor, France.*

We investigated a novel use of the well known audio words representations to detect specific audio events, namely gunshots and explosions, in order to get more robustness towards soundtrack variability in Hollywood movies [51]. An audio stream is processed as a sequence of stationary segments. Each segment is described by one or several audio words obtained by applying product quantization to standard features. Such a representation using multiple audio words constructed via product quantisation is one of the novelties described in this work. Based on this representation, Bayesian networks are used to exploit the contextual information in order to detect audio events. Experiments are performed on a comprehensive set of 15 movies, made publicly available. Results are comparable to the state of the art results obtained on the same dataset but show increased robustness to decision thresholds, however limiting the range of possible operating points in some conditions. Late fusion provides a solution to this issue.

## 6.2. Large scale indexing and classification

### 6.2.1. Parallelism and distribution for very large scale content-based image retrieval

**Participants:** Gylfi Gudmundsson, Diana Moise, Denis Shestakov, Laurent Amsaleg.

Two observations drove the design of the high-dimensional indexing technique developed in the framework of the Ph. D. thesis of Gylfi Gudmundson. Firstly, the collections are so huge, typically several terabytes, that they must be kept on secondary storage. Addressing disk related issues is thus central to our work. Secondly, all CPUs are now multi-core and clusters of machines are a commonplace. Parallelism and distribution are both key for fast indexing and high-throughput batch-oriented searching.

We developed a high-dimensional indexing technique called eCP. Its design includes the constraints associated to using disks, parallelism and distribution. At its core is an non-iterative unstructured vectorial quantization scheme. eCP builds on an existing indexing scheme that is main memory oriented. The first contribution in eCP is a set of extensions for processing very large data collections, reducing indexing costs and best using disks. The second contribution proposes multi-threaded algorithms for both building and searching, harnessing the power of multi-core processors. Datasets for evaluation contain about 25 million images or over 8 billion SIFT descriptors. The third contribution addresses distributed computing. We adapt eCP to the MapReduce programming model and use the Hadoop framework and HDFS for our experiments. This time we evaluate eCP's ability to scale-up with a collection of 100 million images, more than 30 billion SIFT descriptors, and its ability to scale-out by running experiments on more than 100 machines.

### 6.2.2. Contributions in image indexing

**Participants:** Hervé Jégou, Giorgos Tolias.

*Partially in collaboration with Yannis Avrithis, National Technical University of Athens, Greece, Cai-Zhi Zhu and Shin'ichi Satoh, Japanese National Institute of Informatics, Japan.*

In [62], we have considered a framework and its associated family of metrics to compare images based on their local descriptors. It encompasses the VLAD descriptor and matching techniques such as Hamming embedding. Making the bridge between these approaches leads us to propose a match kernel that takes the best of existing techniques by combining an aggregation procedure with a selective match kernel. Finally, the representation underpinning this kernel is approximated, providing a large scale image search both precise and scalable, as shown by our experiments on several benchmarks. We give a Matlab package associated with the paper that allows to reproduce the results of the most interesting variant.

On the same topic, we propose in [78] a query expansion technique for image search that is faster and more precise than the existing ones. An enriched representation of the query is obtained by exploiting the binary representation offered by the Hamming embedding image matching approach: The initial local descriptors are refined by aggregating those of the database, while new descriptors are produced from the images that are deemed relevant. This approach 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 79 ms on a desktop machine. Overall, our technique significantly outperforms the visual query expansion state of the art on popular benchmarks. It is also the first query expansion technique shown effective on the UKB benchmark, which has few relevant images per query.

Finally, in [67] we considered a problem related to object retrieval, where we aim at retrieving, from a collection of images, all those in which a given query object appears. This problem is inherently asymmetric: the query object is mostly included in the database image, while the converse is not necessarily true. However, existing approaches mostly compare the images with symmetrical measures, without considering the different roles of query and database. This paper first measures the extent of asymmetry on large-scale public datasets reflecting this task. Considering the standard bag-of-words representation, we then propose new asymmetrical dissimilarities accounting for the different inlier ratios associated with query and database images. These asymmetrical measures depend on the query, yet they are compatible with an inverted file structure, without noticeably impacting search efficiency. Our experiments show the benefit of our approach, and show that the visual object retrieval task is better treated asymmetrically, in the spirit of state-of-the-art text retrieval.

### 6.2.3. Outlier detection applied to content-based image retrieval

**Participants:** Teddy Furon, Hervé Jégou.

The primary target of content based image retrieval is to return a list of images that are the most similar to a query image, which is usually done by ordering the images based on a similarity score. In most state-of-the-art systems, the magnitude of this score is very different from one query to another. This prevents us from making a proper decision about the correctness of the returned images. Our work [74] considers the applications where a confidence measurement is required, such as in copy detection or when a re-ranking stage is applied on a

short-list such as geometrical verification. For this purpose, we formulate image search as an outlier detection problem, and propose a framework derived from extreme values theory. We translate the raw similarity score returned by the system into a relevance score related to the probability that a raw score deviates from the estimated model of scores of random images. The method produces a relevance score which is normalized in the sense that it is more consistent across queries. Experiments performed on several popular image retrieval benchmarks and state-of-the-art image representations show the interest of our approach.

#### **6.2.4. Exploiting motion characteristics for action classification in videos**

**Participants:** Mihir Jain, Hervé Jégou.

*In collaboration with Patrick Bouthemy, Inria/Serpico, France.*

Several recent studies on action recognition have attested the importance of explicitly integrating motion characteristics in video description. In this work [43], we have re-visited the use of motion in videos, in order to better exploit it and improve action recognition systems. First, we established that adequately decomposing visual motion into dominant and residual motions, both in the extraction of the space-time trajectories and for the computation of descriptors, significantly improves action recognition algorithms. Then, we designed a new motion descriptor, the DCS descriptor, based on differential motion scalar quantities, divergence, curl and shear features. It captures additional information on the local motion patterns enhancing results. Finally, applying the recent VLAD coding technique proposed in image retrieval provides a substantial improvement for action recognition. Our three contributions are complementary and lead to significantly outperform all reported results on three challenging datasets, namely Hollywood 2, HMDB51 and Olympic Sports.

#### **6.2.5. Recognizing events in videos**

**Participant:** Hervé Jégou.

*In collaboration with Matthijs Douze, Jérôme Revaud and Cordelia Schmid, Inria/LEAR, France.*

We have addressed the problem of event retrieval for large-scale video collection. Given a video clip of a specific event, e.g., the wedding of Prince William and Kate Middleton, the goal is to retrieve other videos representing the same event from a dataset of over 100k videos.

Our first approach [55] encodes the frame descriptors of a video to jointly represent their appearance and temporal order. It exploits the properties of circulant matrices to compare the videos in the frequency domain. This offers a significant gain in complexity and accurately localizes the matching parts of videos. Furthermore, we extend product quantization to complex vectors in order to compress our descriptors, and to compare them in the compressed domain. Our method outperforms the state of the art both in search quality and query time on two large-scale video benchmarks for copy detection, Trecvid and CCweb. The evaluation has also been done on a new challenging dataset for event retrieval that we introduce: EVVE.

In a subsequent paper [39], we have made two other contributions to event retrieval in large collections of videos. First, we propose hyper-pooling strategies that encode the frame descriptors into a representation of the video sequence in a stable manner. Our best choices compare favorably with regular pooling techniques based on k-means quantization. Second, we introduce a technique to improve the ranking. It can be interpreted either as a query expansion method or as a similarity adaptation based on the local context of the query video descriptor. Experiments on public benchmarks show that our methods are complementary and improve event retrieval results, without sacrificing efficiency.

#### **6.2.6. Large-scale SVM image classification**

**Participants:** Thanh Nghi Doan, François Poulet.



Visual recognition remains an extremely challenging problem in computer vision research. Large datasets with millions images for thousands categories poses more challenges. We extend the state-of-the-art large scale linear classifier LIBLINEAR SVM and nonlinear classifier Power Mean SVM in two ways. The first one is to build a balanced bagging classifier with sampling strategy. The second one is to parallelize the training process of all binary classifiers with several multi-core computers [35]. We also applied the same approach to the stochastic gradient descent support vector machines (SVM-SGD) and to both state-of-the-art large linear classifier LIBLINEAR-CDBLOCK and nonlinear classifier Power Mean SVM in an incremental and parallel way [36].

### 6.2.7. Video copy detection with SNAP, a DNA indexing algorithm

**Participants:** Laurent Amsaleg, Guillaume Gravier.

*In collaboration with Leonardo S. De Oliveira, Zenilton Kleber G. Do Patrocínio Jr. and Silvio Jamil F. Guimarães, PUC Minas, Brazil.*

Near-duplicate video sequence identification consists in identifying real positions of a specific video clip in a video stream stored in a database. To address this problem, we proposed a new approach based on a scalable sequence aligner borrowed from proteomics [79]. Sequence alignment is performed on symbolic representations of features extracted from the input videos, based on an algorithm originally applied to bio-informatics. Experimental results demonstrate that our method performance achieved 94 % recall with 100 % precision, with an average searching time of about 1 second.

## 6.3. Security of multimedia contents and applications

### 6.3.1. Approximate nearest neighbors search with security and privacy requirements

**Participants:** Benjamin Mathon, Laurent Amsaleg, Teddy Furon.

*In collaboration with Julien Bringer, Morpho, France.*

This work presents a moderately secure but highly scalable and fast approximate nearest neighbors search. Our philosophy is to start from a state-of-the-art technique in this field based on approximate metrics: Euclidean distance based search in [47], [70], and cosine similarity based search in [42]. We then analyze the threats, and patch them avoiding as much as possible bricks penalizing too much the scalability and the speed. On the other hand, we do not completely prevent the players to infer some knowledge, but these limitations are well explained and experimentally assessed. The experimental body uses database of size much bigger than what the past secure solutions can handle.

### 6.3.2. A privacy-preserving framework for large-scale content-based information retrieval

**Participants:** Ewa Kijak, Laurent Amsaleg, Teddy Furon.

*In close cooperation with Stéphane Marchand-Maillet, Li Weng and April Morton, University of Geneva, Switzerland.*

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 instead of 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. Two different constructions of robust hash algorithms are used. One is based on random projections; the other is based on the discrete wavelet transform. Both algorithms exhibit satisfactory performance in comparison with state-of-the-art reference schemes. The results show that the privacy enhancement slightly improves the retrieval performance.

We consider the *majority voting attack* for estimating the query category and ID. Experiment results show that this attack is a threat when there are near-duplicates, but the success rate decreases with the number of omitted bits and the number of distinct items.

### 6.3.3. Privacy preserving data aggregation and service personalization using highly-scalable indexing techniques

**Participants:** Raghavendran Balu, Laurent Amsaleg, Hervé Jégou, Teddy Furon.

*In collaboration with Armen Aghasaryan, Dimitre Davidov and Makram Bouzid, Alcatel-Lucent, and Sébastien Gambs, Inria/CIDRE, in the framework of the Alcatel-Lucent / Inria common Lab.*

A challenging approach to the problem of privacy preserving data aggregation and service personalization has recently been proposed in Bell Labs, which introduces a privacy-preserving intermediation layer between end-users and service providers. It uses a distributed variant of a Locality Sensitive Hashing (LSH) techniques of doing scalable nearest-neighbor search, adapted in a novel way, to discover similar users while preserving their privacy. This approach faces however several important challenges that will be targeted in the scope of this collaboration. The challenges are:

- *LSH optimization:* Definitions of hash functions as well as various LSH parameters need to be automatically tuned in order to achieve a good quality of generated recommendations with an expected level of the procured user anonymity. An interesting issue is the possibility of supervised machine learning. If some public profiles are available, more efficient clustering methods boost the quality of the recommendation service but their levels of anonymity have never been assessed so far.
- *Irreversibility of anonymization:* This needs to be evaluated for different attack models, e.g. exploiting the knowledge of LSH hashing functions or any other publically available information on users. It is equivalent as being able to define the region of the super high-dimensional space mapped into the same hashing results. This attack is bound to fail as this region is too large to leak information. However, the prior knowledge about the sparseness of the profiles might drastically reduce this region, and hence weaken the privacy.
- *System dynamics:* Dealing with the cold-start problem or controlling the dynamics of a running system when the profiles and the cluster assignments evolve over the time is yet another challenge this approach is confronted with. If these temporal issues are well studied in conventional relational databases, no clear solution is efficient in the recommendation area, and a fortiori in privacy enhancing recommendation systems.

## 6.4. Structuring multimedia content and summarization

### 6.4.1. Stream labeling for TV Structuring

**Participants:** Vincent Claveau, Guillaume Gravier, Patrick Gros, Emmanuelle Martienne, Abir Ncibi.

In this application, we focus on the problem of labeling the segments of TV streams according to their types (eg. programs, commercial breaks, sponsoring...). During this year, following the work initiated in 2012, we have proposed an in-depth analysis of the use of conditional random fields (CRF) for our task [50]. Through several experiments conducted on real TV streams, we have shown that the CRF yields high results compared with state-of-the-art approaches. In particular, CRF offers several ways to efficiently take the sequencibility of our stream labeling problem into account. We also showed that it is robust when dealing with few training data or few features.

#### 6.4.2. *Statistical tests for repetition detection in TV streams*

**Participant:** Patrick Gros.

Detecting all repeated sequences in a TV stream is the first step of all techniques of TV stream structuring. We have improved our technique in several ways. First, a statistical hypothesis test with a corrected risk of Bonferroni was used to clean the repetitions of small sequences. Second, a content-based test is used to clean the remaining sequences, but also to complete the repeated sequences to their maximal length. One of our objective is to reduce the number of descriptor needed to achieve this task, given that this computation is the most expensive of the method. As a matter of fact, the method required computing the descriptors of 15.4 % of the images only.

#### 6.4.3. *Video summarization with constraint programming*

**Participants:** Mohamed-Haykel Boukadida, Patrick Gros.

*Joint work with Sid-Ahmed Berrani, Orange labs.*

Up to now, most video summarization methods are based on concepts like saliency and often use a single modality. In order to develop a more general framework, we propose to use a constraint programming approach, where summarizing a video is seen as a constraint resolution problem, which consists in choosing certain excerpts with respect to various criteria. This year we studied several ways to model the problem in order to gain a maximum flexibility in the summary. A first model was based on the selection of shots, the second one on the selection of parts of shots; The third one does not relies on shots and select image sequences directly. The challenge is to express the useful constraints with these models and the limited possibilities of the solver.

#### 6.4.4. *Transcript-free spoken content summarization using motif discovery*

**Participants:** Sébastien Champion, Guillaume Gravier.

*Joint work with Frédéric Bimbot and Nathan Souviraa-Labastié, Inria/PANAMA, France.*

Exploiting previous results on the unsupervised discovery of repeating words in speech signals, we proposed a method dedicated to transcript-free spoken content summarization. Extractive summarization is performed by selecting a small number of segments, typically one or two, which contains most of the repeated fragments [77]. Audio summaries were included in the Texmix demonstration and are currently being evaluated.

#### 6.4.5. *TV program structure discovery using grammatical inference*

**Participants:** Guillaume Gravier, Bingqing Qu.

*Joint work with Félicien Vallet and Jean Carrive, Institut National de l'Audiovisuel.*

Video structuring, in particular applied to TV programs which have strong editing structures, mostly relies on supervised approaches either to retrieve a known structure for which a model has been obtained or to detect key elements from which a known structure is inferred. We investigated an unsupervised approach to recurrent TV program structuring, exploiting the repetitiveness of key structural elements across episodes of the same show. We cast the problem of structure discovery as a grammatical inference problem and show that a suited symbolic representation can be obtained by filtering generic events based on their reoccurring property [92]. The method follows three steps: *i*) generic event detection, *ii*) selection of events relevant to the structure and *iii*) grammatical inference from a symbolic representation. Experimental evaluation is performed on three types of shows, viz., game shows, news and magazines, demonstrating that grammatical inference can be used to discover the structure of recurrent programs with very limited supervision.

#### 6.4.6. *Discovering and linking related images in large collections*

**Participants:** Guillaume Gravier, Hervé Jégou, Wanlei Zhao.

We have tackled the problem of image linking. One of the most successful method to link all similar images within a large collection is min-Hash, which is a way to significantly speed-up the comparison of images when the underlying image representation is bag-of-words. However, the quantization step of min-Hash introduces important information loss. In [66], we proposed a generalization of min-Hash, called Sim-min-Hash, to compare sets of real-valued vectors. We demonstrated the effectiveness of our approach when combined with the Hamming embedding similarity. Experiments on large-scale popular benchmarks demonstrated that Sim-min-Hash is more accurate and faster than min-Hash for similar image search. Linking a collection of one million images described by 2 billion local descriptors is done in 7 minutes on a single core machine.

## 6.5. Natural language processing in multimedia data

### 6.5.1. Text detection in videos

**Participants:** Khaoula Elagouni, Pascale Sébillot.

Texts embedded in multimedia documents often provide high level semantic clues that can be used in several applications or services. We thus aim at designing efficient Optical Character Recognition (OCR) systems able to recognize these texts. During the last three years, we have proposed three novel approaches, robust to text variability (different fonts, colors, sizes, etc.) and acquisition conditions (complex background, non-uniform lighting, low resolution, etc.). The first approach relies on a segmentation step and computes nonlinear separations between characters well adapted to the local morphology of images. The two other ones, called segmentation-free approaches, avoid the segmentation step by integrating a multi-scale scanning scheme: The first one relies on a graph model, while the second one uses a particular connectionist recurrent model able to handle spatial constraints between characters. In 2013, a precise evaluation and comparison between these approaches was conducted and published in [16].

### 6.5.2. Combining lexical cohesion and disruption for topic segmentation

**Participants:** Guillaume Gravier, Pascale Sébillot, Anca-Roxana Simon.

Topic segmentation classically relies on one of two criteria, either finding areas with coherent vocabulary use or detecting discontinuities. We proposed a segmentation criterion combining both lexical cohesion and disruption, enabling a trade-off between the two [58]. We provide the mathematical formulation of the criterion and an efficient graph based decoding algorithm for topic segmentation. Experimental results on standard textual data sets and on a more challenging corpus of automatically transcribed broadcast news shows demonstrate the benefit of such a combination. Gains were observed in all conditions, with segments of either regular or varying length and abrupt or smooth topic shifts. Long segments benefit more than short segments. However the algorithm has proven robust on automatic transcripts with short segments and limited vocabulary reoccurrences.

#### 6.5.2.1. Information extraction and text mining

**Participants:** Vincent Claveau, Marie Béatrice Arnulphy.

Following the work initiated in the previous period, we have kept on working on relation extraction. During this year, we have proposed a new prototype that still relies on a supervised machine learning approach but we now rely on the sequence built from the shortest syntactic path between the entities, as it is done in many studies. These paths of lemmas are then used in a kNN whose similarity score is based on language modeling techniques. Based on this new prototype, we have participated to several tracks of the BioNLP challenges concerning the automatic extraction of relations in a specialized corpus. Results obtained with this simple and non-domain specific technique were relatively good, with a second and fourth ranks among the participants for the two tasks concerned [26].

We also pursued previous work on supervised techniques for entity extraction and classification. Instead of working on complex machine learning approaches, we rather use simple methods but the focus is set on clever similarity computing between training examples and candidates for which we make the most of existing information retrieval techniques. Our approach has been evaluated through our participation to BioNLP-ST13 competition, where it has been ranked first [26].

We have also proposed unsupervised techniques for knowledge discovery, more precisely, to bring out coherent groups of entities. Existing techniques are usually based on clustering; the challenge is then to define a notion of similarity between the relevant entities. In this work, we have proposed to divert conditional random fields (CRF) in order to calculate indirectly the similarities among text sequences. Our approach consists in generating artificial labeling problems on the data to be processed to reveal regularities in the labeling of the entities. The good results obtained shows the validity of our approach [27] and opens many research avenues for other knowledge discovery tasks.

### 6.5.3. *Unsupervised approaches to fine-grained morphological analysis*

**Participants:** Vincent Claveau, Ewa Kijak.

Following the work initiated in the previous years, we have kept on studying fine-grained morphological analysis for biomedical information retrieval. In the biomedical field, the key to access information is the use of specialized terms (like *photochemotherapy*). These complex morphological structures may prevent a user querying for *gastrodynia* to retrieve texts containing *stomachalgia*. The original unsupervised technique proposed in 2012 has been further developed and tested. In particular, during this year, we have shown that it largely outperforms state-of-the-art tools (e.g., Morfessor and Derif) for morphological segmentation tasks. It also offers indirect morpho-lexical resources that are more reliable than hand-coded ones used in most state-of-the-art tools [11].

### 6.5.4. *Tree-structured named entities recognition*

**Participants:** Christian Raymond, Davy Weissenbacher.

Many natural language processing tasks needs the production of tree-structured outputs, like syntactic parsing, named entities recognition or language understanding. Currently, only machine learning based systems are robust enough to process the raw and noisy automatic transcribed speech while no machine learning paradigm are able to learn directly the tree structure in a reasonable time. In this work, we studied a solution to tackle the problem of predicting tree structured named entities from speech contents. We investigate a fast and robust decomposition strategy that was implemented and ranked best at the ETAPE NER evaluation campaign with results far better than those of the other participant systems [54].

### 6.5.5. *Fast machine learning algorithm for efficient combination of various features*

**Participant:** Christian Raymond.

Currently, in the field of natural language processing the machine learning algorithm "boosting over decision stumps" is often designed as the best off-the-shell classifier. It's actually widely used for his abilities to work on relatively big dataset, to operate intrinsically feature selection and to produce very good decision rules. We investigated a slight modification of this algorithm where the decision stumps are replaced by bonsai trees. Bonsai trees are small decision trees (with low depth) that can capture some structure in the data that decision stumps can not. This modification allows the boosting algorithm to exhibits better (or in the worst case similar) performances with a lower number of iteration the original algorithm needs. Thus allows in some cases a big improvement in term of performance for a lower cost in term of learning time. An application on image processing (typed/hand classification) exhibited interesting results in [94]

## 6.6. Competitions and international evaluation benchmarks

### 6.6.1. *FGcomp'2013, in conjunction with Imagenet*

**Participants:** Philippe-Henri Gosselin, Hervé Jégou.

*Joint participation with Naila Murray and Florent Perronnin, Xerox Research Center Europe.*

We have participated the the FGCOMP'2013 challenge and obtained the best results among all participants, see <http://sites.google.com/site/fgcomp2013> Although the proposed system follows most of the standard Fisher classification pipeline, we have evaluated and used several key features and good practices that improve the accuracy when specifically considering fine-grained classification tasks [75]. In particular, we consider the late fusion of two systems both based on Fisher vectors, but that employ drastically different design choices that make them very complementary. Moreover, we show that a simple yet effective filtering strategy significantly boosts the performance for several class domains. The method is described in a technical report.

### 6.6.2. *Hyperlink generation in broadcast videos*

**Participants:** Guillaume Gravier, Pascale Sébillot, Anca-Roxana Simon.

*Joint participation with Camille Guinaudeau, Heidelberg Institute of Technology (currently LIMSI-CNRS).*

Following up on our 2012 participation, we participated in the Search and hyperlinking task implemented in the framework of the Mediaeval 2013 benchmark initiative. We limited ourselves to hyperlink generation, building on research results in natural language processing, information retrieval and topic segmentation, focusing our contribution on the selection of precise target segments for hyperlinks.

### 6.6.3. *Maurdor campaign*

**Participant:** Christian Raymond.

*Joint participation with Yann Ricquebourg, Baptiste Poirriez, Aurélie Lemaitre and Bertrand Couïasnon, IRISA/Intuidoc.*

We are participating to the ongoing MAURDOR campaign <http://www.maurdor-campaign.org> which aims at evaluating systems for automatic processing of written documents. The contribution of TEXMEX comes from the machine learning system based on boosting over bonsai trees we implemented. In the context of this campaign, we investigate the usefulness of this algorithm to combine efficiently features on a relatively big dataset. The very first result shows that this system get state-of-the-art performance while it is much faster than traditional SVM approaches.

### 6.6.4. *Information extraction challenge at BioNLP-ST13*

**Participant:** Vincent Claveau.

BioNLP Shared Task is a community-wide effort to address fine-grained, structural information extraction from biomedical literature. This year, several tasks were proposed and 22 teams participated. TexMex has proposed runs for three main tasks concerning entity extraction and categorization, and relation extraction. The methods proposed by our team are based on machine learning and information retrieval components. Although they do not exploit specialized or domain-specific knowledge, we obtained good results and ranked first, first and third according to the tasks.

## WILLOW Project-Team

### 6. New Results

#### 6.1. 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.*

### 6.1.1. *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 accepted for publication to the ACM Transactions on Graphics (ACM ToG). The problem addressed in this work is illustrated in Figure 1 and example results are shown in figure 2. The pre-print is available online at [10].

### 6.1.2. *Learning and Calibrating Per-Location Classifiers for Visual Place Recognition*

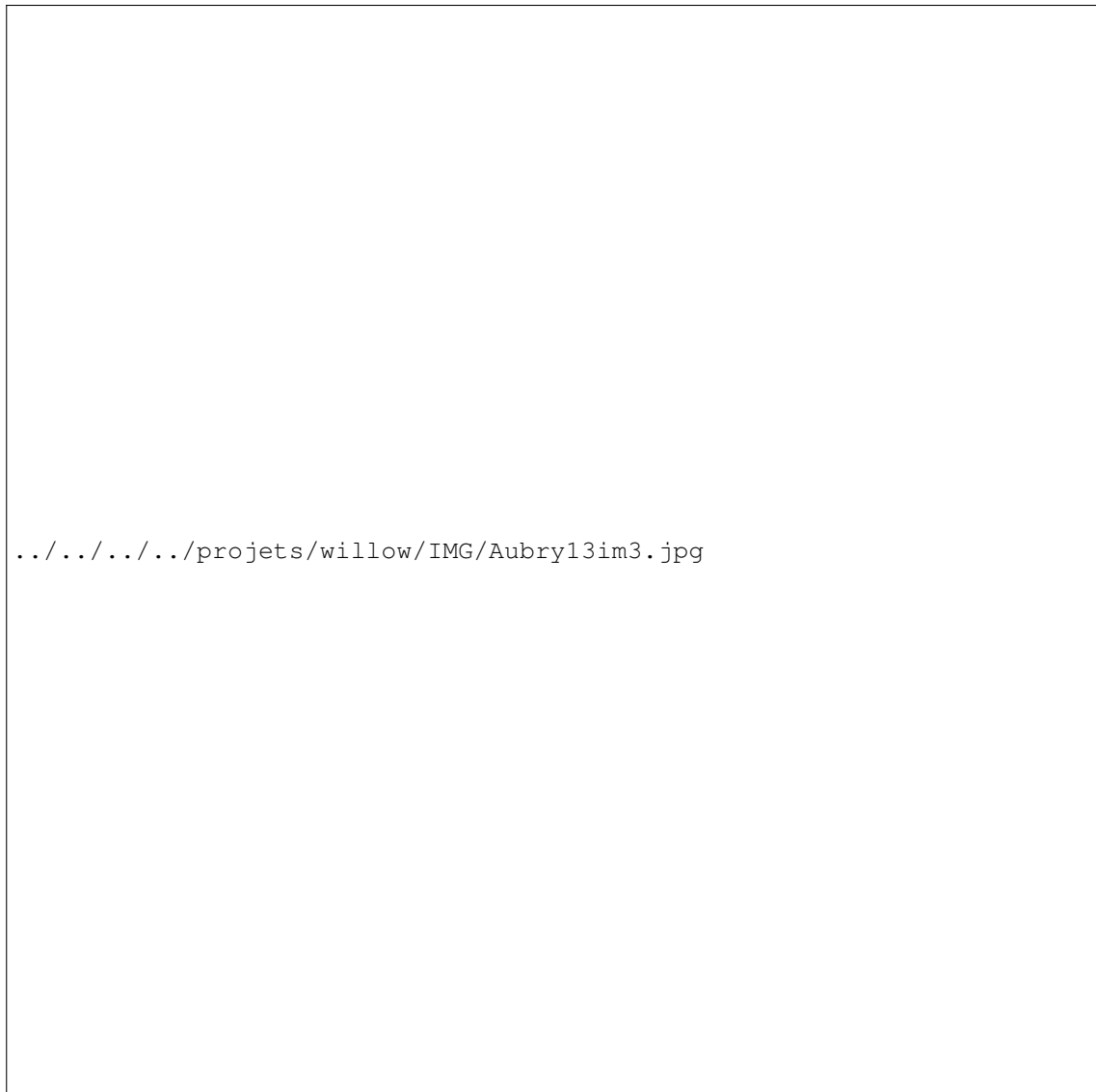
**Participants:** Petr Gronat, Josef Sivic, Guillaume Obozinski [ENPC / Inria SIERRA], Tomáš Pajdla [CTU in Prague].

The aim of this work is to localize a query photograph by finding other images depicting the same place in a large geotagged image database. This is a challenging task due to changes in viewpoint, imaging conditions and the large size of the image database. The contribution of this work is two-fold. First, we cast the place recognition problem as a classification task and use the available geotags to train a classifier for each location in the database in a similar manner to per-exemplar SVMs in object recognition. Second, as only few positive training examples are available for each location, we propose a new approach to calibrate all the per-location SVM classifiers using *only* the negative examples. The calibration we propose relies on a significance measure essentially equivalent to the p-values classically used in statistical hypothesis testing. Experiments are performed on a database of 25,000 geotagged street view images of Pittsburgh and demonstrate improved place recognition accuracy of the proposed approach over the previous work. This work has been published at CVPR 2013 [6].

### 6.1.3. *Visual Place Recognition with Repetitive Structures*

**Participants:** Akihiko Torii [Tokyo Institute of Technology], Josef Sivic, Tomáš Pajdla [CTU in Prague], Masatoshi Okutomi [Tokyo Institute of Technology].

Repeated structures such as building facades, fences or road markings often represent a significant challenge for place recognition. Repeated structures are notoriously hard for establishing correspondences using multi-view geometry. Even more importantly, they violate the feature independence assumed in the bag-of-visual-words representation which often leads to over-counting evidence and significant degradation of retrieval performance. In this work we show that repeated structures are not a nuisance but, when appropriately represented, they form an important distinguishing feature for many places. We describe a representation of repeated structures suitable for scalable retrieval. It is based on robust detection of repeated image structures and a simple modification of weights in the bag-of-visual-word model. Place recognition results are shown on datasets of street-level imagery from Pittsburgh and San Francisco demonstrating significant gains in recognition performance compared to the standard bag-of-visual-words baseline and more recently proposed burstiness weighting. This work has been published at CVPR 2013 [9].



*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.*



### 6.1.4. *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 submitted to CVPR 2014.

## 6.2. Category-level object and scene recognition

### 6.2.1. *Learning Graphs to Match*

**Participants:** Minsu Cho, Karteek Alahari, Jean Ponce.

Many tasks in computer vision are formulated as graph matching problems. Despite the NP-hard nature of the problem, fast and accurate approximations have led to significant progress in a wide range of applications. Learning graph models from observed data, however, still remains a challenging issue. This work presents an effective scheme to parameterize a graph model, and learn its structural attributes for visual object matching. For this, we propose a graph representation with histogram-based attributes, and optimize them to increase the matching accuracy. Experimental evaluations on synthetic and real image datasets demonstrate the effectiveness of our approach, and show significant improvement in matching accuracy over graphs with pre-defined structures. The work is illustrated in Figure 3. This work has been published ICCV 2013 [3].

### 6.2.2. *Finding Matches in a Haystack: A Max-Pooling Strategy for Graph Matching in the Presence of Outliers*

**Participants:** Minsu Cho, Olivier Duchenne [Intel], Jian Sun, Jean Ponce.

A major challenge in real-world 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 hard to distinguish inliers from outliers due to clutters. In this project, we propose a novel approach to graph matching, which is not only resilient to deformations but also remarkably tolerant to outliers. By adopting a max-pooling strategy within the graph matching framework, 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 submitted to CVPR 2014.

### 6.2.3. *Decomposing Bag of Words Histograms*

**Participants:** Ankit Gandhi [IIIT India], Karteek Alahari, C.v. Jawahar [IIIT India].

We aim to decompose a global histogram representation of an image into histograms of its associated objects and regions. This task is formulated as an optimization problem, given a set of linear classifiers, which can effectively discriminate the object categories present in the image. Our decomposition bypasses harder problems associated with accurately localizing and segmenting objects. We evaluate our method on a wide variety of composite histograms, and also compare it with MRF-based solutions. In addition to merely measuring the accuracy of decomposition, we also show the utility of the estimated object and background histograms for the task of image classification on the PASCAL VOC 2007 dataset. This work has been published at ICCV 2013 [5].

### 6.2.4. *Image Retrieval using Textual Cues*

**Participants:** Anand Mishra [IIIT India], Karteek Alahari, C.v. Jawahar [IIIT India].



*Figure 3. Graph learning for matching. Our approach learns a graph model from labeled data to provide the best match to instances of a target class. It shows significant improvement over previous approaches for matching. (Best viewed in color.)*

We present an approach for the text-to-image retrieval problem based on textual content present in images. Given the recent developments in understanding text in images, an appealing approach to address this problem is to localize and recognize the text, and then query the database, as in a text retrieval problem. We show that such an approach, despite being based on state-of-the-art methods, is insufficient, and propose a method, where we do not rely on an exact localization and recognition pipeline. We take a query-driven search approach, where we find approximate locations of characters in the text query, and then impose spatial constraints to generate a ranked list of images in the database. The retrieval performance is evaluated on public scene text datasets as well as three large datasets, namely IIT scene text retrieval, Sports-10K and TV series-1M, we introduce. This work has been published at ICCV 2013 [7].

### **6.2.5. Learning Discriminative Part Detectors for Image Classification and Cosegmentation**

**Participants:** Jian Sun, Jean Ponce.

In this work, we address the problem of learning discriminative part detectors from image sets with category labels. We propose a novel latent SVM model regularized by group sparsity to learn these part detectors. Starting from a large set of initial parts, the group sparsity regularizer forces the model to jointly select and optimize a set of discriminative part detectors in a max-margin framework. We propose a stochastic version of a proximal algorithm to solve the corresponding optimization problem. We apply the proposed method to image classification and cosegmentation, and quantitative experiments with standard benchmarks show that it matches or improves upon the state of the art. This work has been published at CVPR 2013 [8].

### **6.2.6. 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. The pre-print of this work is available online [11]. Results are illustrated in Figure 4.

### **6.2.7. 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.

We present an approach for the text-to-image retrieval problem based on textual content present in images. Given the recent developments in understanding text in images, an appealing approach to address this problem is to localize and recognize the text, and then query the database, as in a text retrieval problem. We show that such an approach, despite being based on state-of-the-art methods, is insufficient, and propose a method, where we do not rely on an exact localization and recognition pipeline. We take a query-driven search approach, where we find approximate locations of characters in the text query, and then impose spatial constraints to generate a ranked list of images in the database. The retrieval performance is evaluated on public scene text datasets as well as three large datasets, namely IIT scene text retrieval, Sports-10K and TV series-1M, we introduce. This work has been submitted to CVPR 2014.



*Figure 4. Recognition and localization results of our method for a Pascal VOC test image. Output maps are shown for six object categories with the highest responses.*

## 6.3. Image restoration, manipulation and enhancement

### 6.3.1. Learning to Estimate and Remove Non-uniform Image Blur

**Participants:** Florent Couzinie-Devy, Jian Sun, Karteek Alahari, Jean Ponce.

This work addresses the problem of restoring images subjected to unknown and spatially varying blur caused by defocus or linear (say, horizontal) motion. The estimation of the global (non-uniform) image blur is cast as a multi-label energy minimization problem. The energy is the sum of unary terms corresponding to learned local blur estimators, and binary ones corresponding to blur smoothness. Its global minimum is found using Ishikawa's method by exploiting the natural order of discretized blur values for linear motions and defocus. Once the blur has been estimated, the image is restored using a robust (non-uniform) deblurring algorithm based on sparse regularization with global image statistics. The proposed algorithm outputs both a segmentation of the image into uniform-blur layers and an estimate of the corresponding sharp image. We present qualitative results on real images, and use synthetic data to quantitatively compare our approach to the publicly available implementation of Chakrabarti et al. 2010. This work has been published at CVPR 2013 [4] and example results are shown in figure 5 .

### 6.3.2. Efficient, Blind, Spatially-Variant Deblurring for Shaken Images

**Participants:** Oliver Whyte [Microsoft Redmond], Josef Sivic, Andrew Zisserman, Jean Ponce.

In this chapter we discuss modeling and removing spatially-variant blur from photographs. We describe a compact global parameterization of camera shake blur, based on the 3D rotation of the camera during the exposure. Our model uses three-parameter homographies to connect camera motion to image motion and, by assigning weights to a set of these homographies, can be seen as a generalization of the standard, spatially-invariant convolutional model of image blur. As such we show how existing algorithms, designed for spatially-invariant deblurring, can be "upgraded" in a straightforward manner to handle spatially-variant blur instead. We demonstrate this with algorithms working on real images, showing results for blind estimation of blur parameters from single images, followed by non-blind image restoration using these parameters. Finally, we introduce an efficient approximation to the global model, which significantly reduces the computational cost of modeling the spatially-variant blur. By approximating the blur as locally-uniform, we can take advantage of fast Fourier-domain convolution and deconvolution, reducing the time required for blind deblurring by an order of magnitude.

This work has been accepted for publication as a book chapter in the upcoming book "Motion Deblurring: Algorithms and Systems" to be published by Cambridge University Press in May 2014. <sup>2</sup> The demo implementing deblurring of images degraded by camera shake is available online at: <http://www.di.ens.fr/willow/research/saturation/>.

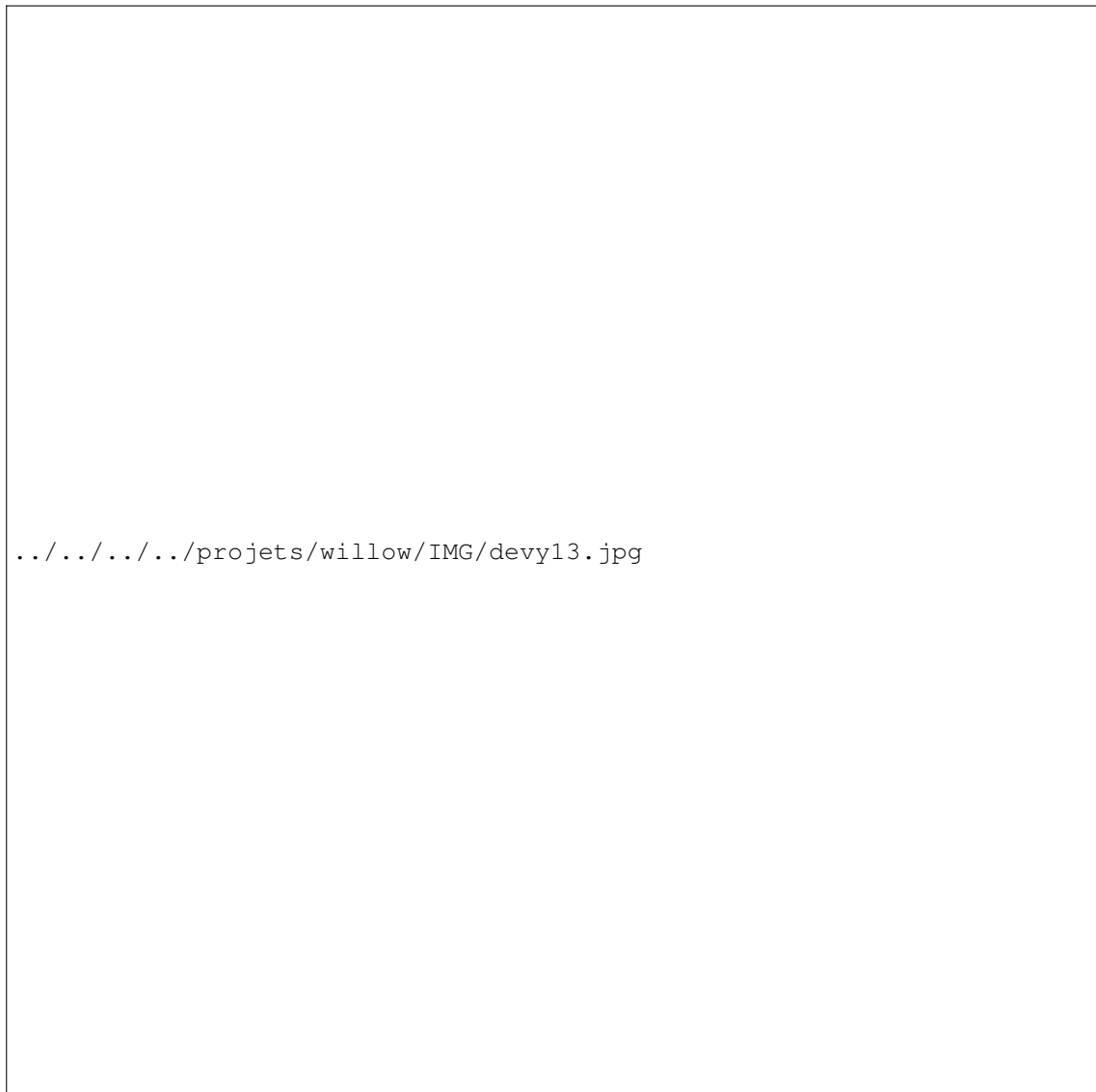
## 6.4. Human activity capture and classification

### 6.4.1. Layered Segmentation of People in Stereoscopic Movies

**Participants:** Karteek Alahari, Guillaume Seguin, Josef Sivic, Ivan Laptev.

In this work we seek to obtain a pixel-wise segmentation and pose estimation of multiple people in a stereoscopic video. This involves challenges such as dealing with unconstrained stereoscopic video, non-stationary cameras, and complex indoor and outdoor dynamic scenes. The contributions of our work are two-fold: First, we develop a segmentation model incorporating person detection, pose estimation, as well as colour, motion, and disparity cues. Our new model explicitly represents depth ordering and occlusion. Second, we introduce a stereoscopic dataset with frames extracted from feature-length movies "StreetDance 3D" and "Pina". The dataset contains 2727 realistic stereo pairs and includes annotation of human poses, person bounding boxes, and pixel-wise segmentations for hundreds of people. The dataset is composed of indoor and outdoor scenes depicting multiple people with frequent occlusions. We demonstrate results on our new challenging dataset, as well as on the H2view dataset from (Sheasby et al. ACCV 2012). This work has been published at ICCV 2013 [1].

<sup>2</sup><http://www.cambridge.org/fr/academic/subjects/engineering/image-processing-and-machine-vision/motion-deblurring-algorithms-and-systems>



*Figure 5. Sample deblurring results on real images. From left to right: blurry image, deblurred image, close-up corresponding to the boxes shown in red. Note that our estimated deblurred image has more detail.*

#### **6.4.2. Finding Actors and Actions in Movies**

**Participants:** Piotr Bojanowski, Francis Bach [Inria Sierra], Ivan Laptev, Jean Ponce, Cordelia Schmid [Inria Lear], Josef Sivic.

We address the problem of learning a joint model of actors and actions in movies using weak supervision provided by scripts. Specifically, we extract actor/action pairs from the script and use them as constraints in a discriminative clustering framework. The corresponding optimization problem is formulated as a quadratic program under linear constraints. People in video are represented by automatically extracted and tracked faces together with corresponding motion features. First, we apply the proposed framework to the task of learning names of characters in the movie and demonstrate significant improvements over previous methods used for this task. Second, we explore the joint actor/action constraint and show its advantage for weakly supervised action learning. We validate our method in the challenging setting of localizing and recognizing characters and their actions in feature length movies *Casablanca* and *American Beauty*. This work has been published at ICCV 2013 [2] and example results are shown in figure 6. The corresponding software has been also made publicly available (see the software section of this report).

#### **6.4.3. Highly-Efficient Video Features for Action Recognition and Counting**

**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 steadily improved over the recent years, the low speed of feature extraction remains to be a major bottleneck preventing current methods from addressing large-scale applications. In this work we demonstrate that local video features can be computed very efficiently by exploiting motion information readily-available from standard video compression schemes. We show experimentally that the use of sparse motion vectors provided by the video compression improves the speed of existing optical-flow based methods by two orders of magnitude while resulting in limited drops of recognition performance. Building on this representation, we next address the problem of event counting in video and present a method providing accurate counts of human actions and enabling to process 100 years of video on a modest computer cluster. This work has been submitted to CVPR 2014.



*Figure 6. Examples of automatically assigned names and actions in the movie Casablanca. Top row: Correct name and action assignments for tracks that have an actor/action constraint in the script. Bottom row: Correct name and action assignments for tracks that do not have a corresponding constraint in the script, but are still correctly classified. Note that even very infrequent characters are correctly classified (Annina and Yvonne). See more examples on the project web-page: <http://www.di.ens.fr/willow/research/actoraction/>*