



RESEARCH CENTER

FIELD

**Perception, Cognition and Interaction**

Activity Report 2016

**Section Scientific Foundations**

Edition: 2017-08-25



## DATA AND KNOWLEDGE REPRESENTATION AND PROCESSING

1. CEDAR Team	5
2. DAHU Project-Team	7
3. EXMO Project-Team	8
4. GRAPHIK Project-Team	10
5. LACODAM Team	12
6. LINKS Project-Team	17
7. MAGNET Project-Team	21
8. ORPAILLEUR Project-Team	26
9. SMIS Project-Team	28
10. TYREX Project-Team	31
11. WIMMICS Project-Team	33
12. ZENITH Project-Team	35

## INTERACTION AND VISUALIZATION

13. ALICE Project-Team	40
14. AVIZ Project-Team	42
15. EX-SITU Team	46
16. GRAPHDECO Project-Team	47
17. HYBRID Project-Team	50
18. ILDA Project-Team	53
19. IMAGINE Project-Team	57
20. MANAO Project-Team	59
21. MAVERICK Project-Team	67
22. MIMETIC Project-Team	70
23. MINT Project-Team	73
24. Mjolnir Team	75
25. POTIOC Project-Team	77
26. TITANE Project-Team	81

## LANGUAGE, SPEECH AND AUDIO

27. ALPAGE Project-Team	84
28. MULTISPEECH Project-Team	89
29. PANAMA Project-Team	93
30. SEMAGRAMME Project-Team	96

## ROBOTICS AND SMART ENVIRONMENTS

31. CHROMA Team	97
32. DEFROST Team	104
33. FLOWERS Project-Team	106
34. HEPHAISTOS Project-Team	110
35. LAGADIC Project-Team	113
36. LARSEN Team	116
37. PERVASIVE INTERACTION Team	120

38. RITS Project-Team .....	128
VISION, PERCEPTION AND MULTIMEDIA INTERPRETATION	
39. AYIN Team .....	137
40. LINKMEDIA Project-Team .....	139
41. MAGRIT Project-Team .....	141
42. MORPHEO Project-Team .....	144
43. PERCEPTION Project-Team .....	146
44. SIROCCO Project-Team .....	148
45. STARS Project-Team .....	151
46. THOTH Project-Team .....	157
47. WILLOW Project-Team .....	163

## CEDAR Team

### 3. Research Program

#### 3.1. Scalable Heterogeneous Stores

Big Data applications increasingly involve *diverse* data sources, such as: structured or unstructured documents, data graphs, relational databases etc. and it is often impractical to load (consolidate) diverse data sources in a single repository. Instead, interesting data sources need to be exploited “as they are”, with the added value of the data being realized especially through the ability to combine (join) together data from several sources. Systems capable of exploiting diverse Big Data in this fashion are usually termed *polystores*. A current limitation of polystores is that data stays captive of its original storage system, which may limit the data exploitation performance. We work to devise highly efficient storage systems for heterogeneous data across a variety of data stores.

#### 3.2. Semantic Query Answering

In the presence of data semantics, query evaluation techniques are insufficient as they only take into account the database, but do not provide the reasoning capabilities required in order to reflect the semantic knowledge. In contrast, (ontology-based) query answering takes into account both the data and the semantic knowledge in order to compute the full query answers, blending query evaluation and semantic reasoning.

We aim at designing efficient semantic query answering algorithms, both building on cost-based reformulation algorithms developed in the team and exploring new approaches mixing materialization and reformulation.

#### 3.3. Multi-Model Querying

As the world’s affairs get increasingly more digital, a large and varied set of data sources becomes available: they are either structured databases, such as government-gathered data (demographics, economics, taxes, elections, ...), legal records, stock quotes for specific companies, un-structured or semi-structured, including in particular graph data, sometimes endowed with semantics (see e.g. the Linked Open Data cloud). Modern data management applications, such as data journalism, are eager to combine in innovative ways both static and dynamic information coming from structured, semi-structured, and un-structured databases and social feeds. However, current content management tools for this task are not suited for the task, in particular when they require a lengthy rigid cycle of data integration and consolidation in a warehouse. Thus, we see a need for flexible tools allowing to interconnect various kinds of data sources and to query them together.

#### 3.4. Interactive Data Exploration at Scale

In the Big Data era we are faced with an increasing gap between the fast growth of data and the limited human ability to comprehend data. Consequently, there has been a growing demand of data management tools that can bridge this gap and help users retrieve high-value content from data more effectively. To respond to such user information needs, we aim to build interactive data exploration as a new database service, using an approach called “explore-by-example”.

#### 3.5. Exploratory Querying of Semantic Graphs

Semantic graphs including data and knowledge are hard to apprehend for users, due to the complexity of their structure and oftentimes to their large volumes. To help tame this complexity, in prior research (2014), we have presented a full framework for RDF data warehousing, specifically designed for heterogeneous and semantic-rich graphs. However, this framework still leaves to the users the burden of choosing the most interesting warehousing queries to ask. More user-friendly data management tools are needed, which help the user discover the interesting structure and information hidden within RDF graphs.

### **3.6. Representative Semantic Query Answering**

Top-k search is a classical topic, studied in relational databases, semantic web, recommendation systems,... It is extremely useful, among other, when a human user face a large number of query results, allowing the user to reformulate the query if necessary. However, we argue that top-k search incurs a bias on the perception of the set of results which is out of the control of the user. Our goal is to provide the user with k answers as well which are chosen so as to represent the diversity of the answer set. We will first consider this problem in the setting of relational or RDF databases. We will then extend to more heterogeneous sources, including in particular plain text.

## **DAHU Project-Team**

### **3. Research Program**

#### **3.1. Research Program**

Dahu aims at developing mechanisms for high-level specifications of systems built around DBMS, that are easy to understand while also facilitating verification of critical properties. This requires developing tools that are suitable for reasoning about systems that manipulate data. Some tools for specifying and reasoning about data have already been studied independently by the database community and by the verification community, with various motivations. However, this work is still in its infancy and needs to be further developed and unified.

Most current proposals for reasoning about DBMS over XML documents are based on tree automata, taking advantage of the tree structure of XML documents. For this reason, the Dahu team is studying a variety of tree automata. This ranges from restrictions of “classical” tree automata in order to understand their expressive power, to extensions of tree automata in order to understand how to incorporate the manipulation of data.

Moreover, Dahu is also interested in logical frameworks that explicitly refer to data. Such logical frameworks can be used as high level declarative languages for specifying integrity constraints, format change during data exchange, web service functionalities and so on. Moreover, the same logical frameworks can be used to express the critical properties we wish to verify.

In order to achieve its goals, Dahu brings together world-class expertise in both databases and verification.

## EXMO Project-Team

### 3. Research Program

#### 3.1. Knowledge representation semantics

We work with semantically defined knowledge representation languages (like description logics, conceptual graphs and object-based languages). Their semantics is usually defined within model theory initially developed for logics. The languages dedicated to the semantic web (RDF and OWL) follow that approach. RDF is a knowledge representation language dedicated to the description of resources; OWL is designed for expressing ontologies: it describes concepts and relations that can be used within RDF.

We consider a language  $L$  as a set of syntactically defined expressions (often inductively defined by applying constructors over other expressions). A representation ( $o \subseteq L$ ) is a set of such expressions. It is also called an ontology. An interpretation function ( $I$ ) is inductively defined over the structure of the language to a structure called interpretation domain ( $D$ ). This expresses the construction of the “meaning” of an expression in function of its components. A formula is satisfied by an interpretation if it fulfills a condition (in general being interpreted over a particular subset of the domain). A model of a set of expressions is an interpretation satisfying all these expressions. An expression ( $\delta$ ) is then a consequence of a set of expressions ( $o$ ) if it is satisfied by all of their models (noted  $o \models \delta$ ).

A computer must determine if a particular expression (taken as a query, for instance) is the consequence of a set of axioms (a knowledge base). For that purpose, it uses programs, called provers, that can be based on the processing of a set of inference rules, on the construction of models or on procedural programming. These programs are able to deduce theorems (noted  $o \vdash \delta$ ). They are said to be sound if they only find theorems which are indeed consequences and to be complete if they find all the consequences as theorems. However, depending on the language and its semantics, the decidability, i.e., the ability to create sound and complete provers, is not warranted. Even for decidable languages, the algorithmic complexity of provers may prohibit their exploitation.

To solve this problem a trade-off between the expressivity of the language and the complexity of its provers has to be found. These considerations have led to the definition of languages with limited complexity – like conceptual graphs and object-based representations – or of modular families of languages with associated modular prover algorithms – like description logics.

EXMO mainly considers languages with well-defined semantics (such as RDF and OWL that we contributed to define), and defines the semantics of some languages such as the SPARQL query language and alignment languages, in order to establish the properties of computer manipulations of the representations.

#### 3.2. Ontology matching and alignments

When different representations are used, it is necessary to identify their correspondences. This task is called ontology matching and its result is an alignment [4]. It can be described as follows: given two ontologies, each describing a set of discrete entities (which can be classes, properties, rules, predicates, etc.), find the relationships, if any, holding between these entities.

An alignment between two ontologies  $o$  and  $o'$  is a set of correspondences  $\langle e, e', r \rangle$  such that:

- $e$  and  $e'$  are the entities between which a relation is asserted by the correspondence, e.g., formulas, terms, classes, individuals;
- $r$  is the relation asserted to hold between  $e$  and  $e'$ . This relation can be any relation applying to these entities, e.g., equivalence, subsumption.

In addition, a correspondence may support various types of metadata, in particular measures of the confidence in a correspondence.



Given the semantics of the two ontologies provided by their consequence relation, we define an interpretation of two aligned ontologies as a pair of interpretations  $\langle m, m' \rangle$ , one for each ontology. Such a pair of interpretations is a model of the aligned ontologies  $o$  and  $o'$  if and only if each respective interpretation is a model of the ontology and they satisfy all correspondences of the alignment.

This definition is extended to networks of ontologies: a collection of ontologies and associated alignments. A model of such an ontology network is a tuple of local models such that each alignment is valid for the models involved in the tuple. In such a system, alignments play the role of model filters which select the local models that are compatible with all alignments. So, given an ontology network, it is possible to interpret it.

However, given a set of ontologies, it is necessary to find the alignments between them and the semantics does not tell which ones they are. Ontology matching aims at finding these alignments. A variety of methods is used for this task. They perform pairwise comparisons of entities from each of the ontologies and select the most similar pairs. Most matching algorithms provide correspondences between named entities, more rarely between compound terms. The relationships are generally equivalence between these entities. Some systems are able to provide subsumption relations as well as other relations in the support language (like incompatibility or instantiation). Confidence measures are usually given a value between 0 and 1 and are used for expressing preferences between two correspondences.

### 3.3. Data interlinking

Links are important for the publication of RDF data on the web. We call data interlinking the process of generating links identifying the same resource described in two data sets. Data interlinking parallels ontology matching: from two datasets  $(d$  and  $d')$  it generates a set of links (also called a link set,  $L$ ).

We have extended 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 introduced the notion of a link key [4], [1] which is a combination of such keys with alignments. More precisely, a link key is a structure  $\langle K^{eq}, K^{in}, C \rangle$  such that:

- $K^{eq}$  is a set of pairs of property expressions;
- $K^{in}$  is a set of pairs of property expressions;
- $C$  is a correspondence between classes.

Such a link key holds if and only if for any pair of resources belonging to the classes in correspondence such that the values of their property in  $K^{eq}$  are pairwise equal and the values of those in  $K^{in}$  pairwise intersect, the resources are the same.

As can be seen, link key validity is only relying on pairs of objects in two different data sets. We further qualify link keys as weak, plain and strong depending on them satisfying further constraints: a weak link key is only valid on pairs of individuals of different data sets, a plain link key has to apply in addition to pairs of individuals of the same data set as soon as one of them is identified with another individual of the other data set, a strong link key is a link key which is also a key for each data set, it can be thought of as a link key which is made of two keys.

Link keys can then be used for finding equal individuals across two data sets and generating the corresponding owl:sameAs links.

## GRAPHIK Project-Team

### 3. Research Program

#### 3.1. Logic-based Knowledge Representation and Reasoning

We follow the mainstream *logic-based* approach to the KR domain. First-order logic (FOL) is the reference logic in KR and most formalisms in this area can be translated into fragments (i.e., particular subsets) of FOL. This is in particular the case for description logics and existential rules, two well-known KR formalisms studied in the team.

A large part of research in this domain can be seen as studying the *trade-off* between the expressivity of languages and the complexity of (sound and complete) reasoning in these languages. The fundamental problem in KR languages is entailment checking: is a given piece of knowledge entailed by other pieces of knowledge, for instance from a knowledge base (KB)? Another important problem is *consistency* checking: is a set of knowledge pieces (for instance the knowledge base itself) consistent, i.e., is it sure that nothing absurd can be entailed from it? The *ontology-mediated query answering* problem is a topical problem (see Section 3.3). It asks for the set of answers to a query in the KB. In the case of Boolean queries (i.e., queries with a yes/no answer), it can be recast as entailment checking.

#### 3.2. Graph-based Knowledge Representation and Reasoning

Besides logical foundations, we are interested in KR formalisms that comply, or aim at complying with the following requirements: to have good *computational* properties and to allow users of knowledge-based systems to have a maximal *understanding and control* over each step of the knowledge base building process and use.

These two requirements are the core motivations for our graph-based approach to KR. We view labelled graphs as an *abstract representation* of knowledge that can be expressed in many KR languages (different kinds of conceptual graphs —historically our main focus—, the Semantic Web language RDF (Resource Description Framework), its extension RDFS (RDF Schema), expressive rules equivalent to the so-called tuple-generating-dependencies in databases, some description logics dedicated to query answering, etc.). For these languages, reasoning can be based on the structure of objects, thus based on graph-theoretic notions, while staying logically founded.

More precisely, our basic objects are labelled graphs (or hypergraphs) representing entities and relationships between these entities. These graphs have a natural translation in first-order logic. Our basic reasoning tool is graph homomorphism. The fundamental property is that graph homomorphism is sound and complete with respect to logical entailment *i.e.*, given two (labelled) graphs  $G$  and  $H$ , there is a homomorphism from  $G$  to  $H$  if and only if the formula assigned to  $G$  is entailed by the formula assigned to  $H$ . In other words, logical reasoning on these graphs can be performed by graph mechanisms. These knowledge constructs and the associated reasoning mechanisms can be extended (to represent rules for instance) while keeping this fundamental correspondence between graphs and logics.

#### 3.3. Ontology-Mediated Query Answering

Querying knowledge bases has become a central problem in knowledge representation and in databases. A knowledge base (KB) is classically composed of a terminological part (metadata, ontology) and an assertional part (facts, data). Queries are supposed to be at least as expressive as the basic queries in databases, i.e., conjunctive queries, which can be seen as existentially closed conjunctions of atoms or as labelled graphs. The challenge is to define good trade-offs between the expressivity of the ontological language and the complexity of querying data in presence of ontological knowledge. Description logics have been so far the prominent family of formalisms for representing and reasoning with ontological knowledge. However, classical description logics were not designed for efficient data querying. On the other hand, database languages are able to process complex queries on huge databases, but without taking the ontology into account. There is thus a need for new languages and mechanisms, able to cope with the ever growing size of knowledge bases in the Semantic Web or in scientific domains.

This problem is related to two other problems identified as fundamental in KR:

- *Query-answering with incomplete information.* Incomplete information means that it might be unknown whether a given assertion is true or false. Databases classically make the so-called closed-world assumption: every fact that cannot be retrieved or inferred from the base is assumed to be false. Knowledge bases classically make the open-world assumption: if something cannot be inferred from the base, and neither can its negation, then its truth status is unknown. The need of coping with incomplete information is a distinctive feature of querying knowledge bases with respect to querying classical databases (however, as explained above, this distinction tends to disappear). The presence of incomplete information makes the query answering task much more difficult.
- *Reasoning with rules.* Researching types of rules and adequate manners to process them is a mainstream topic in the Semantic Web, and, more generally a crucial issue for knowledge-based systems. For several years, we have been studying some rules, both in their logical and their graph form, which are syntactically very simple but also very expressive. These rules, known as existential rules or Datalog+, can be seen as an abstraction of ontological knowledge expressed in the main languages used in the context of KB querying. See Section 6.1 for details on the results obtained.

A problem generalizing the above described problems, and particularly relevant in the context of multiple data/metadata sources, is *querying hybrid knowledge bases*. In a hybrid knowledge base, each component may have its own formalism and its own reasoning mechanisms. There may be a common ontology shared by all components, or each component may have its own ontology, with mappings being defined among the ontologies. The question is what kind of interactions between these components and/or what limitations on the languages preserve the decidability of basic problems and if so, a “reasonable” complexity. Note that there are strong connections with the issue of data integration in databases.

### 3.4. Imperfect Information and Priorities

While classical FOL is the kernel of many KR languages, to solve real-world problems we often need to consider features that cannot be expressed purely (or not naturally) in classical logic. The logic- and graph-based formalisms used for previous points have thus to be extended with such features. The following requirements have been identified from scenarios in decision making in the agronomy domain.

1. to cope with vague and uncertain information and preferences in queries;
2. to cope with multi-granularity knowledge;
3. to take into account different and potentially conflicting viewpoints ;
4. to integrate decision notions (priorities, gravity, risk, benefit);
5. to integrate argumentation-based reasoning.

Although the solutions we develop need to be validated on the applications that motivated them, we also want them to be sufficiently generic to be applied in other contexts. One angle of attack (but not the only possible one) consists in increasing the expressivity of our core languages, while trying to preserve their essential combinatorial properties, so that algorithmic optimizations can be transferred to these extensions.

## LACODAM Team

### 3. Research Program

#### 3.1. Introduction

The three research axes of the Lacodam project-team are the following. First, we briefly introduce these axes, as well as their interplay:

- The first research axis is dedicated to the design of *novel pattern mining methods*. Pattern mining is one of the most important approaches to discover novel knowledge in data, and one of our strongest areas of expertise. Work in this axis will be the most fundamental of all three axes, and is expected to serve as foundations for work on the other two axes.
- The second axis tackles another aspect of knowledge discovery in data: the *interaction between the user and the system*, in order to co-discover novel knowledge. Our team has a long experience to collaborate with domain experts, and is thus especially aware of the need to improve such interaction.
- The third axis concerns *decision support*. With the help of methods from the two previous axes, our goal here is to design systems that can either help humans to take better decisions in precise applicative contexts, or to allow machines to automatically take relevant decisions in situations where extremely fast reaction time is required.

The following figure sums up the detailed work presented in the next few pages: on the sides are the three research axes of the team (X-axis) and our main applications areas (Y-axis). In the middle are colored squares that represent the precise research topics of the team that will be described in this section, placed relatively to their axis and main application area. Lines represent projects that can link several topics, and that are also connected to their main application area.

#### 3.2. Pattern mining algorithms

Twenty years of research in pattern mining have resulted in efficient approaches to handle the algorithmic complexity of the problem. Existing algorithms are now able to efficiently extract patterns with complex structures (ex: sequences, graphs, co-variations) from large datasets. However, when dealing with large, real world datasets, these methods still output a huge set of patterns, which is impractical for human analysis. This problem is called pattern explosion. The ongoing challenge of pattern mining research is to extract fewer but more meaningful patterns. The Lacodam team is committed to solve the pattern explosion problem following four research topics:

- the design of dedicated algorithms for mining temporal patterns
- the design of flexible pattern mining approaches
- the selection of interesting data mining results
- the design of parallel pattern algorithms to ensure scalability

The originality of our contributions relies on the exploration of knowledge-based approaches whose principle is to incorporate dedicated domain knowledge (aka application background knowledge) deep into the mining process. While most of the data mining approaches are based on agnostic approaches that are designed to cope with the pattern explosion, we propose to develop data mining techniques relying on knowledge-based artificial intelligence techniques. This covers the use of structured knowledge representations, as well as reasoning methods, in combination with mining.

The first approach concerns the classical approach of pattern mining which consists in using expert knowledge to define new pattern types (and related algorithms) that can solve applicative issues. In particular, we investigate how to handle temporality in pattern representations which turns out to be important in many real world applications (in particular for decision support) and deserves particular attention.

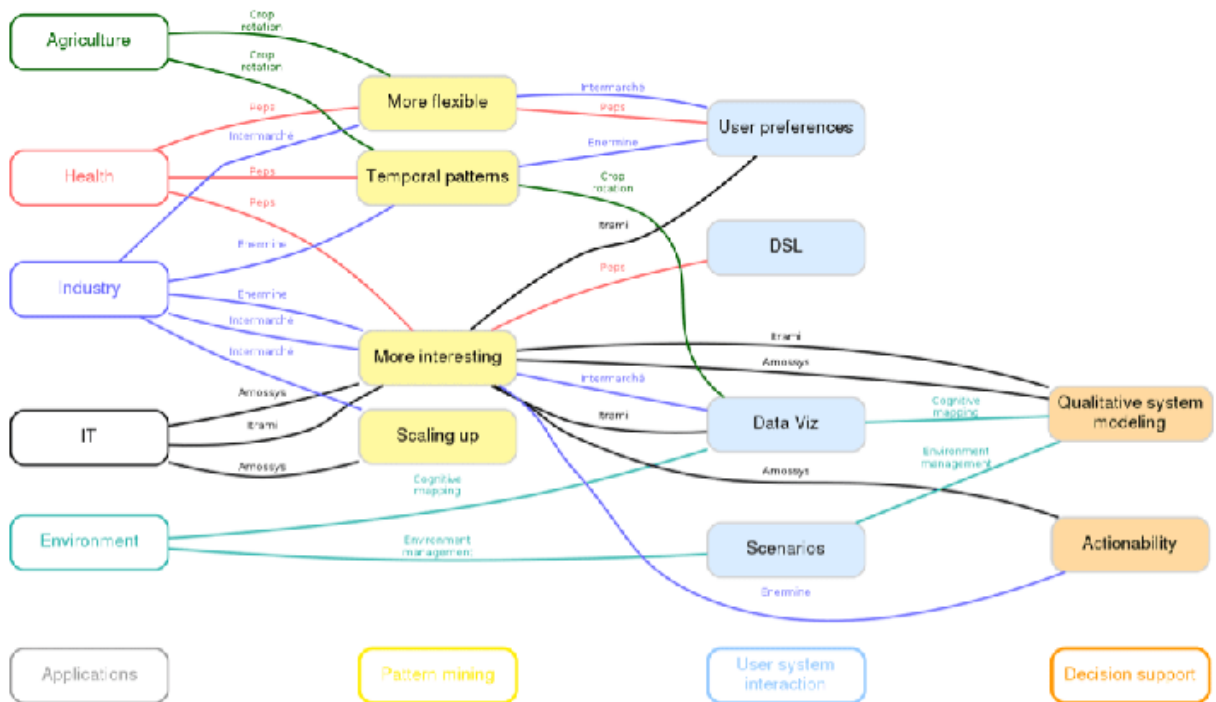


Figure 1. Lacadam research topics organized by axis and application

The two other approaches aim at proposing alternative pattern mining methods to let the user incorporate, by her own, knowledge that will help define her pattern domain of interest. Flexible pattern mining approaches enable analysts to easily incorporate extra knowledge, for example domain related constraints, in order to extract only the most relevant patterns. On the other hand, the selection of interesting data mining results aims at devising strategies to filter out the results that are useless for the data analyst. Beside the challenge related to algorithmic efficiency of such approaches, we are interested in formalizing the foundations of interestingness, according to background knowledge modeled with logic knowledge representation paradigms.

Last, pattern mining algorithms are computation-intensive, it is thus important to exploit all the available computing power. Parallelism is for a foreseeable future one of the main ways to speed up computations, and we have a strong competence on the design of parallel pattern mining algorithms. We will exploit this competence in order to guarantee that our approaches scale up to the real data provided by our partners.

### **3.3. User/system interaction**

As we pointed out before, there is a strong need to present relevant patterns to the user. This can be done by using more specific constraints, background knowledge and/or tailor-made optimization functions. Due to the difficulty of determining these elements beforehand, one of the most promising solutions is that the system and the user co-construct the definition of most relevant patterns, i.e., to have a human in the loop. This requires to have means to present intermediate results to the user, and to get user feedback in order to guide the search space exploration process in the right direction. This is an important research axis for Lacodam, which will be tackled in several complementary ways:

- **Domain Specific Languages:** one way to interact with the user is to propose a Domain Specific Language (DSL) tailored to the domain at hand and to the analysis tasks to perform. The challenge is to propose a DSL allowing the users to easily express the required processing workflows, to deploy those workflows for mining large volumes of data and to offer as much automation as possible.
- **What if / What for scenarios:** we are also investigating the use of scenarios to query results from data mining processes, as well as other complex processes such as complex system simulations or model predictions. Such scenarios are answers to questions of the type “what if [situation]” or “what [should be done] for [expected outcome]”.
- **User preferences:** in exploratory analysis, users often do not have a precise enough idea of what they want, and are not able to formulate such queries. Lacodam is thus investigating simple ways for letting users express their interests and preferences, either during the mining process to guide the search space exploration, or after, to help in getting the most relevant results.
- **Data visualization:** most of the research directions presented in this document require users to examine patterns at some point. The output of most pattern mining algorithms is simply a (long) list of patterns. While this presentation can be sufficient in some applications, it is often not enough to provide a complete understanding, especially for non-experts in pattern mining. A transversal research topic that we want to develop in Lacodam is to propose data visualization techniques adequate to understanding output results. Numerous (failed) experiments have shown that data mining and data visualization are fields which require distinct skills, where researchers in one field usually do not make significant advances in the other field (this is detailed in [Keim 2010]). Thus, our strategy is to establish collaborations with prominent data visualization teams for this line of research, with a long term goal to recruit a specialist in data visualization if the opportunity arises.

### **3.4. Decision support**

Patterns, especially predictive sequential patterns, resulting from mining a dataset have often a direct application in diagnosis. Lacodam inherits from the former Dream team a strong background in decision support systems, with an internationally recognized expertise in diagnosis. This AI subfield is concerned with determining if a system is operating normally or not, and if the system is in an abnormal state, to determine the cause of the faulty behavior. The considered system can as well be an agro- or eco-system, a software system or an animal or human being, as well.

The increasing volumes of data coming from a wide range of different systems (ex: sensor data from agro-environmental systems, log data from software systems, biological data coming from health monitoring systems) show that it is possible to gather more and more observations for such systems. Thus, it should be possible to exploit such observations to help human or software agents to take better decisions. Hence, while keeping the strong interest on decision support (and especially diagnosis) that existed in Dream, Lacodam adds the idea that the decision support systems should take advantage of the huge volumes of data available. This third and last research axis is thus a meeting point for all members of the team, as it requires to integrate AI techniques of traditional decision support systems with results from data mining techniques.

Two main research axes are investigated in Lacodam:

- **Diagnosis-based approaches.** We are exploring how to integrate knowledge found from pattern mining approaches, possibly with the help of interactive methods, into the qualitative models. The goal of such work is to partly automate the construction of the model, which can require a lot of human effort otherwise.
- **Actionable patterns and rules.** In many settings of “exploratory data mining”, the actual interest of a pattern is hard to assess, as it may be hard to measure or may be subjective (resulting from introducing the user in the mining process). However, there exist applications where once patterns are found, there are well defined measures to define what this pattern will bring to the user. Further, patterns and rules that can lead to actual actions beneficial to the user are called actionable patterns. Such actionable patterns and rules are especially important for industry.

### **3.5. Long-term goals**

The following perspectives are at the convergence of the three research axes presented before, and can be seen as the ideal towards which our efforts tend:

- **Automating data science workflow discovery.** The current methods for extracting knowledge from data and building decision support systems require a lot of human effort. Our three research axes aim at alleviating this effort, by devising methods that are more generic and by improving the interaction between the user and the system. An ideal solution would be that the user could forget completely about the existence of pattern mining or decision support methods. Instead the user would only loosely specify her problem, while the system would construct for her various data science / decision support workflows, possibly further refined via interactions.

We consider that this is a second order AI task, where AI techniques such as planning are used to explore the workflow search space, the workflow itself being composed of data mining and/or decision support components. This is a strategic evolution for data science endeavors, were the demand far exceeds the available human skilled manpower.

- **Logic argumentation based on epistemic interest.** Having increasingly automated approaches will require better and better ways to handle the interactions with the user. Our second long term goal is to explore the use of logic argumentation as an interaction tool between users and a data analysis tool. Alongside visualization and interactive data mining tools, it can be a way for users to query in an intuitive manner both the results and the way they were obtained. Such querying can also help the expert to reformulate her query in an interactive analysis setting.

This research direction continues the work on “epistemic interest” presented before. Its goal is to exploit principles of interactive data analysis in the context of epistemic interest measures. Logic argumentation [Besnard 2014] can be a natural tool for interactions between the user and the system: display of possibly exhaustive list of arguments, relationships – whether reinforcement, compatibility or conflict – between arguments, variable degrees of arguments, and possible solutions for argument conflicts.

The first step is to define a formal argumentation framework for explaining data mining results. This implies to continue theoretical work on the foundations of argumentation in order to identify the most adapted framework (either existing or a new one to be defined). Logic argumentation may be implemented and deeply explored in ASP, allowing us to build on our expertise in this logic language.

- Collaborative feedback and knowledge management. We are convinced that improving the data science process, and possibly automating it, will rely at some point in the near future on the vast feedback that can be obtained by communities of user seamlessly collaborating over the web. Consider for example what has been achieved by collaborative platforms such as StackOverflow: it has become the reference site for any programming question.

Data science is a more complex problem than programming, as in order to get help from the community, the user has to share her data and workflow, or at least some parts of them. This raises obvious privacy issues that may prevent this idea to succeed. As our research on automating the production of data science workflows should enable more people to have access to data science results, we are interested to investigate the design of collaborative platforms to exchange expert advices over data, workflows and analysis results, with an aim at exploiting this human feedback to improve the automated system with machine learning.



## **LINKS Project-Team**

### **3. Research Program**

#### **3.1. Background**

The main objective of LINKS is to develop methods for querying and managing linked data collections. Even though open linked data is the most prominent example, we will focus on hybrid linked data collections, which are collections of semi-structured datasets in hybrid formats: graph-based, RDF, relational, and NOSQL. The elements of these datasets may be linked, either by pointers or by additional relations between the elements of the different datasets, for instance the “same-as” or “member-of” relations as in RDF.

The advantage of traditional data models is that there exist powerful querying methods and technologies that one might want to preserve. In particular, they come with powerful schemas that constraint the possible manners in which knowledge is represented to a finite number of patterns. The exhaustiveness of these patterns is essential for writing of queries that cover all possible cases. Pattern violations are excluded by schema validation. In contrast, RDF schema languages such as RDFS can only enrich the relations of a dataset by new relations, which also helps for query writing, but which cannot constraint the number of possible patterns, so that they do not come with any reasonable notion of schema validation.

The main weakness of traditional formats, however, is that they do not scale to large data collections as stored on the Web, while the RDF data models scales well to very big collections such as linked open data. Therefore, our objective is to study mixed data collections, some of which may be in RDF format, in which we can lift the advantages of smaller datasets in traditional formats to much larger linked data collections. Such data collections are typically distributed over the internet, that some data sources have rigid query facilities that cannot be easily adapted or extended.

The main assumption that we impose in order to enable the logical approach, is that the given linked data collection must be correct in most dimensions. This means that all datasets are well-formed with respect to their available constraints and schemas, and clean with respect to the data values in most of the components of the relations in the datasets. One of the challenges is to integrate good quality RDF datasets into this setting, another is to clean the incorrect data in those dimensions that are less proper. It remains to be investigated in how far these assumptions can be maintained in realistic applications, and how much they can be weakened otherwise.

For querying linked data collections, the main problems are to resolve the heterogeneity of data formats and schemas, to understand the efficiency and expressiveness of recursive queries, that can follow links repeatedly, to answer queries under constraints, and to optimize query answering algorithms based on static analysis. When linked data is dynamically created, exchanged, or updated, the problems are how to process linked data incrementally, and how to manage linked data collections that change dynamically. In any case (static and dynamic) one needs to find appropriate schema mappings for linking semi-structured datasets. We will study how to automatize parts of this search process by developing symbolic machine learning techniques for linked data collections.

#### **3.2. Querying Heterogeneous Linked Data**

Our main objective is to query collections of linked datasets. In the static setting, we consider two kinds of links: explicit links between elements of the datasets, such as equalities or pointers, and logical links between relations of different datasets such as schema mappings. In the dynamic setting, we permit a third kind of links that point to “intentional” relations computable from a description, such as the application of a Web service or the application of a schema mapping.

We believe that collections of linked datasets are usually too big to ensure a global knowledge of all datasets. Therefore, schema mappings and constraints should remain between pairs of datasets. Our main goal is to be able to pose a query on a collection of datasets, while accounting for the possible recursive effects of schema mappings. For illustration, consider a ring of datasets  $D_1, D_2, D_3$  linked by schema mappings  $M_1, M_2, M_3$  that tell us how to complete a database  $D_i$  by new elements from the next database in the cycle.

The mappings  $M_i$  induce three intentional datasets  $I_1, I_2$ , and  $I_3$ , such that  $I_i$  contains all elements from  $D_i$  and all elements implied by  $M_i$  from the next intentional dataset in the ring:

$$I_1 = D_1 \cup M_1(I_2), \quad I_2 = D_2 \cup M_2(I_3), \quad I_3 = D_3 \cup M_3(I_1)$$

Clearly, the global information collected by the intentional datasets depends recursively on all three original datasets  $D_i$ . Queries to the global information can now be specified as standard queries to the intentional databases  $I_i$ . However, we will never materialize the intentional databases  $I_i$ . Instead, we can rewrite queries on one of the intentional datasets  $I_i$  to recursive queries on the union of the original datasets  $D_1, D_2$ , and  $D_3$  with their links and relations. Therefore, a query answering algorithm is needed for recursive queries, that chases the “links” between the  $D_i$  in order to compute the part of  $I_i$  needed for the purpose of query answering.

This illustrates that we must account for the graph data models when dealing with linked data collections whose elements are linked, and that query languages for such graphs must provide recursion in order to chase links. Therefore, we will have to study graph databases with recursive queries, such as RDF graphs with SPARQL queries, but also other classes of graph databases and queries.

We study schemas and mappings between datasets with different kinds of data models and the complexity of evaluating recursive queries over graphs. In order to use schema mapping for efficiently querying the different datasets, we need to optimize the queries by taking into account the mappings. Therefore, we will study static analysis of schema mappings and recursive queries. Finally, we develop concrete applications in which our fundamental techniques can be applied.

### 3.3. Managing Dynamic Linked Data

With the quick growth of the information technology on the Web, more and more Web data gets created dynamically every day, for instance by smartphones, industrial machines, users of social networks, and all kinds of sensors. Therefore, large amounts of dynamic data need to be exchanged and managed by various data-centric web services, such as online shops, online newspapers, and social networks.

Dynamic data is often created by the application of some kind of service on the Web. This kind of data is intentional in the same spirit as the intentional data specified by the application of a schema mapping, or the application of some query to the hidden Web. Therefore, we will consider a third kind of links in the dynamic setting, that map to intentional data specified by whatever kind of function application. Such a function can be defined in data-centric programming languages, in the style of Active XML, XSLT, and NOSQL languages.

The dynamicity of data adds a further dimension to the challenges for linked data collections that we described before, while all the difficulties remain valid. One of the new aspects is that intentional data may be produced incrementally, as for instance when exchanged over data streams. Therefore, one needs incremental algorithms able to evaluate queries on incomplete linked data collections, that are extended or updated incrementally. Note that incremental data may be produced without end, such as a Twitter stream, so that one cannot wait for its completion. Instead, one needs to query and manage dynamic data with as low latency as possible. Furthermore, all static analysis problems are to be re-investigated in the presence of dynamic data.

Another aspect of dynamic data is distribution over the Web, and thus parallel processing as in the cloud. This raises the typical problems coming with data distribution: huge data sources cannot be moved without very high costs, while data must be replicated for providing efficient parallel access. This makes it difficult, if not impossible, to update replicated data consistently. Therefore, the consistency assumption has been removed by NOSQL databases for instance, while parallel algorithmic is limited to naive parallelisation (i.e. map/reduce) where only few data needs to be exchanged.

We will investigate incremental query evaluation for distributed data-centered programming languages for linked data collections, dynamic updates as needed for linked data management, and static analysis for linked data workflows.

### 3.4. Linking Graphs

When datasets from independent sources are not linked with existing schema mappings, we would like to investigate symbolic machine learning solutions for inferring such mappings in order to define meaningful links between data from separate sources. This problem can be studied for various kinds of linked data collections. Before presenting the precise objectives, we will illustrate our approach on the example of linking data in two independent graphs: an address book of a research institute containing detailed personnel information and a (global) bibliographic database containing information on papers and their authors.

We remind that a schema allows to identify a collection of types each grouping objects from the same semantic class e.g., the collection of all persons in the address book and the collection of all authors in the bibliography database. As a schema is often lacking or underspecified in graph data models, we intend to investigate inference methods based on structural similarity of graph fragments used to describe objects from the same class in a given document e.g., in the bibliographic database every author has a name and a number of affiliations, while a paper has a title and a number of authors. Furthermore, our inference methods will attempt to identify, for every type, a set of possible keys, where by key we understand a collection of attributes of an object that uniquely identifies such an object in its semantic class. For instance, for a person in the address book two examples of a key are the name of the person and the office phone number of that person.

In the next step, we plan to investigate employing existing entity linkage solutions to identify pairs of types from different databases whose instances should be linked using compatible keys. For instance, persons in the address book should be linked with authors in the bibliographical database using the name as the compatible key. Linking the same objects (represented in different ways) in two databases can be viewed as an instance of a mapping between the two databases. Such mapping is, however, discriminatory because it typically maps objects from a specific subset of objects of given types. For instance, the mapping implied by linking persons in the address book with authors in the bibliographic database involves in fact researchers, a subgroup of personnel of the research institute, and authors affiliated with the research institute. Naturally, a subset of objects of a given type, or a subtype, can be viewed as a result of a query on the set of all objects, which on very basic level illustrates how learning data mappings can be reduced to learning queries.

While basic mappings link objects of the same type, more general mappings define how the same type of information is represented in two different databases. For instance, the email address and the postal address of an individual may be represented in one way in the address book and in another way in the bibliographic databases, and naturally, the query asking for the email address and the postal address of a person identified by a given name will differ from one database to the other. While queries used in the context of linking objects of compatible types are essentially unary, queries used in the context of linking information are  $n$ -ary and we plan to approach inference of general database mappings by investigating and employing algorithms for inference of  $n$ -ary queries.

An important goal in this research is elaborating a formal definition of *learnability* (feasibility of inference) of a given class of concepts (schemas of queries). We plan to following the example of Gold (1967), which requires not only the existence of an efficient algorithm that infers concepts consistent with the given input but the ability to infer every concept from the given class with a sufficiently informative input. Naturally, learnability depends on two parameters. The first parameter is the class of concepts i.e., a class of schema and

a class of queries, from which the goal concept is to be inferred. The second parameter is the type of input that an inference algorithm is given. This can be a set of examples of a concept e.g., instances of RDF databases for which we wish to construct a schema or a selection of nodes that a goal query is to select. Alternatively, a more general interactive scenario can be used where the learning algorithm inquires the user about the goal concept e.g., by asking to indicate whether a given node is to be selected or not (as membership queries of Angluin (1987) ). In general, the richer the input is, the richer class of concepts can be handled, however, the richer class of queries is to be handled, the higher computational cost is to be expected. The primary task is to find a good compromise and identify classes of concepts that are of high practical value, allow efficient inference with possibly simple type of input.

The main open problem for graph-shaped data studied by Links are how to infer queries, schemas, and schema-mappings for graph-structured data.

## **MAGNET Project-Team**

### **3. Research Program**

#### **3.1. Introduction**

The main objective of MAGNET is to develop original machine learning methods for networked data in order to build applications like browsing, monitoring and recommender systems, and more broadly information extraction in information networks. We consider information networks in which the data consist of both feature vectors and texts. We model such networks as (multiple) (hyper)graphs wherein nodes correspond to entities (documents, spans of text, users, ...) and edges correspond to relations between entities (similarity, answer, co-authoring, friendship, ...). Our main research goal is to propose new on-line and batch learning algorithms for various problems (node classification / clustering, link classification / prediction) which exploit the relationships between data entities and, overall, the graph topology. We are also interested in searching for the best hidden graph structure to be generated for solving a given learning task. Our research will be based on generative models for graphs, on machine learning for graphs and on machine learning for texts. The challenges are the dimensionality of the input space, possibly the dimensionality of the output space, the high level of dependencies between the data, the inherent ambiguity of textual data and the limited amount of human labeling. An additional challenge will be to design scalable methods for large information networks. Hence, we will explore how sampling, randomization and active learning can be leveraged to improve the scalability of the proposed algorithms.

Our research program is organized according to the following questions:

1. How to go beyond vectorial classification models in Natural Language Processing (NLP) tasks?
2. How to adaptively build graphs with respect to the given tasks? How to create networks from observations of information diffusion processes?
3. How to design methods able to achieve a good trade-off between predictive accuracy and computational complexity?
4. How to go beyond strict node homophilic/similarity assumptions in graph-based learning methods?

#### **3.2. Beyond Vectorial Models for NLP**

One of our overall research objectives is to derive graph-based machine learning algorithms for natural language and text information extraction tasks. This section discusses the motivations behind the use of graph-based ML approaches for these tasks, the main challenges associated with it, as well as some concrete projects. Some of the challenges go beyond NLP problems and will be further developed in the next sections. An interesting aspect of the project is that we anticipate some important cross-fertilizations between NLP and ML graph-based techniques, with NLP not only benefiting from but also pushing ML graph-based approaches into new directions.

Motivations for resorting to graph-based algorithms for texts are at least threefold. First, online texts are organized in networks. With the advent of the web, and the development of forums, blogs, and micro-blogging, and other forms of social media, text productions have become strongly connected. Interestingly, NLP research has been rather slow in coming to terms with this situation, and most of the literature still focus on document-based or sentence-based predictions (wherein inter-document or inter-sentence structure is not exploited). Furthermore, several multi-document tasks exist in NLP (such as multi-document summarization and cross-document coreference resolution), but most existing work typically ignore document boundaries and simply apply a document-based approach, therefore failing to take advantage of the multi-document dimension [40], [43].

A second motivation comes from the fact that most (if not all) NLP problems can be naturally conceived as graph problems. Thus, NLP tasks often involve discovering a relational structure over a set of text spans (words, phrases, clauses, sentences, etc.). Furthermore, the *input* of numerous NLP tasks is also a graph; indeed, most end-to-end NLP systems are conceived as pipelines wherein the output of one processor is in the input of the next. For instance, several tasks take POS tagged sequences or dependency trees as input. But this structured input is often converted to a vectorial form, which inevitably involves a loss of information.

Finally, graph-based representations and learning methods appear to address some core problems faced by NLP, such as the fact that textual data are typically not independent and identically distributed, they often live on a manifold, they involve very high dimensionality, and their annotations is costly and scarce. As such, graph-based methods represent an interesting alternative to, or at least complement, structured prediction methods (such as CRFs or structured SVMs) commonly used within NLP. Graph-based methods, like label propagation, have also been shown to be very effective in semi-supervised settings, and have already given some positive results on a few NLP tasks [22], [45].

Given the above motivations, our first line of research will be to investigate how one can leverage an underlying network structure (e.g., hyperlinks, user links) between documents, or text spans in general, to enhance prediction performance for several NLP tasks. We think that a “network effect”, similar to the one that took place in Information Retrieval (with the Page Rank algorithm), could also positively impact NLP research. A few recent papers have already opened the way, for instance in attempting to exploit Twitter follower graph to improve sentiment classification [44].

Part of the challenge here will be to investigate how adequately and efficiently one can model these problems as instances of more general graph-based problems, such as node clustering/classification or link prediction discussed in the next sections. In a few cases, like text classification or sentiment analysis, graph modeling appears to be straightforward: nodes correspond to texts (and potentially users), and edges are given by relationships like hyperlinks, co-authorship, friendship, or thread membership. Unfortunately, modeling NLP problems as networks is not always that obvious. From the one hand, the right level of representation will probably vary depending on the task at hand: the nodes will be sentences, phrases, words, etc. From the other hand, the underlying graph will typically not be given a priori, which in turn raises the question of how we construct it. A preliminary discussion of the issue of optimal graph construction for semi-supervised learning in NLP is given in [22], [48]. We identify the issue of adaptive graph construction as an important scientific challenge for machine learning on graphs in general, and we will discuss it further in Section 3.3 .

As noted above, many NLP tasks have been recast as structured prediction problems, allowing to capture (some of the) output dependencies. How to best combine structured output and graph-based ML approaches is another challenge that we intend to address. We will initially investigate this question within a semi-supervised context, concentrating on graph regularization and graph propagation methods. Within such approaches, labels are typically binary or in a small finite set. Our objective is to explore how one propagates an exponential number of *structured labels* (like a sequence of tags or a dependency tree) through graphs. Recent attempts at blending structured output models with graph-based models are investigated in [45], [33]. Another related question that we will address in this context is how does one learn with *partial labels* (like partially specified tag sequence or tree) and use the graph structure to complete the output structure. This last question is very relevant to NLP problems where human annotations are costly; being able to learn from partial annotations could therefore allow for more targeted annotations and in turn reduced costs [35].

The NLP tasks we will mostly focus on are coreference resolution and entity linking, temporal structure prediction, and discourse parsing. These tasks will be envisioned in both document and cross-document settings, although we expect to exploit inter-document links either way. Choices for these particular tasks is guided by the fact that they are still open problems for the NLP community, they potentially have a high impact for industrial applications (like information retrieval, question answering, etc.), and we already have some expertise on these tasks in the team (see for instance [34], [30], [32]). As a midterm goal, we also plan to work on tasks more directly relating to micro-blogging, such sentiment analysis and the automatic thread structuring of technical forums; the latter task is in fact an instance of rhetorical structure prediction [47].

We have already initiated some work on the coreference resolution with graph-based learning, by casting the problem as an instance of spectral clustering [32].

### 3.3. Adaptive Graph Construction

In most applications, edge weights are computed through a complex data modeling process and convey crucially important information for classifying nodes, making it possible to infer information related to each data sample even exploiting the graph topology solely. In fact, a widespread approach to several classification problems is to represent the data through an undirected weighted graph in which edge weights quantify the similarity between data points. This technique for coding input data has been applied to several domains, including classification of genomic data [42], face recognition [31], and text categorization [36].

In some cases, the full adjacency matrix is generated by employing suitable similarity functions chosen through a deep understanding of the problem structure. For example for the TF-IDF representation of documents, the affinity between pairs of samples is often estimated through the cosine measure or the  $\chi^2$  distance. After the generation of the full adjacency matrix, the second phase for obtaining the final graph consists in an edge sparsification/reweighting operation. Some of the edges of the clique obtained in the first step are pruned and the remaining ones can be reweighted to meet the specific requirements of the given classification problem. Constructing a graph with these methods obviously entails various kinds of loss of information. However, in problems like node classification, the use of graphs generated from several datasets can lead to an improvement in accuracy ([49], [23], [24]). Hence, the transformation of a dataset into a graph may, at least in some cases, partially remove various kinds of irregularities present in the original datasets, while keeping some of the most useful information for classifying the data samples. Moreover, it is often possible to accomplish classification tasks on the obtained graph using a running time remarkably lower than is needed by algorithms exploiting the initial datasets, and a suitable sparse graph representation can be seen as a compressed version of the original data. This holds even when input data are provided in an online/stream fashion, so that the resulting graph evolves over time.

In this project we will address the problem of adaptive graph construction towards several directions. The first one is about how to choose the best similarity measure given the objective learning task. This question is related to the question of metric and similarity learning ([25], [26]) which has not been considered in the context of graph-based learning. In the context of structured prediction, we will develop approaches where output structures are organized in graphs whose similarity is given by top- $k$  outcomes of greedy algorithms.

A different way we envision adaptive graph construction is in the context of semi-supervised learning. Partial supervision can take various forms and an interesting and original setting is governed by two currently studied applications: detection of brain anomaly from connectome data and polls recommendation in marketing. Indeed, for these two applications, a partial knowledge of the information diffusion process can be observed while the network is unknown or only partially known. An objective is to construct (or complete) the network structure from some local diffusion information. The problem can be formalized as a graph construction problem from partially observed diffusion processes. It has been studied very recently in [38]. In our case, the originality comes either from the existence of different sources of observations or from the large impact of node contents in the network.

We will study how to combine graphs defined by networked data and graphs built from flat data to solve a given task. This is of major importance for information networks because, as said above, we will have to deal with multiple relations between entities (texts, spans of texts, ...) and also use textual data and vectorial data.

### 3.4. Prediction on Graphs and Scalability

As stated in the previous sections, graphs as complex objects provide a rich representation of data. Often enough the data is only partially available and the graph representation is very helpful in predicting the unobserved elements. We are interested in problems where the complete structure of the graph needs to be recovered and only a fraction of the links is observed. The link prediction problem falls into this category. We are also interested in the recommendation and link classification problems which can be seen as graphs

where the structure is complete but some labels on the links (weights or signs) are missing. Finally we are also interested in labeling the nodes of the graph, with class or cluster memberships or with a real value, provided that we have (some information about) the labels for some of the nodes.

The semi-supervised framework will be also considered. A midterm research plan is to study how graph regularization models help for structured prediction problems. This question will be studied in the context of NLP tasks, as noted in Section 3.2, but we also plan to develop original machine learning algorithms that have a more general applicability. Inputs are networks whose nodes (texts) have to be labeled by structures. We assume that structures lie in some manifold and we want to study how labels can propagate in the network. One approach is to find a smooth labeling function corresponding to an harmonic function on both manifolds in input and output.

Scalability is one of the main issues in the design of new prediction algorithms working on networked data. It has gained more and more importance in recent years, because of the growing size of the most popular networked data that are now used by millions of people. In such contexts, learning algorithms whose computational complexity scales quadratically, or slower, in the number of considered data objects (usually nodes or edges, depending on the task) should be considered impractical.

These observations lead to the idea of using graph sparsification techniques in order to work on a part of the original network for getting results that can be easily extended and used for the whole original input. A sparsified version of the original graph can often be seen as a subset of the initial input, i.e. a suitably selected input subgraph which forms the training set (or, more in general, it is included in the training set). This holds even for the active setting. A simple example could be to find a spanning tree of the input graph, possibly using randomization techniques, with properties such that we are allowed to obtain interesting results for the initial graph dataset. We have started to explore this research direction for instance in [46].

At the level of mathematical foundations, the key issue to be addressed in the study of (large-scale) random networks also concerns the segmentation of network data into sets of independent and identically distributed observations. If we identify the data sample with the whole network, as it has been done in previous approaches [37], we typically end up with a set of observations (such as nodes or edges) which are highly interdependent and hence overly violate the classic i.i.d. assumption. In this case, the data scale can be so large and the range of correlations can be so wide, that the cost of taking into account the whole data and their dependencies is typically prohibitive. On the contrary, if we focus instead on a set of subgraphs independently drawn from a (virtually infinite) target network, we come up with a set of independent and identically distributed observations—namely the subgraphs themselves, where subgraph sampling is the underlying ergodic process [28]. Such an approach is one principled direction for giving novel statistical foundations to random network modeling. At the same time, because one shifts the focus from the whole network to a set of subgraphs, complexity issues can be restricted to the number of subgraphs and their size. The latter quantities can be controlled much more easily than the overall network size and dependence relationships, thus allowing to tackle scalability challenges through a radically redesigned approach.

Another way to tackle scalability problems is to exploit the inherent decentralized nature of very large graphs. Indeed, in many situations very large graphs are the abstract view of the digital activities of a very large set of users equipped with their own device. Nowadays, smartphones, tablets and even sensors have storage and computation power and gather a lot of data that serve to analytics, prediction, suggestion and personalized recommendation. Gathering all user data in large data centers is costly because it requires oversized infrastructures with huge energy consumption and large bandwidth networks. Even though cloud architectures can optimize such infrastructures, data concentration is also prone to security leaks, lost of privacy and data governance for end users. The alternative we have started to develop in Magnet is to devise decentralized, private and personalized machine learning algorithms so that they can be deployed in the personal devices. The key challenges are therefore to learn in a collaborative way in a network of learners and to preserve privacy and control on personal data.

### **3.5. Beyond Homophilic Relationships**



In many cases, algorithms for solving node classification problems are driven by the following assumption: linked entities tend to be assigned to the same class. This assumption, in the context of social networks, is known as homophily ([29], [39]) and involves ties of every type, including friendship, work, marriage, age, gender, and so on. In social networks, homophily naturally implies that a set of individuals can be parted into subpopulations that are more cohesive. In fact, the presence of homogeneous groups sharing common interests is a key reason for affinity among interconnected individuals, which suggests that, in spite of its simplicity, this principle turns out to be very powerful for node classification problems in general networks.

Recently, however, researchers have started to consider networked data where connections may also carry a negative meaning. For instance, disapproval or distrust in social networks, negative endorsements on the Web. Although the introduction of signs on graph edges appears like a small change from standard weighted graphs, the resulting mathematical model, called signed graphs, has an unexpectedly rich additional complexity. For example, their spectral properties, which essentially all sophisticated node classification algorithms rely on, are different and less known than those of graphs. Signed graphs naturally lead to a specific inference problem that we have discussed in previous sections: link classification. This is the problem of predicting signs of links in a given graph. In online social networks, this may be viewed as a form of sentiment analysis, since we would like to semantically categorize the relationships between individuals.

Another way to go beyond homophily between entities will be studied using our recent model of hypergraphs with bipartite hyperedges [41]. A bipartite hyperedge connects two ends which are disjoint subsets of nodes. Bipartite hyperedges is a way to relate two collections of (possibly heterogeneous) entities represented by nodes. In the NLP setting, while hyperedges can be used to model bags of words, bipartite hyperedges are associated with relationships between bags of words. But each end of bipartite hyperedges is also a way to represent complex entities, gathering several attribute values (nodes) into hyperedges viewed as records. Our hypergraph notion naturally extends directed and undirected weighted graph. We have defined a spectral theory for this new class of hypergraphs and opened a way to smooth labeling on sets of nodes. The weighting scheme allows to weigh the participation of each node to the relationship modeled by bipartite hyperedges accordingly to an equilibrium condition. This condition provides a competition between nodes in hyperedges and allows interesting modeling properties that go beyond homophily and similarity over nodes (the theoretical analysis of our hypergraphs exhibits tight relationships with signed graphs). Following this competition idea, bipartite hyperedges are like matches between two teams and examples of applications are team creation. The basic tasks we are interested in are hyperedge classification, hyperedge prediction, node weight prediction. Finally, hypergraphs also represent a way to summarize or compress large graphs in which there exists highly connected couples of (large) subsets of nodes.

---

## ORPAILLEUR Project-Team

### 3. Research Program

#### 3.1. Knowledge Discovery guided by Domain Knowledge

**Keywords:** knowledge discovery in databases, knowledge discovery in databases guided by domain knowledge, data mining formal concept analysis, classification, pattern mining second-order Hidden Markov Models

Knowledge discovery in databases (KDD) is aimed at discovering patterns in large databases. These patterns can then be interpreted as knowledge units to be reused in knowledge systems. From an operational point of view, the KDD process is based on three main steps: (i) selection and preparation of the data, (ii) data mining, (iii) interpretation of the discovered patterns. The KDD process –as implemented in the Orpailleur team– is based on data mining methods which are either symbolic or numerical. Symbolic methods are based on pattern mining (e.g. mining frequent itemsets, association rules, sequences...), Formal Concept Analysis (FCA [80]) and extensions of FCA such as Pattern Structures [83] and Relational Concept Analysis (RCA [90]). Numerical methods are based on probabilistic approaches such as second-order Hidden Markov Models (HMM [85]), which are well adapted to the mining of temporal and spatial data.

Domain knowledge, when available, can improve and guide the KDD process, materializing the idea of *Knowledge Discovery guided by Domain Knowledge* or KDDK. In KDDK, domain knowledge plays a role at each step of KDD: the discovered patterns can be interpreted as knowledge units and reused for problem-solving activities in knowledge systems, implementing the operational sequence “mining, interpreting (modeling), representing, and reasoning”. In this way, knowledge discovery appears as a core task in knowledge engineering, with an impact in various semantic activities, e.g. information retrieval, recommendation and ontology engineering. Usual application domains for the team include agronomy, astronomy, biology, chemistry, and medicine.

One main operation in the research work of Orpailleur on KDDK is *classification*, which is a polymorphic process involved in modeling, mining, representing, and reasoning tasks. Classification problems can be formalized by means of a class of objects (or individuals), a class of attributes (or properties), and a binary correspondence between the two classes, indicating for each individual-property pair whether the property applies to the individual or not. The properties may be features that are present or absent, or the values of a property that have been transformed into binary variables. Formal Concept Analysis (FCA) relies on the analysis of such binary tables and may be considered as a symbolic data mining technique to be used for extracting a set of formal concepts then organized within a concept lattice [80] (concept lattices are also known as “Galois lattices” [68]).

In parallel, the search for frequent itemsets and the extraction of association rules are well-known symbolic data mining methods, related to FCA (actually searching for frequent itemsets can be understood as traversing a concept lattice). Both processes usually produce a large number of items and rules, leading to the associated problems of “mining the sets of extracted items and rules”. Some subsets of itemsets, e.g. frequent closed itemsets (FCIs), allow finding interesting subsets of association rules, e.g. informative association rules. This explains why several algorithms are needed for mining data depending on specific applications [92].

#### 3.2. Text Mining

**Keywords:** text mining, knowledge discovery from collection of texts, annotation, ontology engineering from texts

The objective of a text mining process is to extract useful knowledge units from large collections of texts [78]. The text mining process shows specific characteristics due to the fact that texts are complex objects written in natural language. The information in a text is expressed in an informal way, following linguistic rules, making text mining a particular task. To avoid information dispersion, a text mining process has to take into account –as much as possible– paraphrases, ambiguities, specialized vocabulary and terminology. This is why the preparation of texts for text mining is usually dependent on linguistic resources and methods.

From a KDDK perspective, text mining is aimed at extracting “interesting units” (nouns and relations) from texts with the help of domain knowledge encoded within an ontology (also useful for text annotation). Text mining is especially useful in the context of semantic web for ontology engineering. In the Orpailleur team, the focus is put on the mining of real-world texts in application domains such as biology and medicine, using mainly symbolic data mining methods, and especially Formal Concept Analysis. Accordingly, the text mining process may be involved in a loop used to enrich and to extend linguistic resources. In turn, linguistic and ontological resources can be exploited to guide a “knowledge-based text mining process”.

### 3.3. Knowledge Systems and Web of Data

**Keywords:** knowledge engineering, web of data, semantic web, ontology, description logics, classification-based reasoning, case-based reasoning, information retrieval

The web of data constitutes a good platform for experimenting ideas on knowledge engineering and knowledge discovery, in relation with the principles of semantic web. A software agent may be able to read, understand, and manipulate information on the web, if and only if the knowledge necessary for achieving those tasks is available: this is why domain knowledge and ontologies are of main importance. The knowledge representation language recommended by W3C to design ontologies and knowledge bases is OWL, which is based on description logics (DLs [65]). In OWL, knowledge units are represented by classes (DL concepts) having properties (DL roles) and instances. Concepts are organized within a partial order based on a subsumption relation, and the inference services are based on classification-based reasoning and case-based reasoning (CBR).

Actually, there are many interconnections between concept lattices in FCA and ontologies, e.g. the partial order underlying an ontology can be supported by a concept lattice. Moreover, a pair of implications within a concept lattice can be adapted for designing concept definitions in ontologies. Accordingly, we are interested here in two main challenges: how the web of data, as a set of potential knowledge sources (e.g. DBpedia, Wikipedia, Yago, Freebase...) can be mined for helping the design of definitions and knowledge bases and how knowledge discovery techniques can be applied for providing a better usage of the web of data (e.g. LOD classification).

Accordingly, a part of the research work in Knowledge Engineering is oriented towards knowledge discovery in the web of data, as, with the increased interest in machine processable data, more and more data is now published in RDF (Resource Description Framework) format. Particularly, we are interested in the completeness of the data and their potential to provide concept definitions in terms of necessary and sufficient conditions [66]. We have proposed a novel technique based on FCA which allows data exploration as well as the discovery of definition (bidirectional implication rules).

## SMIS Project-Team

### 3. Research Program

#### 3.1. Embedded Data Management

The challenge tackled in this research action is twofold: (1) to design embedded database techniques matching the hardware constraints of (current and future) smart objects and (2) to set up co-design rules helping hardware manufacturers to calibrate their future platforms to match the requirements of data driven applications. While a large body of work has been conducted on data management techniques for high-end servers (storage, indexation and query optimization models minimizing the I/O bottleneck, parallel DBMS, main memory DBMS, etc.), less research efforts have been placed on embedded database techniques. Light versions of popular DBMS have been designed for powerful handheld devices; yet DBMS vendors have never addressed the complex problem of embedding database components into chips. Proposals dedicated to databases embedded on chip usually consider small databases, stored in the non-volatile memory of the microcontroller –hundreds of kilobytes– and rely on NOR Flash or EEPROM technologies. Conversely, SMIS is pioneering the combination of microcontrollers and NAND Flash constraints to manage Gigabyte(s) size embedded databases. We present below the positioning of SMIS with respect to international teams conducting research on topics which may be connected to the addressed problem, namely work on electronic stable storage, RAM consumption and specific hardware platforms.

Major database teams are investigating data management issues related to hardware advances (EPFL: A. Ailamaki, CWI: M. Kersten, U. Of Wisconsin: J. M. Patel, Columbia: K. Ross, UCSB: A. El Abbadi, IBM Almaden: C. Mohan, etc.). While there are obvious links with our research on embedded databases, these teams target high-end computers and do not consider highly constrained architectures with non traditional hardware resources balance. At the other extreme, sensors (ultra-light computing devices) are considered by several research teams (e.g., UC Berkeley: D. Culler, ITU: P. Bonnet, Johns Hopkins University: A. Terzis, MIT: S. Madden, etc.). The focus is on the processing of continuous streams of collected data. Although the devices we consider share some hardware constraints with sensors, the objectives of both environments strongly diverge in terms of data cardinality and complexity, query complexity and data confidentiality requirements. Several teams are looking at efficient indexes on flash (HP LABS: G. Graefe, U. Minnesota: B. Debnath, U. Massachusetts: Y. Diao, Microsoft: S. Nath, etc.). Some studies try to minimize the RAM consumption, but the considered RAM/stable storage ratio is quite large compared to the constraints of the embedded context. Finally, a large number of teams have focused on the impact of flash memory on database system design (we presented an exhaustive state of the art in a VLDB tutorial [34]). The work conducted in the SMIS team on bi-modal flash devices takes the opposite direction, proposing to influence the design of flash devices by the expression of database requirements instead of running after the constantly evolving flash device technology.

#### 3.2. Access and Usage Control Models

Access control management has been deeply studied for decades. Different models have been proposed to declare and administer access control policies, like DAC, MAC, RBAC, TMAC, and OrBAC. While access control management is well established, new models are being defined to cope with privacy requirements. Privacy management distinguishes itself from traditional access control in the sense that the data to be protected is personal. Hence, the user's consent must be reflected in the access control policies, as well as the usage of the data, its collection rules and its retention period, which are principles safeguarded by law and must be controlled carefully.

The research community working on privacy models is broad, and involves many teams worldwide including in France ENST-B, LIRIS, Inria LICIT, and LRI, and at the international level IBM Almaden, Purdue Univ., Politecnico di Milano and Univ. of Milano, George Mason Univ., Univ. of Massachusetts, Univ. of Texas and Colorado State Univ. to cite a few. Pioneer attempts towards privacy wary systems include the P3P Platform for Privacy Preservation [36] and Hippocratic databases [29]. In the last years, many other policy languages have been proposed for different application scenarios, including EPAL [40], XACML [39] and WSPL [32]. Hippocratic databases are inspired by the axiom that databases should be responsible for the privacy preservation of the data they manage. The architecture of a Hippocratic database is based on ten guiding principles derived from privacy laws.

The trend worldwide has been to propose enhanced access control policies to capture finer behavior and bridge the gap with privacy policies. To cite a few, Ardagna *et al.* (Univ. Milano) enables actions to be performed after data collection (like notification or removal), purpose binding features have been studied by Lefevre *et al.* (IBM Almaden), and Ni *et al.* (Purdue Univ.) have proposed obligations and have extended the widely used RBAC model to support privacy policies.

The positioning of the SMIS team within this broad area is rather (1) to focus on intuitive or automatic tools helping the individual to control some facets of her privacy (e.g., data retention, minimal collection) instead of increasing the expressiveness but also the complexity of privacy models and (2) to push concrete models enriched by real-case (e.g., medical) scenarios and by a joint work with researchers in Law.

### 3.3. Tamper-resistant Data Management

Tamper-resistance refers to the capacity of a system to defeat confidentiality and integrity attacks. This problem is complementary to access control management while being (mostly) orthogonal to the way access control policies are defined. Security surveys regularly point out the vulnerability of database servers against external (i.e., by intruders) and internal (i.e., by employees) attacks. Several attempts have been made in commercial DBMSs to strengthen server-based security, e.g., by separating the duty between DBA and DSA (Data Security Administrator), by encrypting the database footprint and by securing the cryptographic material using Hardware Security Modules (HSM) [35]. To face internal attacks, client-based security approaches have been investigated where the data is stored encrypted on the server and is decrypted only on the client side. Several contributions have been made in this direction, notably by U. of California Irvine (S. Mehrotra, Database Service Provider model), IBM Almaden (R. Agrawal, computation on encrypted data), U. of Milano (E. Damiani, encryption schemes), Purdue U. (E. Bertino, XML secure publication), U. of Washington (D. Suci, provisional access) to cite a few seminal works. An alternative, recently promoted by Stony Brook Univ. (R. Sion), is to augment the security of the server by associating it with a tamper-resistant hardware module in charge of the security aspects. Contrary to traditional HSM, this module takes part in the query computation and performs all data decryption operations. SMIS investigates another direction based on the use of a tamper-resistant hardware module on the client side. Most of our contributions in this area are based on exploiting the tamper-resistance of secure tokens to build new data protection schemes.

While our work on Privacy-Preserving data Publishing (PPDP) is still related to tamper-resistance, a complementary positioning is required for this specific topic. The primary goal of PPDP is to anonymize/sanitize microdata sets before publishing them to serve statistical analysis purposes. PPDP (and privacy in databases in general) is a hot topic since 2000, when it was introduced by IBM Research (IBM Almaden: R. Agrawal, IBM Watson: C.C. Aggarwal), and many teams, mostly north American universities or research centres, study this topic (e.g., PORTIA DB-Privacy project regrouping universities such as Stanford with H. Garcia-Molina). Much effort has been devoted by the scientific community to the definition of privacy models exhibiting better privacy guarantees or better utility or a balance of both (such as differential privacy studied by C. Dwork: Microsoft Research or D. Kifer: Penn-State Univ and J. Gehrke: Cornell Univ) and thorough surveys exist that provide a large overview of existing PPDP models and mechanisms [37]. These works are however orthogonal to our approach in that they make the hypothesis of a trustworthy central server that can execute the anonymization process. In our work, this is not the case. We consider an architecture composed of a large population of tamper-resistant devices weakly connected to an untrusted infrastructure and study how to compute

PPDP problems in this context [1]. Hence, our work has some connections with the works done on Privacy Preserving Data Collection (Stevens Institute of Tech. / Rutgers Univ,NJ: R.N.Wright, Univ Austin Texas: V. Shmatikov), on Secure Multi-party Computing for Privacy Preserving Data Mining (Rutgers Univ: J. Vaidya, Purdue Univ: C. Clifton) and on distributed PPDP algorithms (Univ Wisconsin: D. DeWitt, Univ Michigan: K. Lefevre, Rutgers Univ: J. Vaidya, Purdue Univ: C. Clifton) while none of them share the same architectural hypothesis as us.

## **TYREX Project-Team**

### **3. Research Program**

#### **3.1. Modeling**

Modeling consists in capturing various aspects of document and data processing and communication in a unifying model. Our modeling research direction mainly focuses on three aspects.

The first aspect aims at reducing the impedance mismatch. The impedance mismatch refers to the complexity, difficulty and lack of performance induced by various web application layers which require the same piece of information to be represented and processed differently. The mismatch occurs because programming languages use different native data models than those used for documents in browsers and for storage in databases. This results in complex and multi-tier software architectures whose different layers are incompatible in nature. This, in turn, results in expensive, inefficient, and error-prone web development. For reducing the impedance mismatch, we will focus on the design of a unifying software stack and programming framework, backed by generic and solid logical foundations similar in spirit to the NoSQL approach.

The second aspect aims at harnessing heterogeneity. Web applications increasingly use diverse data models: ordered and unordered tree-like structures (such as XML), nested records and arrays (such as JSON), graphs (like RDF), and tables. Furthermore, these data models involve a variety of languages for expressing constraints over data (e.g. XML schema, RelaxNG, and RDFS to name just a few). We believe that this heterogeneity is here to stay and is likely to increase. These differences in representations imply loads of error-prone and costly conversions and transformations. Furthermore, some native formats (e.g. JSON) are repurposed from an internal representation to a format for data exchange. This often results in a loss of information and in errors that need to be tracked and corrected. In this context, it is important to seek methods for reducing risks of information loss during data transformation and exchange. For harnessing heterogeneity, we will focus on the integration of data models through unified formal semantics and in particular logical interpretation. This allows using the same programming language constructs on different data models. At the programming language level, this is similar to languages such as JSoNq for JSON and XML.

Finally, the third aspect aims at making applications and data more compositional. Most web programming technologies are currently limited from a compositional point of view. For example, tree grammars (like schema languages for XML) are monolithic in the sense that they require the full description of the considered structures, instead of allowing the assembly of smaller and reusable building blocks. As a consequence, this translates into monolithic web applications, which makes their automated verification harder by making modular analyses more difficult. The need for compositionality is illustrated in the industry by the increasing development of fragmented W3C specifications organised in ad-hoc modules. For making applications and data more compositional, we will focus on the design of modular schema and programming languages. For this purpose, we will notably rely on succinct yet expressive formalisms (like two-way logics, polymorphic types, session types) that ease the process of expressing modular specifications.

#### **3.2. Analysis, verification and optimization**

This research direction aims at guaranteeing two different kinds of properties: safety and efficiency.

The first kind of properties concerns the safety of web applications. Software development was traditionally split between critical and non-critical software. Advanced (and costly) formal verification techniques were reserved to the former whereas non-critical software relied almost exclusively on testing, which only offers a 'best-effort' guarantee (removes most bugs but some of them may not be detected). The central idea was that in a non-critical system, the damage a failure may create is not worth the cost of formal verification. However, as web applications grow more pervasive in everyday life and gain momentum in corporates and various social organizations, and touch larger numbers of users, the potential cost of failure is rapidly and

significantly increasing. In that sense, we can consider that web applications are becoming more and more critical. The growing dependency on the web as a tool, combined with the fact that some applications involve very large user bases, is becoming problematic as it seems to increase rapidly but silently. Some errors like crashes and confidential information leaks, if not discovered, can have massive effects and cause significant financial or reputation damage.

The second kind of properties concerns the efficiency of web applications. One particular characteristic of web programming languages is that they are essentially data-manipulation oriented. These manipulations rely on query and transformation languages whose performance is critical. This performance is very sensitive to data size and organization (constraints) and to the execution model (e.g. streaming evaluators). Static analysis can be used to optimize runtime performance by compile-time automated modification of the code (e.g. substitution of queries by more efficient ones). One major scientific difficulty here consists in dealing with problems close to the frontier of decidability, and therefore in finding useful trade-offs between programming ease, expressivity, complexity, succinctness, algorithmic techniques and effective implementations.

### **3.3. Design of advanced (robust, flexible, rich, novel) web applications**

The generalized use of mobile terminals deeply affects the way users perceive and interact with their environment. The ubiquitous use of search engines capable of producing results in fractions of a second raised user expectations to a very high level: users now expect relevant information to be made available to them instantly and directly by context sensitivity to the environment itself. However, the information that needs to be processed is becoming more and more complex compared to the traditional web. In order to unlock the potential introduced by this new generation of the web, a radical rethinking of how web information is produced, organized and processed is necessary.

Until now, content rendering on the web was mainly based on supporting media formats separately. It is still notably the case in HTML5 for example where, for instance, vector graphics, mathematical content, audio and video are supported only as isolated media types. With the increasing use of web content in mobile terminals, we also need to take into account highly dynamic information flowing from sensors (positioning and orientation moves) and cameras. To reach that goal, web development platforms need to ease the manipulation of such content with carefully designed programming interfaces and by developing supporting integrative methods.

More precisely, we will focus on the following aspects: (1) **Build Rich content models**. This requires combining in a single model several content facets such as 3D elements, animations, user interactions, etc. We will focus on feature-compositional methods, which have become a prerequisite for the production of compelling web applications. (2) **Physical environment modeling and integration**. This consists of modeling and representing urban data such as buildings, pathways, points of interest. It requires developing appropriate languages and techniques to represent, process and query such environment models. In particular, we will focus on tracking positional user information and design techniques capable of combining semantic annotations, content, and representation of the physical world. (3) **Native streams support**. This consists of capturing new data flows extracted from various sensors in mobile terminals and various equipments. (4) **Cross-platform abstractions**. We will contribute to the design of appropriate abstractions to make applications run in a uniform way across various devices and environments. Our goal is to provide a viable alternative to current (platform-specific) mobile application development practices.



## **WIMMICS Project-Team**

### **3. Research Program**

#### **3.1. Users Modeling and Designing Interaction on the Web**

Wimmics focuses on interactions of ordinary users with ontology-based knowledge systems with a preference for semantic Web formalisms and Web 2.0 applications. We specialize interaction design and evaluation methods to Web application tasks such as searching, browsing, contributing or protecting data. The team is especially interested in using semantics in assisting the interactions. We propose knowledge graph representations and algorithms to support interaction adaptation for instance for context-awareness or intelligent interactions with machine. We propose and evaluate Web-based visualization techniques for linked data, querying, reasoning, explaining and justifying. Wimmics also integrates natural language processing approaches to support natural language based interactions. We rely on cognitive studies to build models of the system, the user and the interactions between users through the system, in order to support and improve these interactions. We extend the user modeling technique known as *Personas* where user models are represented as specific, individual humans. *Personas* are derived from significant behavior patterns (i.e., sets of behavioral variables) elicited from interviews with and observations of users (and sometimes customers) of the future product. Our user models specialize *Personas* approaches to include aspects appropriate to Web applications. Wimmics also extends user models to capture very different aspects (e.g. emotional states).

#### **3.2. Communities and Social Interactions Analysis**

The domain of social network analysis is a whole research domain in itself and Wimmics targets what can be done with typed graphs, knowledge representations and social models. We also focus on the specificity of social Web and semantic Web applications and in bridging and combining the different social Web data structures and semantic Web formalisms. Beyond the individual user models, we rely on social studies to build models of the communities, their vocabularies, activities and protocols in order to identify where and when formal semantics is useful. We propose models of collectives of users and of their collaborative functioning extending the collaboration personas and methods to assess the quality of coordination interactions and the quality of coordination artifacts. We extend and compare community detection algorithms to identify and label communities of interest with the topics they share. We propose mixed representations containing social semantic representations (e.g. folksonomies) and formal semantic representations (e.g. ontologies) and propose operations that allow us to couple them and exchange knowledge between them. Moving to social interaction we develop models and algorithms to mine and integrate different yet linked aspects of social media contributions (opinions, arguments and emotions) relying in particular on natural language processing and argumentation theory. To complement the study of communities we rely on multi-agent systems to simulate and study social behaviors. Finally we also rely on Web 2.0 principles to provide and evaluate social Web applications.

#### **3.3. Vocabularies, Semantic Web and Linked Data Based Knowledge Representation**

For all the models we identified in the previous sections, we rely on and evaluate knowledge representation methodologies and theories, in particular ontology-based modeling. We also propose models and formalisms to capture and merge representations of different levels of semantics (e.g. formal ontologies and social folksonomies). The important point is to allow us to capture those structures precisely and flexibly and yet create as many links as possible between these different objects. We propose vocabularies and semantic Web formalizations for the whole aspects we model and we consider and study extensions of these formalisms when needed. The results have all in common to pursue the representation and publication of our models as linked

data. We also contribute to the transformation and linking of existing resources (informal models, databases, texts, etc.) to be published on the semantic Web and as linked data. Examples of aspects we formalize include: user profiles, social relations, linguistic knowledge, business processes, derivation rules, temporal descriptions, explanations, presentation conditions, access rights, uncertainty, emotional states, licenses, learning resources, etc. At a more conceptual level we also work on modeling the Web architecture with philosophical tools so as to give a realistic account of identity and reference and to better understand the whole context of our research and its conceptual cornerstones.

### **3.4. Analyzing and Reasoning on Heterogeneous Semantic Graphs**

One of the characteristics of Wimmics is to rely on graph formalisms unified in an abstract graph model and operators unified in an abstract graph machine to formalize and process semantic Web data, Web resources, services metadata and social Web data. In particular Corese, the core software of Wimmics, maintains and implements that abstraction. We propose algorithms to process the mixed representations of the previous section. In particular we are interested in allowing cross-enrichment between them and in exploiting the life cycle and specificity of each one to foster the life-cycles of the others. Our results all have in common to pursue analyzing and reasoning on heterogeneous semantic graphs issued from social and semantic Web applications. Many approaches emphasize the logical aspect of the problem especially because logics are close to computer languages. We defend that the graph nature of Linked Data on the Web and the large variety of types of links that compose them call for typed graphs models. We believe the relational dimension is of paramount importance in these representations and we propose to consider all these representations as fragments of a typed graph formalism directly built above the Semantic Web formalisms. Our choice of a graph based programming approach for the semantic and social Web and of a focus on one graph based formalism is also an efficient way to support interoperability, genericity, uniformity and reuse.

## ZENITH Project-Team

### 3. Research Program

#### 3.1. Data Management

Data management is concerned with the storage, organization, retrieval and manipulation of data of all kinds, from small and simple to very large and complex. It has become a major domain of computer science, with a large international research community and a strong industry. Continuous technology transfer from research to industry has led to the development of powerful DBMS, now at the heart of any information system, and of advanced data management capabilities in many kinds of software products (search engines, application servers, document systems, etc.).

The fundamental principle behind data management is *data independence*, which enables applications and users to deal with the data at a high conceptual level while ignoring implementation details. The relational model, by resting on a strong theory (set theory and first-order logic) to provide data independence, has revolutionized data management. The major innovation of relational DBMS has been to allow data manipulation through queries expressed in a high-level (declarative) language such as SQL. Queries can then be automatically translated into optimized query plans that take advantage of underlying access methods and indices. Many other advanced capabilities have been made possible by data independence : data and metadata modeling, schema management, consistency through integrity rules and triggers, transaction support, etc.

This data independence principle has also enabled DBMS to continuously integrate new advanced capabilities such as object and XML support and to adapt to all kinds of hardware/software platforms from very small smart devices (smart phone, PDA, smart card, etc.) to very large computers (multiprocessor, cluster, etc.) in distributed environments. For a long time, the research focus was on providing advanced database capabilities with good performance, for both transaction processing and decision support applications. And the main objective was to support all these capabilities within a single DBMS.

The problems of scientific data management (massive scale, complexity and heterogeneity) go well beyond the traditional context of DBMS. To address them, we capitalize on scientific foundations in closely related domains: distributed data management, cloud data management, big data, big data integration, scientific workflows, data analytics and search.

#### 3.2. Distributed Data Management

To deal with the massive scale of scientific data, we exploit large-scale distributed systems, with the objective of making distribution transparent to the users and applications. Thus, we capitalize on the principles of large-scale distributed systems such as clusters, peer-to-peer (P2P) and cloud.

Data management in distributed systems has been traditionally achieved by distributed database systems which enable users to transparently access and update several databases in a network using a high-level query language (e.g. SQL) [12]. Transparency is achieved through a global schema which hides the local databases' heterogeneity. In its simplest form, a distributed database system is a centralized server that supports a global schema and implements distributed database techniques (query processing, transaction management, consistency management, etc.). This approach has proved to be effective for applications that can benefit from centralized control and full-fledge database capabilities, e.g. information systems. However, it cannot scale up to more than tens of databases.

Parallel database systems extend the distributed database approach to improve performance (transaction throughput or query response time) by exploiting database partitioning using a multiprocessor or cluster system. Although data integration systems and parallel database systems can scale up to hundreds of data sources or database partitions, they still rely on a centralized global schema and strong assumptions about the network.

In contrast, peer-to-peer (P2P) systems [8] adopt a completely decentralized approach to data sharing. By distributing data storage and processing across autonomous peers in the network, they can scale without the need for powerful servers. Popular examples of P2P systems such as Gnutella and BitTorrent have millions of users sharing petabytes of data over the Internet. Although very useful, these systems are quite simple (e.g. file sharing), support limited functions (e.g. keyword search) and use simple techniques (e.g. resource location by flooding) which have performance problems. A P2P solution is well-suited to support the collaborative nature of scientific applications as it provides scalability, dynamicity, autonomy and decentralized control. Peers can be the participants or organizations involved in collaboration and may share data and applications while keeping full control over their (local) data sources. But for very-large scale scientific data analysis, we believe cloud computing (see next section), is the right approach as it can provide virtually infinite computing, storage and networking resources. However, current cloud architectures are proprietary, ad-hoc, and may deprive users of the control of their own data. Thus, we postulate that a hybrid P2P/cloud architecture is more appropriate for scientific data management, by combining the best of both approaches. In particular, it will enable the clean integration of the users' own computational resources with different clouds.

### 3.3. Cloud Data Management

Cloud computing encompasses on demand, reliable services provided over the Internet (typically represented as a cloud) with easy access to virtually infinite computing, storage and networking resources. Through very simple Web interfaces and at small incremental cost, users can outsource complex tasks, such as data storage, system administration, or application deployment, to very large data centers operated by cloud providers. However, cloud computing has some drawbacks and not all applications are good candidates for being "cloudified". The major concern is w.r.t. data security and privacy, and trust in the provider (which may use not so trustful providers to operate). One earlier criticism of cloud computing was that customers get locked in proprietary clouds. It is true that most clouds are proprietary and there are no standards for cloud interoperability.

There is much more variety in cloud data than in scientific data since there are many different kinds of customers (individuals, small companies, large corporations, etc.). However, we can identify common features. Cloud data can be very large, unstructured (e.g. text-based) or semi-structured, and typically append-only (with rare updates). And cloud users and application developers may be in high numbers, but not DBMS experts.

### 3.4. Big Data

Big data (like its relative, data science) has become a buzz word, with different meanings depending on your perspective, e.g. 100 terabytes is big for a transaction processing system, but small for a web search engine. It is also a moving target, as shown by two landmarks in DBMS products: the Teradata database machine in the 1980's and the Oracle Exadata database machine in 2010.

Although big data has been around for a long time, it is now more important than ever. We can see overwhelming amounts of data generated by all kinds of devices, networks and programs, e.g. sensors, mobile devices, connected objects (IoT), social networks, computer simulations, satellites, radiotelescopes, etc. Storage capacity has doubled every 3 years since 1980 with prices steadily going down (e.g. 1 Gigabyte of Hard Disk Drive for: 1M\$ in 1982, 1K\$ in 1995, 0.02\$ in 2015), making it affordable to keep more data around. And massive data can produce high-value information and knowledge, which is critical for data analysis, decision support, forecasting, business intelligence, research, (data-intensive) science, etc.

The problem of big data has three main dimensions, quoted as the three big V's:

- Volume: refers to massive amounts of data, making it hard to store, manage, and analyze (big analytics);
- Velocity: refers to continuous data streams being produced, making it hard to perform online processing and analysis;
- Variety: refers to different data formats, different semantics, uncertain data, multiscale data, etc., making it hard to integrate and analyze.

There are also other V's such as: validity (is the data correct and accurate?); veracity (are the results meaningful?); volatility (how long do you need to store this data?).

Many different big data management solutions have been designed, primarily for the cloud, as cloud and big data are synergistic. They typically trade consistency for scalability, simplicity and flexibility, hence the new term Data-Intensive Scalable Computing (DISC). Examples of DISC systems include data processing frameworks (e.g. Hadoop MapReduce, Apache Spark, Pregel), file systems (e.g. Google GFS, HDFS), NoSQL systems (Google BigTable, Hbase, MongoDB), NewSQL systems (Google F1, CockroachDB, LeanXcale). In Zenith, we exploit or extend DISC technologies to fit our needs for scientific workflow management and scalable data analysis.

### 3.5. Data Integration

Scientists can rely on web tools to quickly share their data and/or knowledge. Therefore, when performing a given study, a scientist would typically need to access and integrate data from many data sources (including public databases). Data integration can be either physical or logical. In the former, the source data are integrated and materialized in a data warehouse. In logical integration, the integrated data are not materialized, but accessed indirectly through a global (or mediated) schema using a data integration system. These two approaches have different trade-offs, e.g. efficient analytics but only on historical data for data warehousing versus real-time access to data sources for data integration systems (e.g. web price comparators).

In both cases, to understand a data source content, metadata (data that describe the data) is crucial. Metadata can be initially provided by the data publisher to describe the data structure (e.g. schema), data semantics based on ontologies (that provide a formal representation of the domain knowledge) and other useful information about data provenance (publisher, tools, methods, etc.). Scientific metadata is very heterogeneous, in particular because of the autonomy of the underlying data sources, which leads to a large variety of models and formats. Thus, it is necessary to identify semantic correspondences between the metadata of the related data sources. This requires the matching of the heterogeneous metadata, by discovering semantic correspondences between ontologies, and the annotation of data sources using ontologies. In Zenith, we rely on semantic web techniques (e.g. RDF and SparkQL) to perform these tasks and deal with high numbers of data sources.

Scientific workflow management systems (SWfMS) are also useful for data integration. They allow scientists to describe and execute complex scientific activities, by automating data derivation processes, and supporting various functions such as provenance management, queries, reuse, etc. Some workflow activities may access or produce huge amounts of distributed data. This requires using distributed and parallel execution environments. However, existing workflow management systems have limited support for data parallelism. In Zenith, we use an algebraic approach to describe data-intensive workflows and exploit parallelism.

### 3.6. Data Analytics

Data analytics refers to a set of techniques to draw conclusions through data examination. It involves data mining, statistics, and data management. Data mining provides methods to discover new and useful patterns from very large datasets. These patterns may take different forms, depending on the end-user's request, such as:

- **Frequent itemsets and association rules.** In this case, the data is usually a table with a high number of rows and the data mining algorithm extracts correlations between column values. This problem was first motivated by commercial and marketing purposes (e.g. discovering frequent correlations between items bought in a shop, which could help selling more). A typical example of frequent itemset from a sensor network in a smart building would say that “in 20% rooms, the door is closed, the room is empty, and lights are on.”
- **Frequent sequential pattern extraction.** This problem is very similar to frequent itemset discovery but considering the order between. In the smart building example, a frequent sequence could say that “in 40% of rooms, lights are on at time  $i$ , the room is empty at time  $i+j$  and the door is closed at time  $i+j+k$ ”. Discovering frequent sequences has become critical in marketing, as well as in security (e.g.

detecting network intrusions), in web usage analysis and any domain where data come in a specific order, typically given by timestamps.

- **Clustering.** The goal of clustering is to group together similar data while ensuring that dissimilar data will not be in the same cluster. In our example of smart buildings, we could find clusters of rooms, where offices will be in one category and copy machine rooms in another because of their differences (hours of people presence, number of times lights are turned on/off, etc.).

One main problem in data analytics is to deal with data streams. Existing methods have been designed for very large data sets where complex algorithms from artificial intelligence were not efficient because of data size. However, we now must deal with data streams, sequences of data events arriving at high rate, where traditional data analytics techniques cannot complete in real-time, given the infinite data size. In order to extract knowledge from data streams, the data mining community has investigated approximation methods that could yield good result quality.

### 3.7. Data Search

Technologies for searching information in scientific data have relied on relational DBMS or text-based indexing methods. However, content-based information retrieval has progressed much in the last decade, with much impact on search engines. Rather than restricting the search to the use of metadata, content-based methods index, search and browse digital objects by means of signatures that describe their content. Such methods have been intensively studied in the multimedia community to allow searching massive amounts of multimedia documents that are created every day (e.g. 99% of web data are audio-visual content with very sparse metadata). Scalable content-based methods have been proposed for searching objects in large image collections or detecting copies in huge video archives. Besides multimedia contents, content-based information retrieval methods have expanded their scope to deal with more diverse data such as medical images, 3D models or even molecular data. Potential applications in scientific data management are numerous. First, to allow searching within huge collections of scientific images (earth observation, medical images, botanical images, biology images, etc.) or browsing large datasets of experimental data (e.g. multisensor data, molecular data or instrumental data). However, scalability remains a major issue, involving complex algorithms (such as similarity search, clustering or supervised retrieval), in high dimensional spaces (up to millions of dimensions) with complex metrics ( $L_p$ , Kernels, sets intersections, edit distances, etc.). Most of these algorithms have linear, quadratic or even cubic complexities so that their use at large scale is not affordable without major breakthroughs. In Zenith, we investigate the following challenges:

- **High-dimensional similarity search.** Whereas many indexing methods were designed in the last 20 years to efficiently retrieve multidimensional data with relatively small dimensions, high-dimensional data are challenged by the well-known curse of dimensionality. Only recently have some methods appeared that allow approximate Nearest Neighbors queries in sub-linear time, in particular, Locality Sensitive Hashing methods that offer new theoretical insights in high-dimensional Euclidean spaces and random projections. But there are still challenging issues such as efficient similarity search in any kernel or metric spaces, efficient construction of k-nearest neighbor graphs (k-NNG) or relational similarity queries.
- **Large-scale supervised retrieval.** Supervised retrieval aims at retrieving relevant objects in a dataset by providing some positive and/or negative training samples. Toward this goal, Support Vector Machines (SVM) offer the possibility to construct generalized, non-linear predictors in high-dimensional spaces using small training sets. The prediction time complexity of these methods is usually linear in dataset size. Allowing hyperplane similarity queries in sub-linear time is for example a challenging research issue. A symmetric problem in supervised retrieval consists in retrieving the most relevant object categories that might contain a given query object, providing huge labeled datasets (up to millions of classes and billions of objects) and very few objects per category (from 1 to 100 objects). SVM methods that are formulated as quadratic programming with cubic training time complexity and quadratic space complexity are clearly not usable. Promising solutions include hybrid supervised-unsupervised methods and supervised hashing methods.

- **Distributed content-based retrieval.** Distributed content-based retrieval methods appear as a promising solution to manage masses of data distributed over large networks, in particular when the data cannot be centralized for privacy or cost reasons, which is often the case in scientific social networks. However, current methods are limited to very simple similarity search paradigms. In Zenith, we consider more advanced distributed content-based retrieval and mining methods such as k-NNG construction, large-scale supervised retrieval or multi-source clustering.

## ALICE Project-Team

### 3. Research Program

#### 3.1. Introduction

Computer Graphics is a quickly evolving domain of research. These last few years, both acquisition techniques (*e.g.*, range laser scanners) and computer graphics hardware (the so-called GPU's, for Graphics Processing Units) have made considerable advances. However, despite these advances, fundamental problems still remain open. For instance, a scanned mesh composed of hundred millions triangles cannot be used directly in real-time visualization or complex numerical simulation. To design efficient solutions for these difficult problems, ALICE studies two fundamental issues in Computer Graphics:

- the representation of the objects, *i.e.*, their geometry and physical properties;
- the interaction between these objects and light.

Historically, these two issues have been studied by independent research communities. However, we think that they share a common theoretical basis. For instance, multi-resolution and wavelets were mathematical tools used by both communities [29]. We develop a new approach, which consists in studying the geometry and lighting from the *numerical analysis* point of view. In our approach, geometry processing and light simulation are systematically restated as a (possibly non-linear and/or constrained) functional optimization problem. This type of formulation leads to algorithms that are more efficient. Our long-term research goal is to find a formulation that permits a unified treatment of geometry and illumination over this geometry.

#### 3.2. Geometry Processing for Engineering

**Keywords:** Mesh processing, parameterization, splines

Geometry processing recently emerged (in the middle of the 90's) as a promising strategy to solve the geometric modeling problems encountered when manipulating meshes composed of hundred millions of elements. Since a mesh may be considered to be a *sampling* of a surface - in other words a *signal* - the *digital signal processing* formalism was a natural theoretic background for this subdomain (see *e.g.*, [30]). Researchers of this domain then studied different aspects of this formalism applied to geometric modeling.

Although many advances have been made in the geometry processing area, important problems still remain open. Even if shape acquisition and filtering is much easier than 30 years ago, a scanned mesh composed of hundred million triangles cannot be used directly in real-time visualization or complex numerical simulation. For this reason, automatic methods to convert those large meshes into higher level representations are necessary. However, these automatic methods do not exist yet. For instance, the pioneer Henri Gouraud often mentions in his talks that the *data acquisition* problem is still open [19]. Malcolm Sabin, another pioneer of the "Computer Aided Geometric Design" and "Subdivision" approaches, mentioned during several conferences of the domain that constructing the optimum control-mesh of a subdivision surface so as to approximate a given surface is still an open problem [28]. More generally, converting a mesh model into a higher level representation, consisting of a set of equations, is a difficult problem for which no satisfying solutions have been proposed. This is one of the long-term goals of international initiatives, such as the **AIMShape** European network of excellence.

Motivated by gridding application for finite elements modeling for oil and gas exploration, in the frame of the **Gocad** project, we started studying geometry processing in the late 90's and contributed to this area at the early stages of its development. We developed the LSCM method (Least Squares Conformal Maps) in cooperation with Alias Wavefront [24]. This method has become the de-facto standard in automatic unwrapping, and was adopted by several 3D modeling packages (including Maya and Blender). We explored various applications of the method, including normal mapping, mesh completion and light simulation [21].



However, classical mesh parameterization requires to partition the considered object into a set of topological disks. For this reason, we designed a new method (Periodic Global Parameterization) that generates a continuous set of coordinates over the object [26]. We also showed the applicability of this method, by proposing the first algorithm that converts a scanned mesh into a Spline surface automatically [23].

We are still not fully satisfied with these results, since the method remains quite complicated. We think that a deeper understanding of the underlying theory is likely to lead to both efficient and simple methods. For this reason, in 2012 we studied several ways of discretizing partial differential equations on meshes, including Finite Element Modeling and Discrete Exterior Calculus. In 2013, we also explored Spectral Geometry Processing and Sampling Theory (more on this below).

### 3.3. Computer Graphics

**Keywords:** texture synthesis, shape synthesis, texture mapping, visibility

Content creation is one of the major challenges in Computer Graphics. Modeling shapes and surface appearances which are visually appealing and at the same time enforce precise design constraints is a task only accessible to highly skilled and trained designers.

In this context the team focuses on methods for by-example content creation. Given an input example and a set of constraints, we design algorithms that can automatically generate a new shape (geometry+texture). We formulate the problem of content synthesis as the joint optimization of several objectives: Preserving the local appearance of the example, enforcing global objectives (size, symmetries, mechanical properties), reaching user defined constraints (locally specified geometry, contacts). This results in a wide range of optimization problems, from statistical approaches (Markov Random fields), to combinatorial and linear optimization techniques.

As a complement to the design of techniques for automatic content creation, we also work on the representation of the content, so as to allow for its efficient manipulation. In this context we develop data structures and algorithms targeted at massively parallel architectures, such as GPUs. These are critical to reach the interactive rates expected from a content creation technique. We also propose novel ways to store and access content defined along surfaces [27] or inside volumes [18] [22].

The team also continues research in core topics of computer graphics at the heart of realistic rendering and realistic light simulation techniques; for example, mapping textures on surfaces, or devising visibility relationships between 3D objects populating space.

## AVIZ Project-Team

### 3. Research Program

#### 3.1. Scientific Foundations

The scientific foundations of Visual Analytics lie primarily in the domains of Visualization and Data Mining. Indirectly, it inherits from other established domains such as graphic design, Exploratory Data Analysis (EDA), statistics, Artificial Intelligence (AI), Human-Computer Interaction (HCI), and Psychology.

The use of graphic representation to understand abstract data is a goal Visual Analytics shares with Tukey's Exploratory Data Analysis (EDA) [58], graphic designers such as Bertin [46] and Tufte [57], and HCI researchers in the field of Information Visualization [44].

EDA is complementary to classical statistical analysis. Classical statistics starts from a *problem*, gathers *data*, designs a *model* and performs an *analysis* to reach a *conclusion* about whether the data follows the model. While EDA also starts with a problem and data, it is most useful *before* we have a model; rather, we perform visual analysis to discover what kind of model might apply to it. However, statistical validation is not always required with EDA; since often the results of visual analysis are sufficiently clear-cut that statistics are unnecessary.

Visual Analytics relies on a process similar to EDA, but expands its scope to include more sophisticated graphics and areas where considerable automated analysis is required before the visual analysis takes place. This richer data analysis has its roots in the domain of Data Mining, while the advanced graphics and interactive exploration techniques come from the scientific fields of Data Visualization and HCI, as well as the expertise of professions such as cartography and graphic designers who have long worked to create effective methods for graphically conveying information.

The books of the cartographer Bertin and the graphic designer Tufte are full of rules drawn from their experience about how the meaning of data can be best conveyed visually. Their purpose is to find effective visual representation that describe a data set but also (mainly for Bertin) to discover structure in the data by using the right mappings from abstract dimensions in the data to visual ones.

For the last 25 years, the field of Human-Computer Interaction (HCI) has also shown that interacting with visual representations of data in a tight perception-action loop improves the time and level of understanding of data sets. Information Visualization is the branch of HCI that has studied visual representations suitable to understanding and interaction methods suitable to navigating and drilling down on data. The scientific foundations of Information Visualization come from theories about perception, action and interaction.

Several theories of perception are related to information visualization such as the "Gestalt" principles, Gibson's theory of visual perception [51] and Triesman's "preattentive processing" theory [56]. We use them extensively but they only have a limited accuracy for predicting the effectiveness of novel visual representations in interactive settings.

Information Visualization emerged from HCI when researchers realized that interaction greatly enhanced the perception of visual representations.

To be effective, interaction should take place in an interactive loop faster than 100ms. For small data sets, it is not difficult to guarantee that analysis, visualization and interaction steps occur in this time, permitting smooth data analysis and navigation. For larger data sets, more computation should be performed to reduce the data size to a size that may be visualized effectively.

In 2002, we showed that the practical limit of InfoVis was on the order of 1 million items displayed on a screen [49]. Although screen technologies have improved rapidly since then, eventually we will be limited by the physiology of our vision system: about 20 millions receptor cells (rods and cones) on the retina. Another problem will be the limits of human visual attention, as suggested by our 2006 study on change blindness in large and multiple displays [47]. Therefore, visualization alone cannot let us understand very large data sets. Other techniques such as aggregation or sampling must be used to reduce the visual complexity of the data to the scale of human perception.

Abstracting data to reduce its size to what humans can understand is the goal of Data Mining research. It uses data analysis and machine learning techniques. The scientific foundations of these techniques revolve around the idea of finding a good model for the data. Unfortunately, the more sophisticated techniques for finding models are complex, and the algorithms can take a long time to run, making them unsuitable for an interactive environment. Furthermore, some models are too complex for humans to understand; so the results of data mining can be difficult or impossible to understand directly.

Unlike pure Data Mining systems, a Visual Analytics system provides analysis algorithms and processes compatible with human perception and understandable to human cognition. The analysis should provide understandable results quickly, even if they are not ideal. Instead of running to a predefined threshold, algorithms and programs should be designed to allow trading speed for quality and show the tradeoffs interactively. This is not a temporary requirement: it will be with us even when computers are much faster, because good quality algorithms are at least quadratic in time (e.g. hierarchical clustering methods). Visual Analytics systems need different algorithms for different phases of the work that can trade speed for quality in an understandable way.

Designing novel interaction and visualization techniques to explore huge data sets is an important goal and requires solving hard problems, but how can we assess whether or not our techniques and systems provide real improvements? Without this answer, we cannot know if we are heading in the right direction. This is why we have been actively involved in the design of evaluation methods for information visualization [55], [54], [52], [53], [50]. For more complex systems, other methods are required. For these we want to focus on longitudinal evaluation methods while still trying to improve controlled experiments.

## 3.2. Innovation

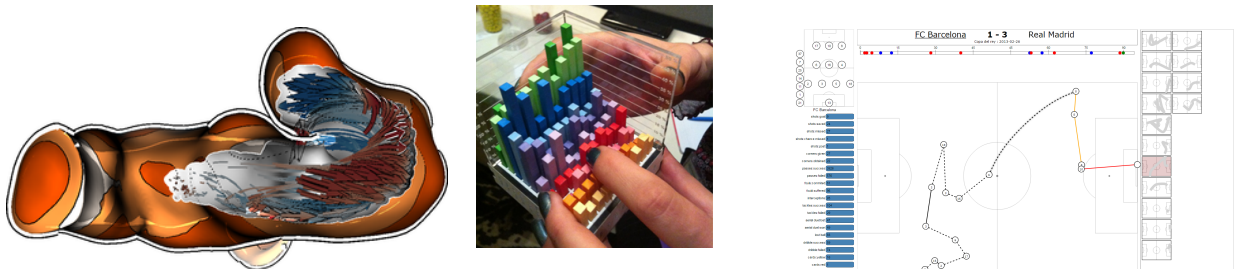


Figure 1. Example novel visualization techniques and tools developed by the team. Left: a non-photorealistic rendering technique that visualizes blood flow and vessel thickness. Middle: a physical visualization showing economic indicators for several countries, right: SoccerStories a tool for visualizing soccer games.

We design novel visualization and interaction techniques (see, for example, Figure 1 ). Many of these techniques are also evaluated throughout the course of their respective research projects. We cover application domains such as sports analysis, digital humanities, fluid simulations, and biology. A focus of Aviz' work is the improvement of graph visualization and interaction with graphs. We further develop individual techniques

for the design of tabular visualizations and different types of data charts. Another focus is the use of animation as a transition aid between different views of the data. We are also interested in applying techniques from illustrative visualization to visual representations and applications in information visualization as well as scientific visualization.

### 3.3. Evaluation Methods

Evaluation methods are required to assess the effectiveness and usability of visualization and analysis methods. Aviz typically uses traditional HCI evaluation methods, either quantitative (measuring speed and errors) or qualitative (understanding users tasks and activities). Moreover, Aviz is also contributing to the improvement of evaluation methods by reporting on the best practices in the field, by co-organizing workshops (BELIV 2010, 2012, 2014, 2016) to exchange on novel evaluation methods, by improving our ways of reporting, interpreting and communicating statistical results, and by applying novel methodologies, for example to assess visualization literacy.

### 3.4. Software Infrastructures

We want to understand the requirements that software and hardware architectures should provide to support exploratory analysis of large amounts of data. So far, “big data” has been focusing on issues related to storage management and predictive analysis: applying a well-known set of operations on large amounts of data. Visual Analytics is about exploration of data, with sometimes little knowledge of its structure or properties. Therefore, interactive exploration and analysis is needed to build knowledge and apply appropriate analyses; this knowledge and appropriateness is supported by visualizations. However, applying analytical operations on large data implies long-lasting computations, incompatible with interactions, and generates large amounts of results, impossible to visualize directly without aggregation or sampling. Visual Analytics has started to tackle these problems for specific applications but not in a general manner, leading to fragmentation of results and difficulties to reuse techniques from one application to the other. We are interested in abstracting-out the issues and finding general architectural models, patterns, and frameworks to address the Visual Analytics challenge in more generic ways.

### 3.5. Emerging Technologies



*Figure 2. Example emerging technology solutions developed by the team for multi-display environments, wall displays, and token-based visualization.*

We want to empower humans to make use of data using different types of display media and to enhance how they can understand and visually and interactively explore information. This includes novel display equipment and accompanying input techniques. The Aviz team specifically focuses on the exploration of the use of large displays in visualization contexts as well as emerging physical and tangible visualizations. In terms of interaction modalities our work focuses on using touch and tangible interaction. Aviz participates to the Digiscope project that funds 11 wall-size displays at multiple places in the Paris area (see <http://www.>

[digiscope.fr](http://digiscope.fr)), connected by telepresence equipment and a Fablab for creating devices. Aviz is in charge of creating and managing the Fablab, uses it to create physical visualizations, and is also using the local wall-size display (called WILD) to explore visualization on large screens. The team also investigates the perceptual, motor and cognitive implications of using such technologies for visualization.

### **3.6. Psychology**

More cross-fertilization is needed between psychology and information visualization. The only key difference lies in their ultimate objective: understanding the human mind vs. helping to develop better tools. We focus on understanding and using findings from psychology to inform new tools for information visualization. In many cases, our work also extends previous work in psychology. Our approach to the psychology of information visualization is largely holistic and helps bridge gaps between perception, action and cognition in the context of information visualization. Our focus includes the perception of charts in general, perception in large display environments, collaboration, perception of animations, how action can support perception and cognition, and judgment under uncertainty.

## EX-SITU Team

### 3. Research Program

#### 3.1. Research Program

We characterize Extreme Situated Interaction as follows:

**Extreme users.** We study extreme users who make extreme demands on current technology. We know that human beings take advantage of the laws of physics to find creative new uses for physical objects. However, this level of adaptability is severely limited when manipulating digital objects. Even so, we find that creative professionals—artists, designers and scientists—often adapt interactive technology in novel and unexpected ways and find creative solutions. By studying these users, we hope to not only address the specific problems they face, but also to identify the underlying principles that will help us to reinvent virtual tools. We seek to shift the paradigm of interactive software, to establish the laws of interaction that significantly empower users and allow them to control their digital environment.

**Extreme situations.** We develop extreme environments that push the limits of today's technology. We take as given that future developments will solve "practical" problems such as cost, reliability and performance and concentrate our efforts on interaction in and with such environments. This has been a successful strategy in the past: Personal computers only became prevalent after the invention of the desktop graphical user interface. Smartphones and tablets only became commercially successful after Apple cracked the problem of a usable touch-based interface for the iPhone and the iPad. Although wearable technologies, such as watches and glasses, are finally beginning to take off, we do not believe that they will create the major disruptions already caused by personal computers, smartphones and tablets. Instead, we believe that future disruptive technologies will include fully interactive paper and large interactive displays.

Our extensive experience with the Digiscope WILD and WILDER platforms places us in a unique position to understand the principles of distributed interaction that extreme environments call for. We expect to integrate, at a fundamental level, the collaborative capabilities that such environments afford. Indeed almost all of our activities in both the digital and the physical world take place within a complex web of human relationships. Current systems only support, at best, passive sharing of information, e.g., through the distribution of independent copies. Our goal is to support active collaboration, in which multiple users are actively engaged in the lifecycle of digital artifacts.

**Extreme design.** We explore novel approaches to the design of interactive systems, with particular emphasis on extreme users in extreme environments. Our goal is to empower creative professionals, allowing them to act as both designers and developers throughout the design process. Extreme design affects every stage, from requirements definition, to early prototyping and design exploration, to implementation, to adaptation and appropriation by end users. We hope to push the limits of participatory design to actively support creativity at all stages of the design lifecycle.

Extreme design does not stop with purely digital artifacts. The advent of digital fabrication tools and FabLabs has significantly lowered the cost of making physical objects interactive. Creative professionals now create hybrid interactive objects that can be tuned to the user's needs. Integrating the design of physical objects into the software design process raises new challenges, with new methods and skills to support this form of extreme prototyping.

Our overall approach is to identify a small number of specific projects, organized around four themes: *Creativity*, *Augmentation*, *Collaboration* and *Infrastructure*. Specific projects may address multiple themes, and different members of the group work together to advance these different topics.

## GRAPHDECO Project-Team

### 3. Research Program

#### 3.1. Introduction

Our research program is oriented around two main axes: 1) Computer-Assisted Design with Heterogeneous Representations and 2) Graphics with Uncertainty and Heterogeneous Content. These two axes are governed by a set of common fundamental goals, share many common methodological tools and are deeply intertwined in the development of applications.

##### 3.1.1. Computer-Assisted Design with Heterogeneous Representations

Designers use a variety of visual representations to explore and communicate about a concept. Figure 2 illustrates some typical representations, including sketches, hand-made prototypes, 3D models, 3D printed prototypes or instructions.

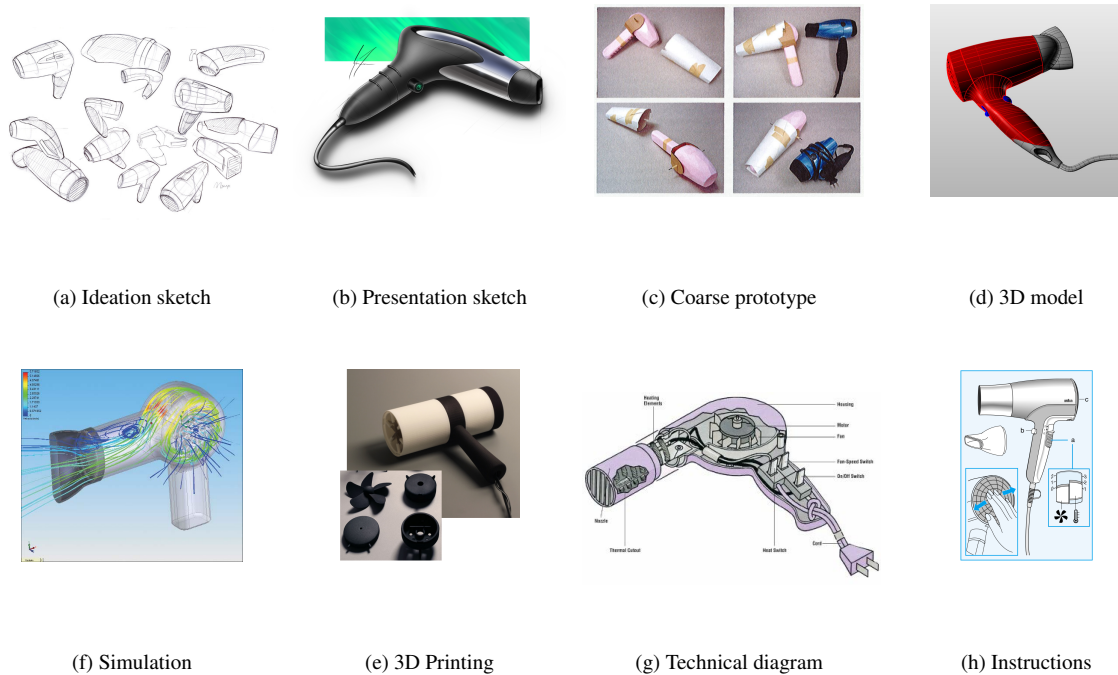


Figure 2. Various representations of a hair dryer at different stages of the design process. Image source, in order: *c-maeng on deviantart.com*, *shauntur on deviantart.com*, "Prototyping and Modelmaking for Product Design" Hallgrímsson, B., Laurence King Publishers, 2012, *samsher511 on turbosquid.com*, *my.solidworks.com*, *weilung tseng on cargocollective.com*, *howstuffworks.com*, *u-manual.com*

The early representations of a concept, such as rough sketches and hand-made prototypes, help designers formulate their ideas and test the form and function of multiple design alternatives. These low-fidelity representations are meant to be cheap and fast to produce, to allow quick exploration of the *design space* of the concept. These representations are also often approximate to leave room for subjective interpretation and to stimulate imagination; in this sense, these representations can be considered *uncertain*. As the concept gets more finalized, time and effort are invested in the production of more detailed and accurate representations, such as high-fidelity 3D models suitable for simulation and fabrication. These detailed models can also be used to create didactic instructions for assembly and usage.

Producing these different representations of a concept requires specific skills in sketching, modeling, manufacturing and visual communication. For these reasons, professional studios often employ different experts to produce the different representations of the same concept, at the cost of extensive discussions and numerous iterations between the actors of this process. The complexity of the multi-disciplinary skills involved in the design process also hinders their adoption by laymen.

Existing solutions to facilitate design have focused on a subset of the representations used by designers. However, no solution considers all representations at once, for instance to directly convert a series of sketches into a set of physical prototypes. In addition, all existing methods assume that the concept is unique rather than ambiguous. As a result, rich information about the variability of the concept is lost during each conversion step.

We plan to facilitate design for professionals and laymen by addressing the following objectives:

- We want to assist designers in the exploration of the *design space* that captures the possible variations of a concept. By considering a concept as a *distribution* of shapes and functionalities rather than a single object, our goal is to help designers consider multiple design alternatives more quickly and effectively. Such a representation should also allow designers to preserve multiple alternatives along all steps of the design process rather than committing to a single solution early on and pay the price of this decision for all subsequent steps. We expect that preserving alternatives will facilitate communication with engineers, managers and clients, accelerate design iterations and even allow mass personalization by the end consumers.
- We want to support the various representations used by designers during concept development. While drawings and 3D models have received significant attention in past Computer Graphics research, we will also account for the various forms of rough physical prototypes made to evaluate the shape and functionality of a concept. Depending on the task at hand, our algorithms will either analyse these prototypes to generate a virtual concept, or assist the creation of these prototypes from a virtual model. We also want to develop methods capable of adapting to the different drawing and manufacturing techniques used to create sketches and prototypes. We envision design tools that conform to the habits of users rather than impose specific techniques to them.
- We want to make professional design techniques available to novices. Affordable software, hardware and online instructions are democratizing technology and design, allowing small businesses and individuals to compete with large companies. New manufacturing processes and online interfaces also allow customers to participate in the design of an object via mass personalization. However, similarly to what happened for desktop publishing thirty years ago, desktop manufacturing tools need to be simplified to account for the needs and skills of novice designers. We hope to support this trend by adapting the techniques of professionals and by automating the tasks that require significant expertise.

### 3.1.2. Graphics with Uncertainty and Heterogeneous Content

Our research is motivated by the observation that traditional CG algorithms have not been designed to account for uncertain data. For example, global illumination rendering assumes accurate virtual models of geometry, light and materials to simulate light transport. While these algorithms produce images of high realism, capturing effects such as shadows, reflections and interreflections, they are not applicable to the growing mass of uncertain data available nowadays.

The need to handle uncertainty in CG is timely and pressing, given the large number of *heterogeneous sources of 3D content* that have become available in recent years. These include data from cheap depth+image sensors (e.g., Kinect or the Tango), 3D reconstructions from image/video data, but also data from large 3D geometry databases, or casual 3D models created using simplified sketch-based modeling tools. Such alternate content has varying levels of *uncertainty* about the scene or objects being modelled. This includes uncertainty in geometry, but also in materials and/or lights – which are often not even available with such content. Since CG algorithms cannot be applied directly, visual effects artists spend hundreds of hours correcting inaccuracies and completing the captured data to make them useable in film and advertising.





Figure 3. Image-Based Rendering (IBR) techniques use input photographs and approximate 3D to produce new synthetic views.

We identify a major scientific bottleneck which is the need to treat *heterogeneous* content, i.e., containing both (mostly captured) uncertain and perfect, traditional content. Our goal is to provide solutions to this bottleneck, by explicitly and formally modeling uncertainty in CG, and to develop new algorithms that are capable of mixed rendering for this content.

We strive to develop methods in which heterogeneous – and often uncertain – data can be handled automatically in CG with a principled methodology. Our main focus is on *rendering* in CG, including dynamic scenes (video/animations).

Given the above, we need to address the following challenges:

- Develop a theoretical model to handle uncertainty in computer graphics. We must define a new formalism that inherently incorporates uncertainty, and must be able to express traditional CG rendering, both physically accurate and approximate approaches. Most importantly, the new formulation must elegantly handle mixed rendering of perfect synthetic data and captured uncertain content. An important element of this goal is to incorporate *cost* in the choice of algorithm and the optimizations used to obtain results, e.g., preferring solutions which may be slightly less accurate, but cheaper in computation or memory.
- The development of rendering algorithms for heterogeneous content often requires preprocessing of image and video data, which sometimes also includes depth information. An example is the decomposition of images into intrinsic layers of reflectance and lighting, which is required to perform relighting. Such solutions are also useful as image-manipulation or computational photography techniques. The challenge will be to develop such “intermediate” algorithms for the uncertain and heterogeneous data we target.
- Develop efficient rendering algorithms for uncertain and heterogeneous content, reformulating rendering in a probabilistic setting where appropriate. Such methods should allow us to develop approximate rendering algorithms using our formulation in a well-grounded manner. The formalism should include probabilistic models of how the scene, the image and the data interact. These models should be data-driven, e.g., building on the abundance of online geometry and image databases, domain-driven, e.g., based on requirements of the rendering algorithms or perceptually guided, leading to plausible solutions based on limitations of perception.

## HYBRID Project-Team

### 3. Research Program

#### 3.1. Research Program

The scientific objective of Hybrid team is to improve 3D interaction of one or multiple users with virtual environments, by making full use of physical engagement of the body, and by incorporating the mental states by means of brain-computer interfaces. We intend to improve each component of this framework individually, but we also want to improve the subsequent combinations of these components.

The "hybrid" 3D interaction loop between one or multiple users and a virtual environment is depicted in Figure 1. Different kinds of 3D interaction situations are distinguished (red arrows, bottom): 1) body-based interaction, 2) mind-based interaction, 3) hybrid and/or 4) collaborative interaction (with at least two users). In each case, three scientific challenges arise which correspond to the three successive steps of the 3D interaction loop (blue squares, top): 1) the 3D interaction technique, 2) the modeling and simulation of the 3D scenario, and 3) the design of appropriate sensory feedback.

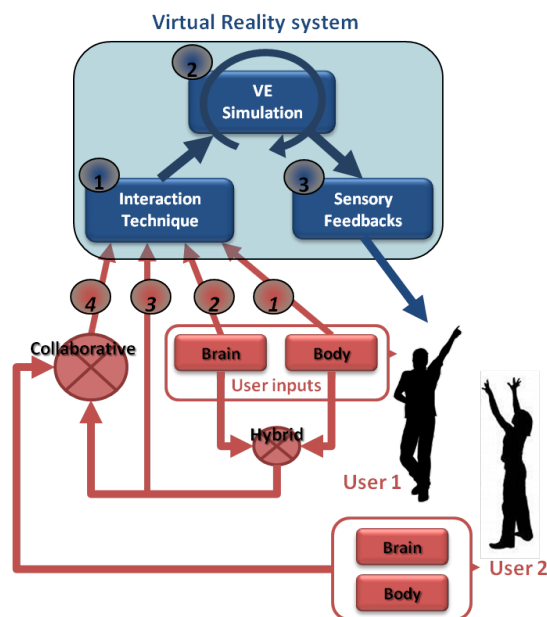


Figure 1. 3D hybrid interaction loop between one or multiple users and a virtual reality system. Top (in blue) three steps of 3D interaction with a virtual environment: (1-blue) interaction technique, (2-blue) simulation of the virtual environment, (3-blue) sensory feedbacks. Bottom (in red) different cases of interaction: (1-red) body-based, (2-red) mind-based, (3-red) hybrid, and (4-red) collaborative 3D interaction.

The 3D interaction loop involves various possible inputs from the user(s) and different kinds of output (or sensory feedback) from the simulated environment. Each user can involve his/her body and mind by means of corporal and/or brain-computer interfaces. A hybrid 3D interaction technique (1) mixes mental and motor inputs and translates them into a command for the virtual environment. The real-time simulation (2) of the

virtual environment is taking into account these commands to change and update the state of the virtual world and virtual objects. The state changes are sent back to the user and perceived by means of different sensory feedbacks (e.g., visual, haptic and/or auditory) (3). The sensory feedbacks are closing the 3D interaction loop. Other users can also interact with the virtual environment using the same procedure, and can eventually “collaborate” by means of “collaborative interactive techniques” (4).

This description is stressing three major challenges which correspond to three mandatory steps when designing 3D interaction with virtual environments:

- **3D interaction techniques:** This first step consists in translating the actions or intentions of the user (inputs) into an explicit command for the virtual environment. In virtual reality, the classical tasks that require such kinds of user command were early categorized in four [38]: navigating the virtual world, selecting a virtual object, manipulating it, or controlling the application (entering text, activating options, etc). The addition of a third dimension, the use of stereoscopic rendering and the use of advanced VR interfaces make however inappropriate many techniques that proved efficient in 2D, and make it necessary to design specific interaction techniques and adapted tools. This challenge is here renewed by the various kinds of 3D interaction which are targeted. In our case, we consider various cases, with motor and/or cerebral inputs, and potentially multiple users.
- **Modeling and simulation of complex 3D scenarios:** This second step corresponds to the update of the state of the virtual environment, in real-time, in response to all the potential commands or actions sent by the user. The complexity of the data and phenomena involved in 3D scenarios is constantly increasing. It corresponds for instance to the multiple states of the entities present in the simulation (rigid, articulated, deformable, fluids, which can constitute both the user’s virtual body and the different manipulated objects), and the multiple physical phenomena implied by natural human interactions (squeezing, breaking, melting, etc). The challenge consists here in modeling and simulating these complex 3D scenarios and meeting, at the same time, two strong constraints of virtual reality systems: performance (real-time and interactivity) and genericity (e.g., multi-resolution, multi-modal, multi-platform, etc).
- **Immersive sensory feedbacks:** This third step corresponds to the display of the multiple sensory feedbacks (output) coming from the various VR interfaces. These feedbacks enable the user to perceive the changes occurring in the virtual environment. They are closing the 3D interaction loop, making the user immersed, and potentially generating a subsequent feeling of presence. Among the various VR interfaces which have been developed so far we can stress two kinds of sensory feedback: visual feedback (3D stereoscopic images using projection-based systems such as CAVE systems or Head Mounted Displays); and haptic feedback (related to the sense of touch and to tactile or force-feedback devices). The Hybrid team has a strong expertise in haptic feedback, and in the design of haptic and “pseudo-haptic” rendering [41]. Note that a major trend in the community, which is strongly supported by the Hybrid team, relates to a “perception-based” approach, which aims at designing sensory feedbacks which are well in line with human perceptual capacities.

These three scientific challenges are addressed differently according to the context and the user inputs involved. We propose to consider three different contexts, which correspond to the three different research axes of the Hybrid research team, namely : 1) body-based interaction (motor input only), 2) mind-based interaction (cerebral input only), and then 3) hybrid and collaborative interaction (i.e., the mixing of body and brain inputs from one or multiple users).

## 3.2. Research Axes

The scientific activity of Hybrid team follows three main axes of research:

- **Body-based interaction in virtual reality.** Our first research axis concerns the design of immersive and effective “body-based” 3D interactions, i.e., relying on a physical engagement of the user’s body. This trend is probably the most popular one in VR research at the moment. Most VR setups make use of tracking systems which measure specific positions or actions of the user in order to interact with a virtual environment. However, in recent years, novel options have emerged for measuring

“full-body” movements or other, even less conventional, inputs (e.g. body equilibrium). In this first research axis we are thus concerned by the emergence of new kinds of “body-based interaction” with virtual environments. This implies the design of novel 3D user interfaces and novel 3D interactive techniques, novel simulation models and techniques, and novel sensory feedbacks for body-based interaction with virtual worlds. It involves real-time physical simulation of complex interactive phenomena, and the design of corresponding haptic and pseudo-haptic feedback.

- **Mind-based interaction in virtual reality.** Our second research axis concerns the design of immersive and effective “mind-based” 3D interactions in Virtual Reality. Mind-based interaction with virtual environments is making use of Brain-Computer Interface technology. This technology corresponds to the direct use of brain signals to send “mental commands” to an automated system such as a robot, a prosthesis, or a virtual environment. BCI is a rapidly growing area of research and several impressive prototypes are already available. However, the emergence of such a novel user input is also calling for novel and dedicated 3D user interfaces. This implies to study the extension of the mental vocabulary available for 3D interaction with VE, then the design of specific 3D interaction techniques “driven by the mind” and, last, the design of immersive sensory feedbacks that could help improving the learning of brain control in VR.
- **Hybrid and collaborative 3D interaction.** Our third research axis intends to study the combination of motor and mental inputs in VR, for one or multiple users. This concerns the design of mixed systems, with potentially collaborative scenarios involving multiple users, and thus, multiple bodies and multiple brains sharing the same VE. This research axis therefore involves two interdependent topics: 1) collaborative virtual environments, and 2) hybrid interaction. It should end up with collaborative virtual environments with multiple users, and shared systems with body and mind inputs.

## ILDA Project-Team

### 3. Research Program

#### 3.1. Introduction

Our ability to acquire or generate, store, process, interlink and query data has increased spectacularly over the last few years. The corresponding advances are commonly grouped under the umbrella of so called *Big Data*. Even if the latter has become a buzzword, these advances are real, and they are having a profound impact in domains as varied as scientific research, commerce, social media, industrial processes or e-government. Yet, looking ahead, emerging technologies related to what we now call the *Web of Data* (a.k.a the Semantic Web) have the potential to create an even larger revolution in data-driven activities, by making information accessible to machines as semistructured data [28] that eventually becomes actionable knowledge. Indeed, novel Web data models considerably ease the interlinking of semi-structured data originating from multiple independent sources. They make it possible to associate machine-processable semantics with the data. This in turn means that heterogeneous systems can exchange data, infer new data using reasoning engines, and that software agents can cross data sources, resolving ambiguities and conflicts between them [71]. Datasets are becoming very rich and very large. They are gradually being made even larger and more heterogeneous, but also much more useful, by interlinking them, as exemplified by the Linked Data initiative [47].

These advances raise research questions and technological challenges that span numerous fields of computer science research: databases, communication networks, security and trust, data mining, as well as human-computer interaction. Our research is based on the conviction that interactive systems play a central role in many data-driven activity domains. Indeed, no matter how elaborate the data acquisition, processing and storage pipelines are, data eventually get processed or consumed one way or another by users. The latter are faced with large, increasingly interlinked heterogeneous datasets (see, e.g., Figure 1 ) that are organized according to complex structures, resulting in overwhelming amounts of both raw data and structured information. Users thus require effective tools to make sense of their data and manipulate them.

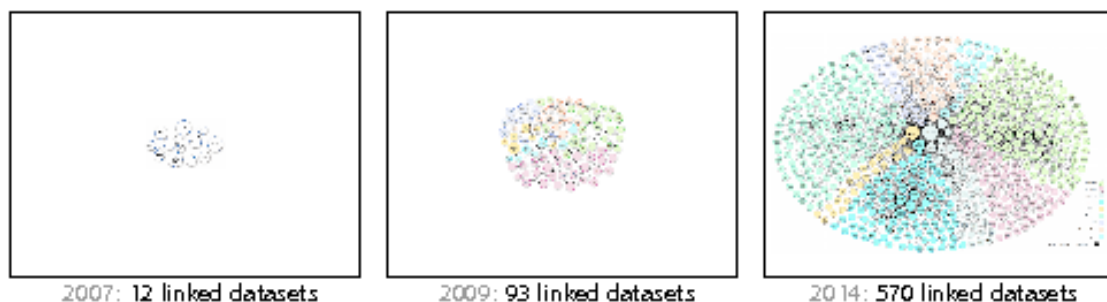


Figure 1. Linking Open Data cloud diagram from 2007 to 2014 – <http://lod-cloud.net>

We approach this problem from the perspective of the Human-Computer Interaction (HCI) field of research, whose goal is to study how humans interact with computers and inspire novel hardware and software designs aimed at optimizing properties such as efficiency, ease of use and learnability, in single-user or cooperative work contexts. More formally, HCI is about designing systems that lower the barrier between users' cognitive model of what they want to accomplish, and computers' understanding of this model. HCI is about the design, implementation and evaluation of computing systems that humans interact with [52], [73]. It is a

highly multidisciplinary field, with experts from computer science, cognitive psychology, design, engineering, ethnography, human factors and sociology.

In this broad context, ILDA aims at designing interactive systems that display [37], [59], [80] the data and let users interact with them, aiming to help users better *navigate* and *comprehend* large webs of data represented visually, as well as *relate* and *manipulate* them.

Our research agenda consists of the three complementary axes detailed in the following subsections. Designing systems that consider interaction in close conjunction with data semantics is pivotal to all three axes. Those semantics will help drive navigation in, and manipulation of, the data, so as to optimize the communication bandwidth between users and data.

## 3.2. Semantics-driven Data Manipulation

**Participants:** Emmanuel Pietriga, Caroline Appert, Hande Ozaygen, Hugo Romat.

The Web of Data has been maturing for the last fifteen years and is starting to gain adoption across numerous application domains (Figure 1 ). Now that most foundational building blocks are in place, from knowledge representation, inference mechanisms and query languages [48], all the way up to the expression of data presentation knowledge [66] and to mechanisms like look-up services [79] or spreading activation [43], we need to pay significant attention to how human beings are going to interact with this new Web, if it is to “*reach its full potential*” [44].

Most efforts in terms of user interface design and development for the Web of data have essentially focused on tools for software developers or subject-matter experts who create ontologies and populate them [54], [42]. Tools more oriented towards end-users are starting to appear [34], [36], [49], [50], [53], [61], including the so-called *linked data browsers* [47]. However, those browsers are in most cases based on quite conventional point-and-click hypertext interfaces that present data to users in a very page-centric, web-of-documents manner that is ill-suited to navigating in, and manipulating, webs of data.

To be successful, interaction paradigms that let users navigate and manipulate data on the Web have to be tailored to the radically different way of browsing information enabled by it, where users directly interact with the data rather than with monolithic documents. The general research question addressed in this part of our research program is how to design novel interaction techniques that help users manipulate their data more efficiently. By data manipulation, we mean all low-level tasks related to manually creating new content, modifying and cleaning existing content, merging data from different sources, establishing connections between datasets, categorizing data, and eventually sharing the end results with other users; tasks that are currently considered quite tedious because of the sheer complexity of the concepts, data models and syntax, and the interplay between all of them.

Our approach is based on the conviction that there is a strong potential for cross-fertilization, as mentioned earlier: on the one hand, user interface design is essential to the management and understanding of webs of data; on the other hand, interlinked datasets enriched with even a small amount of semantics can help create more powerful user interfaces, that provide users with the right information at the right time.

We envision systems that focus on the data themselves, exploiting the underlying *semantics and structure* in the background rather than exposing them – which is what current user interfaces for the Web of Data often do. We envision interactive systems in which the semantics and structure are not exposed directly to users, but serve as input to the system to generate interactive representations that convey information relevant to the task at hand and best afford the possible manipulation actions.

## 3.3. Generalized Multi-scale Navigation

**Participants:** Olivier Chapuis, Emmanuel Pietriga, Caroline Appert, Anastasia Bezerianos, Olivier Gladin, Anna Gogolou, Maria Jesus Lobo Gunther, Arnaud Prouzeau.

The foundational question addressed here is what to display when, where and how, so as to provide effective support to users in their data understanding and manipulation tasks. ILDA targets contexts in which workers have to interact with complementary views on the same data, or with views on different-but-related datasets, possibly at different levels of abstraction. Being able to combine or switch between representations of the data at different levels of detail and merge data from multiple sources in a single representation is central to many scenarios. This is especially true in both of the application domains we consider: mission-critical systems (e.g., natural disaster crisis management) and the exploratory analysis of scientific data (e.g., correlate theories and heterogeneous observational data for an analysis of a given celestial body in Astrophysics).

A significant part of our research over the last ten years has focused on multi-scale interfaces. We designed and evaluated novel interaction techniques, but also worked actively on the development of open-source UI toolkits for multi-scale interfaces (see Section 6.2). These interfaces let users navigate large but relatively homogeneous datasets at different levels of detail, on both workstations [69], [31], [65], [64], [63], [32], [68], [30], [70] and wall-sized displays [5], [55], [67], [60], [33], [39], [38]. This part of the ILDA research program is about extending multi-scale navigation in two directions: 1. Enabling the representation of multiple, spatially-registered but widely varying, multi-scale data layers in Geographical Information Systems (GIS); 2. Generalizing the multi-scale navigation paradigm to interconnected, heterogeneous datasets as found on the Web of Data.

The first research problem is mainly investigated in collaboration with IGN in the context of ANR project MapMuxing (Section 9.2.1), which stands for *multi-dimensional map multiplexing*. Project MapMuxing aims at going beyond the traditional pan & zoom and overview+detail interface schemes, and at designing and evaluating novel cartographic visualizations that rely on high-quality generalization, *i.e.*, the simplification of geographic data to make it legible at a given map scale [76], [77], and symbol specification. Beyond project MapMuxing, we are also investigating multi-scale multiplexing techniques for geo-localized data in the specific context of ultra-high-resolution wall-sized displays, where the combination of a very high pixel density and large physical surface (Figure 2) enable us to explore designs that involve collaborative interaction and physical navigation in front of the workspace. This is work done in cooperation with team Massive Data at Inria Chile.

The second research problem is about the extension of multi-scale navigation to interconnected, heterogeneous datasets. Generalization has a rather straightforward definition in the specific domain of geographical information systems, where data items are geographical entities that naturally aggregate as scale increases. But it is unclear how generalization could work for representations of the more heterogeneous webs of data that we consider in the first axis of our research program. Those data form complex networks of resources with multiple and quite varied relationships between them, that cannot rely on a single, unified type of representation (a role played by maps in GIS applications).

Addressing the limits of current generalization processes is a longer-term, more exploratory endeavor. Here again, the machine-processable semantics and structure of the data give us an opportunity to rethink how users navigate interconnected heterogeneous datasets. Using these additional data, we investigate ways to generalize the multi-scale navigation paradigm to datasets whose layout and spatial relationships can be much richer and much more diverse than what can be encoded with static linear hierarchies as typically found today in interfaces for browsing maps or large imagery. Our goal is thus to design and develop highly dynamic and versatile multi-scale information spaces for heterogeneous data whose structure and semantics are not known in advance, but discovered incrementally.

### 3.4. Novel Forms of Input for Groups and Individuals

**Participants:** Caroline Appert, Anastasia Bezerianos, Olivier Chapuis, Emmanuel Pietriga, André Spritzer, Rafael Morales Gonzalez, Bruno Fruchard.

Analyzing and manipulating large datasets can involve multiple users working together in a coordinated manner in multi-display environments: workstations, handheld devices, wall-sized displays [33]. Those users work towards a common goal, navigating and manipulating data displayed on various hardware surfaces in

a coordinated manner. Group awareness [46], [27] is central in these situations, as users, who may or may not be co-located in the same room, can have an optimal individual behavior only if they have a clear picture of what their collaborators have done and are currently doing in the global context. We work on the design and implementation of interactive systems that improve group awareness in co-located situations [56], making individual users able to figure out what other users are doing without breaking the flow of their own actions.

In addition, users need a rich interaction vocabulary to handle large, structured datasets in a flexible and powerful way, regardless of the context of work. Input devices such as mice and trackpads provide a limited number of input actions, thus requiring users to switch between modes to perform different types of data manipulation and navigation actions. The action semantics of these input devices are also often too much dependent on the display output. For instance, a mouse movement and click can only be interpreted according to the graphical controller (widget) above which it is moved. We focus on designing powerful input techniques based upon technologies such as tactile surfaces (supported by UI toolkits developed in-house), 3D motion tracking systems, or custom-built controllers [58] *to complement (rather than replace) traditional input devices* such as keyboards, that remain the best method so far for text entry, and indirect input devices such as mice or trackpads for pixel-precise pointing actions.

The input vocabularies we investigate enable users to navigate and manipulate large and structured datasets in environments that involve multiple users and displays that vary in their size, position and orientation [33], [45], each having their own characteristics and affordances: wall displays [5], [81], workstations, tabletops [62], [41], tablets [6], [78], smartphones [10], [40], [74], [75], and combinations thereof [2], [9], [60], [33].

We aim at designing rich interaction vocabularies that go far beyond what current touch interfaces offer, which rarely exceeds five gestures such as simple slides and pinches. Designing larger gesture vocabularies requires identifying discriminating dimensions (e.g., the presence or absence of anchor points and the distinction between internal and external frames of reference [6]) in order to structure a space of gestures that interface designers can use as a dictionary for choosing a coherent set of controls. These dimensions should be few and simple, so as to provide users with gestures that are easy to memorize and execute. Beyond gesture complexity, the scalability of vocabularies also depends on our ability to design robust gesture recognizers that will allow users to fluidly chain simple gestures that make it possible to interlace navigation and manipulation actions.

We also plan to study how to further extend input vocabularies by combining touch [10], [6], [62] and mid-air gestures [5] with physical objects [51], [72], [58] and classical input devices such as keyboards to enable users to input commands to the system or to involve other users in their workflow (request for help, delegation, communication of personal findings, etc.) [35], [57]. Gestures and objects encode a lot of information in their shape, dynamics and direction, that can be directly interpreted in relation with the user, independently from the display output. Physical objects can also greatly improve coordination among actors for, e.g., handling priorities or assigning specific roles.



## **IMAGINE Project-Team**

### **3. Research Program**

#### **3.1. Methodology**

As already stressed, thinking of future digital modeling technologies as an Expressive Virtual Pen enabling to seamlessly design, refine and convey animated 3D content, leads to revisit models for shapes, motions and stories from a user-centered perspective. More specifically, inspiring from the user-centered interfaces developed in the Human Computer Interaction domain, we introduced the new concept of user-centered graphical models. Ideally, such models should be designed to behave, under any user action, the way a human user would have predicted. In our case, user's actions may include creation gestures such as sketching to draft a shape or direct a motion, deformation gestures such as stretching a shape in space or a motion in time, or copy-paste gestures to transfer some of the features from existing models to other ones. User-centered graphical models need to incorporate knowledge in order to seamlessly generate the appropriate content from such actions. We are using the following methodology to advance towards these goals:

- Develop high-level models for shapes, motion and stories that embed the necessary knowledge to respond as expected to user actions. These models should provide the appropriate handles for conveying the user's intent while embedding procedural methods that seamlessly take care of the appropriate details and constraints.
- Combine these models with expressive design and control tools such as gesture-based control through sketching, sculpting, or acting, towards interactive environments where users can create a new virtual scene, play with it, edit or refine it, and semi-automatically convey it through a video.

#### **3.2. Validation**

Validation is a major challenge when developing digital creation tools: there is no ideal result to compare with, in contrast with more standard problems such as reconstructing existing shapes or motions. Therefore, we had to think ahead about our validation strategy: new models for geometry or animation can be validated, as usually done in Computer Graphics, by showing that they solve a problem never tackled before or that they provide a more general or more efficient solution than previous methods. The interaction methods we are developing for content creation and editing rely as much as possible on existing interaction design principles already validated within the HCI community. We also occasionally develop new interaction tools, most often in collaboration with this community, and validate them through user studies. Lastly, we work with expert users from various application domains through our collaborations with professional artists, scientists from other domains, and industrial partners: these expert users validate the use of our new tools compared to their usual pipeline.

#### **3.3. Application Domains**

This research can be applied to any situation where users need to create new, imaginary, 3D content. Our work should be instrumental, in the long term, for the visual arts, from the creation of 3D films and games to the development of new digital planning tools for theater or cinema directors. Our models can also be used in interactive prototyping environments for engineering. They can help promoting interactive digital design to scientists, as a tool to quickly express, test and refine models, as well as an efficient way for conveying them to other people. Lastly, we expect our new methodology to put digital modeling within the reach of the general public, enabling educators, media and other practitioners to author their own 3D content.

Our current application domains are:

- Visual arts
  - Modeling and animation for 3D films and games.
  - Virtual cinematography and tools for theater directors.
- Engineering
  - Industrial design.
  - Mechanical & civil engineering.
- Natural Sciences
  - Virtual functional anatomy.
  - Virtual plants.
- Education and Creative tools
  - Sketch-based teaching.
  - Creative environments for novice users.

The diversity of users these domains bring, from digital experts to other professionals and novices, gives us excellent opportunities to validate our general methodology with different categories of users. Our ongoing projects in these various application domains are listed in Section 6.

## MANAO Project-Team

### 3. Research Program

#### 3.1. Related Scientific Domains

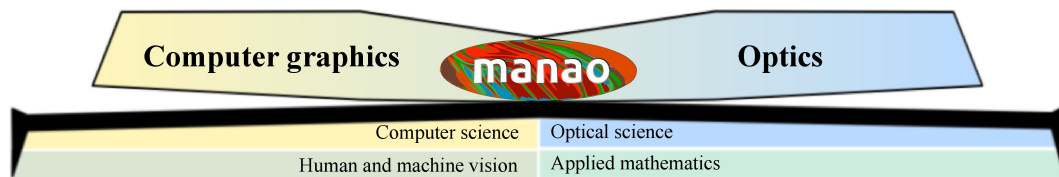


Figure 3. Related scientific domains of the MANAO project.

The *MANAO* project aims at studying, acquiring, modeling, and rendering the interactions between the three components that are light, shape, and matter from the viewpoint of an observer. As detailed more lengthily in the next section, such a work will be done using the following approach: first, we will tend to consider that these three components do not have strict frontiers when considering their impacts on the final observers; then, we will not only work in **computer graphics**, but also at the intersection of computer graphics and **optics**, exploring the mutual benefits that the two domains may provide. It is thus intrinsically a **transdisciplinary** project (as illustrated in Figure 3) and we expect results in both domains.

Thus, the proposed team-project aims at establishing a close collaboration between computer graphics (e.g., 3D modeling, geometry processing, shading techniques, vector graphics, and GPU programming) and optics (e.g., design of optical instruments, and theories of light propagation). The following examples illustrate the strengths of such a partnership. First, in addition to simpler radiative transfer equations [46] commonly used in computer graphics, research in the later will be based on state-of-the-art understanding of light propagation and scattering in real environments. Furthermore, research will rely on appropriate instrumentation expertise for the measurement [59], [60] and display [58] of the different phenomena. Reciprocally, optics researches may benefit from the expertise of computer graphics scientists on efficient processing to investigate interactive simulation, visualization, and design. Furthermore, new systems may be developed by unifying optical and digital processing capabilities. Currently, the scientific background of most of the team members is related to computer graphics and computer vision. A large part of their work have been focused on simulating and analyzing optical phenomena as well as in acquiring and visualizing them. Combined with the close collaboration with the optics laboratory LP2N (<http://www.lp2n.fr>) and with the students issued from the “Institut d’Optique” (<http://www.institutoptique.fr>), this background ensures that we can expect the following results from the project: the construction of a common vocabulary for tightening the collaboration between the two scientific domains and creating new research topics. By creating this context, we expect to attract (and even train) more trans-disciplinary researchers.

At the boundaries of the *MANAO* project lie issues in **human and machine vision**. We have to deal with the former whenever a human observer is taken into account. On one side, computational models of human vision are likely to guide the design of our algorithms. On the other side, the study of interactions between light, shape, and matter may shed some light on the understanding of visual perception. The same kind of connections are expected with machine vision. On the one hand, traditional computational methods for acquisition (such as photogrammetry) are going to be part of our toolbox. On the other hand, new display technologies (such as the ones used for augmented reality) are likely to benefit from our integrated approach

and systems. In the *MANAO* project we are mostly users of results from human vision. When required, some experimentation might be done in collaboration with experts from this domain, like with the European PRISM project. For machine vision, provided the tight collaboration between optical and digital systems, research will be carried out inside the *MANAO* project.

Analysis and modeling rely on **tools from applied mathematics** such as differential and projective geometry, multi-scale models, frequency analysis [48] or differential analysis [85], linear and non-linear approximation techniques, stochastic and deterministic integrations, and linear algebra. We not only rely on classical tools, but also investigate and adapt recent techniques (e.g., improvements in approximation techniques), focusing on their ability to run on modern hardware: the development of our own tools (such as Eigen, see Section 6.3) is essential to control their performances and their abilities to be integrated into real-time solutions or into new instruments.

## 3.2. Research axes

The *MANAO* project is organized around four research axes that cover the large range of expertise of its members and associated members. We briefly introduce these four axes in this section. More details and their inter-influences that are illustrated in the Figure 2 will be given in the following sections.

Axis 1 is the theoretical foundation of the project. Its main goal is to increase the understanding of light, shape, and matter interactions by combining expertise from different domains: optics and human/machine vision for the analysis and computer graphics for the simulation aspect. The goal of our analyses is to identify the different layers/phenomena that compose the observed signal. In a second step, the development of physical simulations and numerical models of these identified phenomena is a way to validate the pertinence of the proposed decompositions.

In Axis 2, the final observers are mainly physical captors. Our goal is thus the development of new acquisition and display technologies that combine optical and digital processes in order to reach fast transfers between real and digital worlds, in order to increase the convergence of these two worlds.

Axes 3 and 4 focus on two aspects of computer graphics: rendering, visualization and illustration in Axis 3, and editing and modeling (content creation) in Axis 4. In these two axes, the final observers are mainly human users, either generic users or expert ones (e.g., archaeologist [89], computer graphics artists).

## 3.3. Axis 1: Analysis and Simulation

**Challenge:** Definition and understanding of phenomena resulting from interactions between light, shape, and matter as seen from an observer point of view.

**Results:** Theoretical tools and numerical models for analyzing and simulating the observed optical phenomena.

To reach the goals of the *MANAO* project, we need to **increase our understanding** of how light, shape, and matter act together in synergy and how the resulting signal is finally observed. For this purpose, we need to identify the different phenomena that may be captured by the targeted observers. This is the main objective of this research axis, and it is achieved by using three approaches: the simulation of interactions between light, shape, and matter, their analysis and the development of new numerical models. This resulting improved knowledge is a foundation for the researches done in the three other axes, and the simulation tools together with the numerical models serve the development of the joint optical/digital systems in Axis 2 and their validation.

One of the main and earliest goals in computer graphics is to faithfully reproduce the real world, focusing mainly on light transport. Compared to researchers in physics, researchers in computer graphics rely on a subset of physical laws (mostly radiative transfer and geometric optics), and their main concern is to efficiently use the limited available computational resources while developing as fast as possible algorithms. For this purpose, a large set of theoretical as well as computational tools has been introduced to take a **maximum benefit of hardware** specificities. These tools are often dedicated to specific phenomena (e.g., direct or indirect lighting, color bleeding, shadows, caustics). An efficiency-driven approach needs such a classification

of light paths [55] in order to develop tailored strategies [101]. For instance, starting from simple direct lighting, more complex phenomena have been progressively introduced: first diffuse indirect illumination [53], [93], then more generic inter-reflections [62], [46] and volumetric scattering [90], [43]. Thanks to this search for efficiency and this classification, researchers in computer graphics have developed a now recognized expertise in fast-simulation of light propagation. Based on finite elements (radiosity techniques) or on unbiased Monte Carlo integration schemes (ray-tracing, particle-tracing, ...), the resulting algorithms and their combination are now sufficiently accurate to be used-back in physical simulations. The MANAO project will continue the search for **efficient and accurate simulation** techniques, but extending it from computer graphics to optics. Thanks to the close collaboration with scientific researchers from optics, new phenomena beyond radiative transfer and geometric optics will be explored.

Search for algorithmic efficiency and accuracy has to be done in parallel with **numerical models**. The goal of visual fidelity (generalized to accuracy from an observer point of view in the project) combined with the goal of efficiency leads to the development of alternative representations. For instance, common classical finite-element techniques compute only basis coefficients for each discretization element: the required discretization density would be too large and to computationally expensive to obtain detailed spatial variations and thus visual fidelity. Examples includes texture for decorrelating surface details from surface geometry and high-order wavelets for a multi-scale representation of lighting [42]. The numerical complexity explodes when considering directional properties of light transport such as radiance intensity (Watt per square meter and per steradian -  $W.m^{-2}.sr^{-1}$ ), reducing the possibility to simulate or accurately represent some optical phenomena. For instance, Haar wavelets have been extended to the spherical domain [92] but are difficult to extend to non-piecewise-constant data [95]. More recently, researches prefer the use of Spherical Radial Basis Functions [98] or Spherical Harmonics [84]. For more complex data, such as reflective properties (e.g., BRDF [77], [63] - 4D), ray-space (e.g., Light-Field [73] - 4D), spatially varying reflective properties (6D - [88]), new models, and representations are still investigated such as rational functions [80] or dedicated models [31] and parameterizations [91], [96]. For each (newly) defined phenomena, we thus explore the space of possible numerical representations to determine the **most suited one for a given application**, like we have done for BRDF [80].

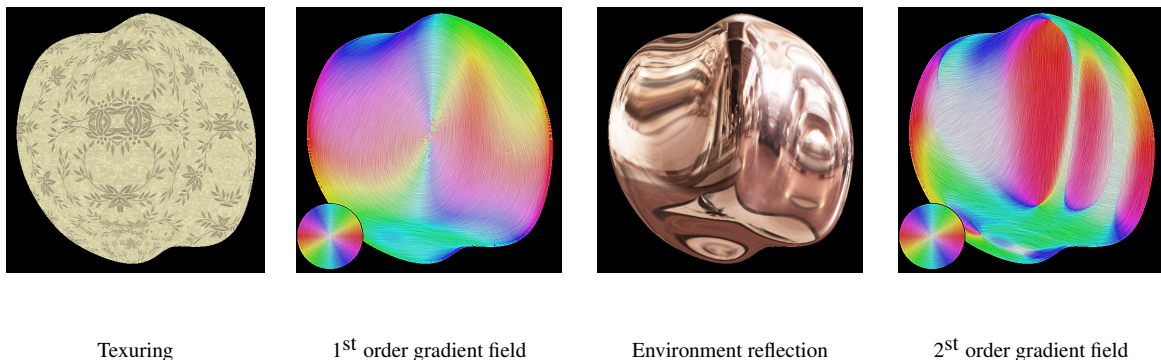


Figure 4. First-order analysis [102] have shown that shading variations are caused by depth variations (first-order gradient field) and by normal variations (second-order fields). These fields are visualized using hue and saturation to indicate direction and magnitude of the flow respectively.

Before being able to simulate or to represent the different **observed phenomena**, we need to define and describe them. To understand the difference between an observed phenomenon and the classical light, shape, and matter decomposition, we can take the example of a highlight. Its observed shape (by a human user or a sensor) is the resulting process of the interaction of these three components, and can be simulated this way. However, this does not provide any intuitive understanding of their relative influence on the final shape: an artist will directly describe the resulting shape, and not each of the three properties. We thus want to

decompose the observed signal into models for each scale that can be easily understandable, representable, and manipulable. For this purpose, we will rely on the **analysis** of the resulting interaction of light, shape, and matter as observed by a human or a physical sensor. We first consider this analysis from an **optical point of view**, trying to identify the different phenomena and their scale according to their mathematical properties (e.g., differential [85] and frequency analysis [48]). Such an approach has led us to exhibit the influence of surfaces flows (depth and normal gradients) into lighting pattern deformation (see Figure 4). For a **human observer**, this correspond to one recent trend in computer graphics that takes into account the human visual systems [49] both to evaluate the results and to guide the simulations.

### 3.4. Axis 2: From Acquisition to Display

**Challenge:** Convergence of optical and digital systems to blend real and virtual worlds.

**Results:** Instruments to acquire real world, to display virtual world, and to make both of them interact.

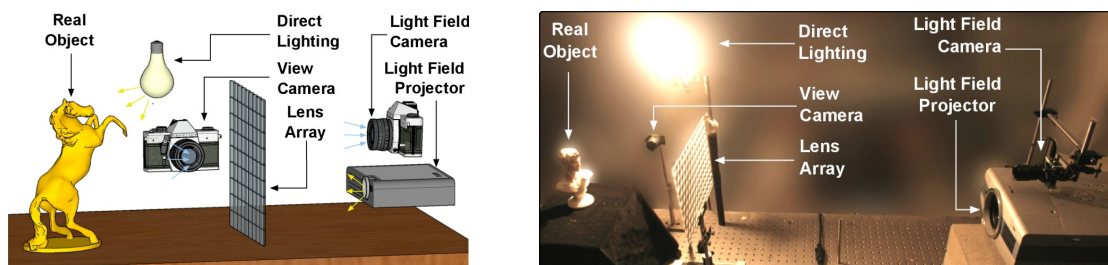


Figure 5. Light-Field transfer: global illumination between real and synthetic objects [41]

In this axis, we investigate *unified acquisition and display systems*, that is systems which combine optical instruments with digital processing. From digital to real, we investigate new display approaches [73], [58]. We consider projecting systems and surfaces [38], for personal use, virtual reality and augmented reality [33]. From the real world to the digital world, we favor direct measurements of parameters for models and representations, using (new) optical systems unless digitization is required [52], [51]. These resulting systems have to acquire the different phenomena described in Axis 1 and to display them, in an efficient manner [56], [32], [57], [60]. By efficient, we mean that we want to shorten the path between the real world and the virtual world by increasing the data bandwidth between the real (analog) and the virtual (digital) worlds, and by reducing the latency for real-time interactions (we have to prevent unnecessary conversions, and to reduce processing time). To reach this goal, the systems have to be designed as a whole, not by a simple concatenation of optical systems and digital processes, nor by considering each component independently [61].

To increase data bandwidth, one solution is to **parallelize more and more the physical systems**. One possible solution is to multiply the number of simultaneous acquisitions (e.g., simultaneous images from multiple viewpoints [60], [82]). Similarly, increasing the number of viewpoints is a way toward the creation of full 3D displays [73]. However, full acquisition or display of 3D real environments theoretically requires a continuous field of viewpoints, leading to huge data size. Despite the current belief that the increase of computational power will fill the missing gap, when it comes to visual or physical realism, if you double the processing power, people may want four times more accuracy, thus increasing data size as well. To reach the best performances, a trade-off has to be found between the amount of data required to represent accurately the reality and the amount of required processing. This trade-off may be achieved using **compressive sensing**. Compressive sensing is a new trend issued from the applied mathematics community that provides tools to accurately reconstruct a signal from a small set of measurements assuming that it is sparse in a transform domain (e.g., [81], [107]).

We prefer to achieve this goal by avoiding as much as possible the classical approach where acquisition is followed by a fitting step: this requires in general a large amount of measurements and the fitting itself may consume consequently too much memory and preprocessing time. By **preventing unnecessary conversion** through fitting techniques, such an approach increase the speed and reduce the data transfer for acquisition but also for display. One of the best recent examples is the work of Cossairt et al. [41]. The whole system is designed around a unique representation of the energy-field issued from (or leaving) a 3D object, either virtual or real: the Light-Field. A Light-Field encodes the light emitted in any direction from any position on an object. It is acquired thanks to a lens-array that leads to the capture of, and projection from, multiple simultaneous viewpoints. A unique representation is used for all the steps of this system. Lens-arrays, parallax barriers, and coded-aperture [70] are one of the key technologies to develop such acquisition (e.g., Light-Field camera<sup>0</sup> [61] and acquisition of light-sources [52]), projection systems (e.g., auto-stereoscopic displays). Such an approach is versatile and may be applied to improve classical optical instruments [68]. More generally, by designing unified optical and digital systems [78], it is possible to leverage the requirement of processing power, the memory footprint, and the cost of optical instruments.

Those are only some examples of what we investigate. We also consider the following approaches to develop new unified systems. First, similar to (and based on) the analysis goal of Axis 1, we have to take into account as much as possible the characteristics of the measurement setup. For instance, when fitting cannot be avoided, integrating them may improve both the processing efficiency and accuracy [80]. Second, we have to integrate signals from multiple sensors (such as GPS, accelerometer, ...) to prevent some computation (e.g., [71]). Finally, the experience of the group in surface modeling help the design of optical surfaces [64] for light sources or head-mounted displays.

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

**Challenge:** How to offer the most legible signal to the final observer in real-time?

**Results:** High-level shading primitives, expressive rendering techniques for object depiction, real-time realistic rendering algorithms

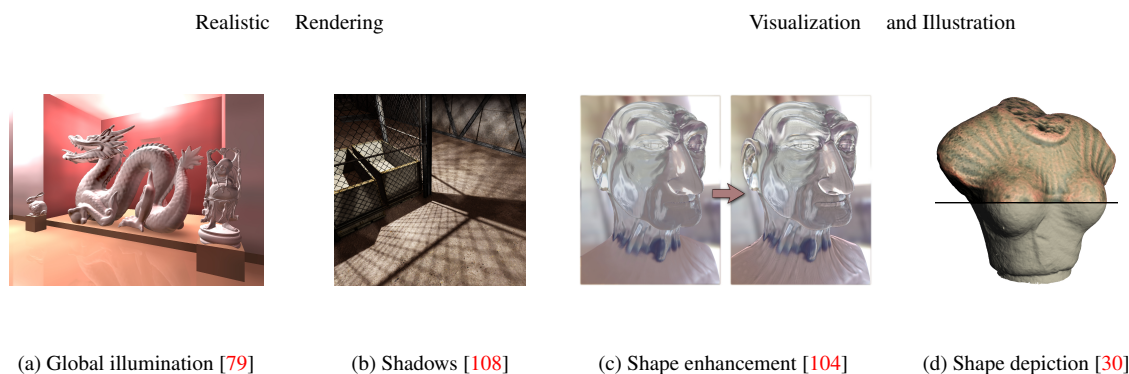


Figure 6. In the MANAO project, we are investigating rendering techniques from realistic solutions (e.g., inter-reflections (a) and shadows (b)) to more expressive ones (shape enhancement (c) with realistic style and shape depiction (d) with stylized style) for visualization.

The main goal of this axis is to offer to the final observer, in this case mostly a human user, the most legible signal in real-time. Thanks to the analysis and to the decomposition in different phenomena resulting from interactions between light, shape, and matter (Axis 1), and their perception, we can use them to convey essential information in the most pertinent way. Here, the word *pertinent* can take various forms depending on the application.

<sup>0</sup>Lytro, <http://www.lytro.com/>

In the context of scientific illustration and visualization, we are primarily interested in tools to convey shape or material characteristics of objects in animated 3D scenes. **Expressive rendering** techniques (see Figure 6 c,d) provide means for users to depict such features with their own style. To introduce our approach, we detail it from a shape-depiction point of view, domain where we have acquired a recognized expertise. Prior work in this area mostly focused on stylization primitives to achieve line-based rendering [105], [67] or stylized shading [36], [104] with various levels of abstraction. A clear representation of important 3D **object features** remains a major challenge for better shape depiction, stylization and abstraction purposes. Most existing representations provide only local properties (e.g., curvature), and thus lack characterization of broader shape features. To overcome this limitation, we are developing higher level descriptions of shape [29] with increased robustness to sparsity, noise, and outliers. This is achieved in close collaboration with Axis 1 by the use of higher-order local fitting methods, multi-scale analysis, and global regularization techniques. In order not to neglect the observer and the material characteristics of the objects, we couple this approach with an analysis of the appearance model. To our knowledge, this is an approach which has not been considered yet. This research direction is at the heart of the *MANAO* project, and has a strong connection with the analysis we plan to conduct in Axis 1. Material characteristics are always considered at the light ray level, but an understanding of **higher-level primitives** (like the shape of highlights and their motion) would help us to produce more legible renderings and permit novel stylizations; for instance, there is no method that is today able to create stylized renderings that follow the motion of highlights or shadows. We also believe such tools also play a fundamental role for geometry processing purposes (such as shape matching, reassembly, simplification), as well as for editing purposes as discussed in Axis 4.

In the context of **real-time photo-realistic rendering** (see Figure 6 a,b), the challenge is to compute the most plausible images with minimal effort. During the last decade, a lot of work has been devoted to design approximate but real-time rendering algorithms of complex lighting phenomena such as soft-shadows [106], motion blur [48], depth of field [94], reflexions, refractions, and inter-reflexions. For most of these effects it becomes harder to discover fundamentally new and faster methods. On the other hand, we believe that significant speedup can still be achieved through more clever use of **massively parallel architectures** of the current and upcoming hardware, and/or through more clever tuning of the current algorithms. In particular, regarding the second aspect, we remark that most of the proposed algorithms depend on several parameters which can be used to **trade the speed over the quality**. Significant speed-up could thus be achieved by identifying effects that would be masked or facilitated and thus devote appropriate computational resources to the rendering [69], [47]. Indeed, the algorithm parameters controlling the quality vs speed are numerous without a direct mapping between their values and their effect. Moreover, their ideal values vary over space and time, and to be effective such an auto-tuning mechanism has to be extremely fast such that its cost is largely compensated by its gain. We believe that our various work on the analysis of the appearance such as in Axis 1 could be beneficial for such purpose too.

Realistic and real-time rendering is closely related to Axis 2: real-time rendering is a requirement to close the loop between real world and digital world. We have to thus develop algorithms and rendering primitives that allow the integration of the acquired data into real-time techniques. We have also to take care of that these real-time techniques have to work with new display systems. For instance, stereo, and more generally multi-view displays are based on the multiplication of simultaneous images. Brute force solutions consist in independent rendering pipeline for each viewpoint. A more energy-efficient solution would take advantages of the computation parts that may be factorized. Another example is the rendering techniques based on image processing, such as our work on augmented reality [40]. Independent image processing for each viewpoint may disturb the feeling of depth by introducing inconsistent information in each images. Finally, more dedicated displays [58] would require new rendering pipelines.

### 3.6. Axis 4: Editing and Modeling

**Challenge:** Editing and modeling appearance using drawing- or sculpting-like tools through high level representations.

**Results:** High-level primitives and hybrid representations for appearance and shape.



During the last decade, the domain of computer graphics has exhibited tremendous improvements in image quality, both for 2D applications and 3D engines. This is mainly due to the availability of an ever increasing amount of shape details, and sophisticated appearance effects including complex lighting environments. Unfortunately, with such a growth in visual richness, even so-called *vectorial* representations (e.g., subdivision surfaces, Bézier curves, gradient meshes, etc.) become very dense and unmanageable for the end user who has to deal with a huge mass of control points, color labels, and other parameters. This is becoming a major challenge, with a necessity for novel representations. This Axis is thus complementary of Axis 3: the focus is the development of primitives that are easy to use for modeling and editing.

More specifically, we plan to investigate *vectorial representations* that would be amenable to the production of rich shapes with a minimal set of primitives and/or parameters. To this end we plan to build upon our insights on dynamic local reconstruction techniques and implicit surfaces [4] [35]. When working in 3D, an interesting approach to produce detailed shapes is by means of procedural geometry generation. For instance, many natural phenomena like waves or clouds may be modeled using a combination of procedural functions. Turning such functions into triangle meshes (main rendering primitives of GPUs) is a tedious process that appears not to be necessary with an adapted vectorial shape representation where one could directly turn procedural functions into implicit geometric primitives. Since we want to prevent unnecessary conversions in the whole pipeline (here, between modeling and rendering steps), we will also consider *hybrid representations* mixing meshes and implicit representations. Such research has thus to be conducted while considering the associated editing tools as well as performance issues. It is indeed important to keep *real-time performance* (cf. Axis 2) throughout the interaction loop, from user inputs to display, via editing and rendering operations. Finally, it would be interesting to add *semantic information* into 2D or 3D geometric representations. Semantic geometry appears to be particularly useful for many applications such as the design of more efficient manipulation and animation tools, for automatic simplification and abstraction, or even for automatic indexing and searching. This constitutes a complementary but longer term research direction.

In the *MANAO* project, we want to investigate representations beyond the classical light, shape, and matter decomposition. We thus want to directly control the appearance of objects both in 2D and 3D applications (e.g., [99]): this is a core topic of computer graphics. When working with 2D vector graphics, digital artists must carefully set up color gradients and textures: examples range from the creation of 2D logos to the photo-realistic imitation of object materials. Classic vector primitives quickly become impractical for creating illusions of complex materials and illuminations, and as a result an increasing amount of time and skill is required. This is only for still images. For animations, vector graphics are only used to create legible appearances composed of simple lines and color gradients. There is thus a need for more complex primitives that are able to accommodate complex reflection or texture patterns, while keeping the ease of use of vector graphics. For instance, instead of drawing color gradients directly, it is more advantageous to draw flow lines that represent local surface concavities and convexities. Going through such an intermediate structure then allows to deform simple material gradients and textures in a coherent way (see Figure 7), and animate them all at once. The manipulation of 3D object materials also raises important issues. Most existing material models are tailored to faithfully reproduce physical behaviors, not to be *easily controllable* by artists. Therefore artists learn to tweak model parameters to satisfy the needs of a particular shading appearance, which can quickly become cumbersome as the complexity of a 3D scene increases. We believe that an alternative approach is required, whereby material appearance of an object in a typical lighting environment is directly input (e.g., painted or drawn), and adapted to match a plausible material behavior. This way, artists will be able to create their own appearance (e.g., by using our shading primitives [99]), and replicate it to novel illumination environments and 3D models. For this purpose, we will rely on the decompositions and tools issued from Axis 1.

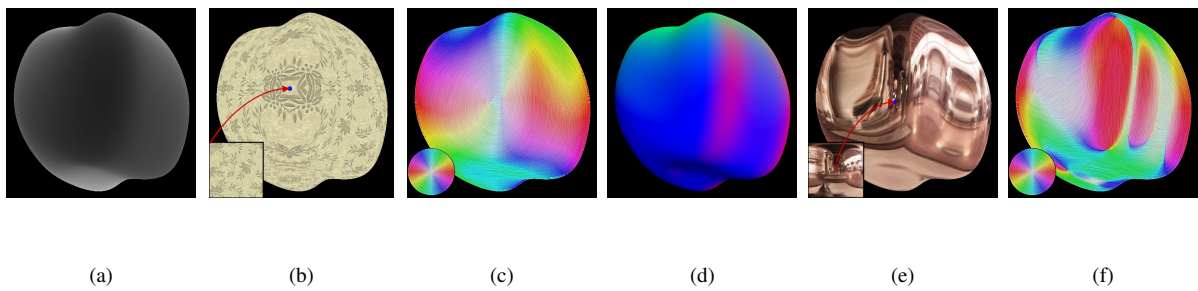


Figure 7. Based on our analysis [102] (Axis 1), we have designed a system that mimics texture (left) and shading (right) effects using image processing alone. It takes depth (a) and normal (d) images as input, and uses them to deform images (b-e) in ways that closely approximate surface flows (c-f). It provides a convincing, yet artistically controllable illusion of 3D shape conveyed through texture or shading cues.

## MAVERICK Project-Team

### 3. Research Program

#### 3.1. Introduction

The Maverick project-team aims at producing representations and algorithms for efficient, high-quality computer generation of pictures and animations through the study of four **research problems**:

- *Computer Visualization* where we take as input a large localized dataset and represent it in a way that will let an observer understand its key properties. Visualization can be used for data analysis, for the results of a simulation, for medical imaging data...
- *Expressive Rendering*, where we create an artistic representation of a virtual world. Expressive rendering corresponds to the generation of drawings or paintings of a virtual scene, but also to some areas of computational photography, where the picture is simplified in specific areas to focus the attention.
- *Illumination Simulation*, where we model the interaction of light with the objects in the scene, resulting in a photorealistic picture of the scene. Research include improving the quality and photorealism of pictures, including more complex effects such as depth-of-field or motion-blur. We are also working on accelerating the computations, both for real-time photorealistic rendering and offline, high-quality rendering.
- *Complex Scenes*, where we generate, manage, animate and render highly complex scenes, such as natural scenes with forests, rivers and oceans, but also large datasets for visualization. We are especially interested in interactive visualization of complex scenes, with all the associated challenges in terms of processing and memory bandwidth.

The fundamental research interest of Maverick is first, *understanding* what makes a picture useful, powerful and interesting for the user, and second *designing* algorithms to create and improve these pictures.

#### 3.2. Research approaches

We will address these research problems through three interconnected research approaches:

##### 3.2.1. *Picture Impact*

Our first research axis deals with the *impact* pictures have on the viewer, and how we can improve this impact. Our research here will target:

- *evaluating user response*: we need to evaluate how the viewers respond to the pictures and animations generated by our algorithms, through user studies, either asking the viewer about what he perceives in a picture or measuring how his body reacts (eye tracking, position tracking).
- *removing artefacts and discontinuities*: temporal and spatial discontinuities perturb viewer attention, distracting the viewer from the main message. These discontinuities occur during the picture creation process; finding and removing them is a difficult process.

##### 3.2.2. *Data Representation*

The data we receive as input for picture generation is often unsuitable for interactive high-quality rendering: too many details, no spatial organisation... Similarly the pictures we produce or get as input for other algorithms can contain superfluous details.

One of our goals is to develop new data representations, adapted to our requirements for rendering. This includes fast access to the relevant information, but also access to the specific hierarchical level of information needed: we want to organize the data in hierarchical levels, pre-filter it so that sampling at a given level also gives information about the underlying levels. Our research for this axis include filtering, data abstraction, simplification and stylization.

The input data can be of any kind: geometric data, such as the model of an object, scientific data before visualization, pictures and photographs. It can be time-dependent or not; time-dependent data bring an additional level of challenge on the algorithm for fast updates.

### 3.2.3. Prediction and simulation

Our algorithms for generating pictures require computations: sampling, integration, simulation... These computations can be optimized if we already know the characteristics of the final picture. Our recent research has shown that it is possible to predict the local characteristics of a picture by studying the phenomena involved: the local complexity, the spatial variations, their direction...

Our goal is to develop new techniques for predicting the properties of a picture, and to adapt our image-generation algorithms to these properties, for example by sampling less in areas of low variation.

Our research problems and approaches are all cross-connected. Research on the *impact* of pictures is of interest in three different research problems: *Computer Visualization*, *Expressive rendering* and *Illumination Simulation*. Similarly, our research on *Illumination simulation* will use all three research approaches: impact, representations and prediction.

## 3.3. Cross-cutting research issues

Beyond the connections between our problems and research approaches, we are interested in several issues, which are present throughout all our research:

**sampling** is an ubiquitous process occurring in all our application domains, whether photorealistic rendering (*e.g.* photon mapping), expressive rendering (*e.g.* brush strokes), texturing, fluid simulation (Lagrangian methods), etc. When sampling and reconstructing a signal for picture generation, we have to ensure both coherence and homogeneity. By *coherence*, we mean not introducing spatial or temporal discontinuities in the reconstructed signal. By *homogeneity*, we mean that samples should be placed regularly in space and time. For a time-dependent signal, these requirements are conflicting with each other, opening new areas of research.

**filtering** is another ubiquitous process, occurring in all our application domains, whether in realistic rendering (*e.g.* for integrating height fields, normals, material properties), expressive rendering (*e.g.* for simplifying strokes), textures (through non-linearity and discontinuities). It is especially relevant when we are replacing a signal or data with a lower resolution (for hierarchical representation); this involves filtering the data with a reconstruction kernel, representing the transition between levels.

**performance and scalability** are also a common requirement for all our applications. We want our algorithms to be usable, which implies that they can be used on large and complex scenes, placing a great importance on scalability. For some applications, we target interactive and real-time applications, with an update frequency between 10 Hz and 120 Hz.

**coherence and continuity** in space and time is also a common requirement of realistic as well as expressive models which must be ensured despite contradictory requirements. We want to avoid flickering and aliasing.

**animation:** our input data is likely to be time-varying (*e.g.* animated geometry, physical simulation, time-dependent dataset). A common requirement for all our algorithms and data representation is that they must be compatible with animated data (fast updates for data structures, low latency algorithms...).

## 3.4. Methodology

Our research is guided by several methodological principles:

**Experimentation:** to find solutions and phenomenological models, we use experimentation, performing statistical measurements of how a system behaves. We then extract a model from the experimental data.

**Validation:** for each algorithm we develop, we look for experimental validation: measuring the behavior of the algorithm, how it scales, how it improves over the state-of-the-art... We also compare our algorithms to the exact solution. Validation is harder for some of our research domains, but it remains a key principle for us.

**Reducing the complexity of the problem:** the equations describing certain behaviors in image synthesis can have a large degree of complexity, precluding computations, especially in real time. This is true for physical simulation of fluids, tree growth, illumination simulation... We are looking for *emerging phenomena* and *phenomenological models* to describe them (see framed box “Emerging phenomena”). Using these, we simplify the theoretical models in a controlled way, to improve user interaction and accelerate the computations.

**Transferring ideas from other domains:** Computer Graphics is, by nature, at the interface of many research domains: physics for the behavior of light, applied mathematics for numerical simulation, biology, algorithmics... We import tools from all these domains, and keep looking for new tools and ideas.

**Develop new fundamental tools:** In situations where specific tools are required for a problem, we will proceed from a theoretical framework to develop them. These tools may in return have applications in other domains, and we are ready to disseminate them.

**Collaborate with industrial partners:** we have a long experiment of collaboration with industrial partners. These collaborations bring us new problems to solve, with short-term or medium-term transfert opportunities. When we cooperate with these partners, we have to find *what they need*, which can be very different from *what they want*, their expressed need.

## MIMETIC Project-Team

### 3. Research Program

#### 3.1. Biomechanics and Motion Control

Human motion control is a very complex phenomenon that involves several layered systems, as shown in Figure 3. Each layer of this controller is responsible for dealing with perceptual stimuli in order to decide the actions that should be applied to the human body and his environment. Due to the intrinsic complexity of the information (internal representation of the body and mental state, external representation of the environment) used to perform this task, it is almost impossible to model all the possible states of the system. Even for simple problems, there generally exists an infinity of solutions. For example, from the biomechanical point of view, there are much more actuators (i.e. muscles) than degrees of freedom leading to an infinity of muscle activation patterns for a unique joint rotation. From the reactive point of view there exists an infinity of paths to avoid a given obstacle in navigation tasks. At each layer, the key problem is to understand how people select one solution among these infinite state spaces. Several scientific domains have addressed this problem with specific points of view, such as physiology, biomechanics, neurosciences and psychology.

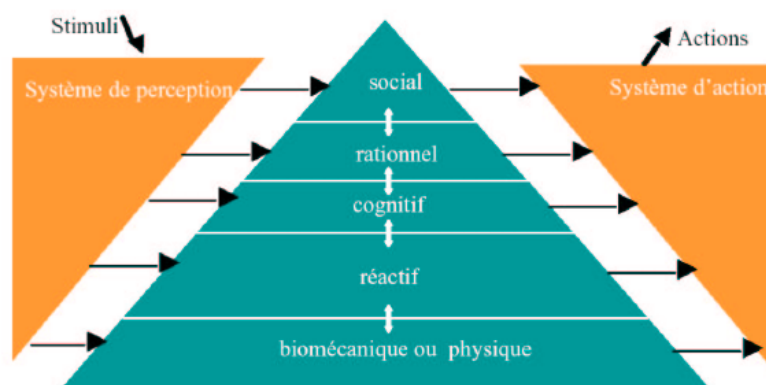


Figure 3. Layers of the motion control natural system in humans.

In biomechanics and physiology, researchers have proposed hypotheses based on accurate joint modeling (to identify the real anatomical rotational axes), energy minimization, force and torques minimization, comfort maximization (i.e. avoiding joint limits), and physiological limitations in muscle force production. All these constraints have been used in optimal controllers to simulate natural motions. The main problem is thus to define how these constraints are composed altogether such as searching the weights used to linearly combine these criteria in order to generate a natural motion. Musculoskeletal models are stereotyped examples for which there exists an infinity of muscle activation patterns, especially when dealing with antagonist muscles. An unresolved problem is to define how to use the above criteria to retrieve the actual activation patterns, while optimization approaches still leads to unrealistic ones. It is still an open problem that will require multidisciplinary skills including computer simulation, constraint solving, biomechanics, optimal control, physiology and neurosciences.

In neurosciences, researchers have proposed other theories, such as coordination patterns between joints driven by simplifications of the variables used to control the motion. The key idea is to assume that instead of controlling all the degrees of freedom, people control higher level variables which correspond to combinations of joint angles. In walking, data reduction techniques such as Principal Component Analysis have shown that lower-limb joint angles are generally projected on a unique plane whose angle in the state space is associated with energy expenditure. Although knowledge exists for specific motions, such as locomotion or grasping, this type of approach is still difficult to generalize. The key problem is that many variables are coupled and it is very difficult to objectively study the behavior of a unique variable in various motor tasks. Computer simulation is a promising method to evaluate such type of assumptions as it enables to accurately control all the variables and to check if it leads to natural movements.

Neurosciences also address the problem of coupling perception and action by providing control laws based on visual cues (or any other senses), such as determining how the optical flow is used to control direction in navigation tasks, while dealing with collision avoidance or interception. Coupling of the control variables is enhanced in this case as the state of the body is enriched by the large amount of external information that the subject can use. Virtual environments inhabited with autonomous characters whose behavior is driven by motion control assumptions is a promising approach to solve this problem. For example, an interesting problem in this field is navigation in an environment inhabited with other people. Typically, avoiding static obstacles together with other people displacing into the environment is a combinatory problem that strongly relies on the coupling between perception and action.

One of the main objectives of MimeTIC is to enhance knowledge on human motion control by developing innovative experiments based on computer simulation and immersive environments. To this end, designing experimental protocols is a key point and some of the researchers in MimeTIC have developed this skill in biomechanics and perception-action coupling. Associating these researchers to experts in virtual human simulation, computational geometry and constraints solving enable us to contribute to enhance fundamental knowledge in human motion control.

### **3.2. Experiments in Virtual Reality**

Understanding interactions between humans is very challenging because it addresses many complex phenomena including perception, decision-making, cognition and social behaviors. Moreover, all these phenomena are difficult to isolate in real situations, and it is therefore very complex to understand their individual influence on these human interactions. It is then necessary to find an alternative solution that can standardize the experiments and that allows the modification of only one parameter at a time. Video was first used since the displayed experiment is perfectly repeatable and cut-offs (stop the video at a specific time before its end) allow having temporal information. Nevertheless, the absence of adapted viewpoint and stereoscopic vision does not provide depth information that are very meaningful. Moreover, during video recording session, the real human is acting in front of a camera and not of an opponent. The interaction is then not a real interaction between humans.

Virtual Reality (VR) systems allow full standardization of the experimental situations and the complete control of the virtual environment. It is then possible to modify only one parameter at a time and to observe its influence on the perception of the immersed subject. VR can then be used to understand what information is picked up to make a decision. Moreover, cut-offs can also be used to obtain temporal information about when information is picked up. When the subject can moreover react as in a real situation, his movement (captured in real time) provides information about his reactions to the modified parameter. Not only is the perception studied, but the complete perception-action loop. Perception and action are indeed coupled and influence each other as suggested by Gibson in 1979.

Finally, VR allows the validation of the virtual human models. Some models are indeed based on the interaction between the virtual character and the other humans, such as a walking model. In that case, there are two ways to validate it. First, they can be compared to real data (e.g. real trajectories of pedestrians). But such data are not always available and are difficult to get. The alternative solution is then to use VR. The validation of the realism of the model is then done by immersing a real subject in a virtual environment in which a virtual

character is controlled by the model. Its evaluation is then deduced from how the immersed subject reacts when interacting with the model and how realistic he feels the virtual character is.

### **3.3. Computational Geometry**

Computational geometry is a branch of computer science devoted to the study of algorithms which can be stated in terms of geometry. It aims at studying algorithms for combinatorial, topological and metric problems concerning sets of points in Euclidian spaces. Combinatorial computational geometry focuses on three main problem classes: static problems, geometric query problems and dynamic problems.

In static problems, some inputs are given and the corresponding outputs need to be constructed or found. Such problems include linear programming, Delaunay triangulations, and Euclidian shortest paths for instance. In geometric query problems, commonly known as geometric search problems, the input consists of two parts: the search space part and the query part, which varies over the problem instances. The search space typically needs to be preprocessed, in a way that multiple queries can be answered efficiently. Some typical problems are range searching, point location in a portioned space, or nearest neighbor queries. In dynamic problems, the goal is to find an efficient algorithm for finding a solution repeatedly after each incremental modification of the input data (addition, deletion or motion of input geometric elements). Algorithms for problems of this type typically involve dynamic data structures. Both of previous problem types can be converted into a dynamic problem, for instance, maintaining a Delaunay triangulation between moving points.

In this context, distance geometry relies solely on distances, instead of points and lines, as in classical geometry. Various applications lead to the definition of problems that can be formulated as a distance geometry, including sensor network localization, robot coordination, the identification of molecular conformations, or as in the context of MimeTIC relations between objects in virtual scenes (e.g., distances between body segments, agents, or cameras). In recent years, scientific research has been oriented to the assumptions allowing for discretizing the search space of a given distance geometry problem. The discretization (which is exact in some situations) allows to conceive ad-hoc and efficient algorithms, and for enumerating the entire solution set of a given instance.

The Mimetic team works on problems such as crowd simulation, spatial analysis, path and motion planning in static and dynamic environments, camera planning with visibility constraints for instance. The core of those problems, by nature, relies on problems and techniques belonging to computational geometry. Proposed models pay attention to algorithms complexity to be compatible with performance constraints imposed by interactive applications.



## MINT Project-Team

### 3. Research Program

#### 3.1. Human-Computer Interaction

The scientific approach that we follow considers user interfaces as means, not an end: our focus is not on interfaces, but on interaction considered as a phenomenon between a person and a computing system [26]. We *observe* this phenomenon in order to understand it, i.e. *describe* it and possibly *explain* it, and we look for ways to significantly *improve* it. HCI borrows its methods from various disciplines, including Computer Science, Psychology, Ethnography and Design. Participatory design methods can help determine users' problems and needs and generate new ideas, for example [30]. Rapid and iterative prototyping techniques allow to decide between alternative solutions [27]. Controlled studies based on experimental or quasi-experimental designs can then be used to evaluate the chosen solutions [32]. One of the main difficulties of HCI research is the doubly changing nature of the studied phenomenon: people can both adapt to the system and at the same time adapt it for their own specific purposes [29]. As these purposes are usually difficult to anticipate, we regularly *create* new versions of the systems we develop to take into account new theoretical and empirical knowledge. We also seek to *integrate* this knowledge in theoretical frameworks and software tools to disseminate it.

#### 3.2. Numerical and algorithmic real-time gesture analysis

Whatever is the interface, user provides some curves, defined over time, to the application. The curves constitute a gesture (positional information, yet may also include pressure). Depending on the hardware input, such a gesture may be either continuous (e.g. data-glove), or not (e.g. multi-touch screens). User gesture can be multi-variate (several fingers captured at the same time, combined into a single gesture, possibly involving two hands, maybe more in the context of co-located collaboration), that we would like, at higher-level, to be structured in time from simple elements in order to create specific command combinations. One of the scientific foundations of the research project is an algorithmic and numerical study of gesture, which we classify into three points:

- *clustering*, that takes into account intrinsic structure of gesture (multi-finger/multi-hand/multi-user aspects), as a lower-level treatment for further use of gesture by application;
- *recognition*, that identifies some semantic from gesture, that can be further used for application control (as command input). We consider in this topic multi-finger gestures, two-handed gestures, gesture for collaboration, on which very few has been done so far to our knowledge. On the contrary, in the case of single gesture case (i.e. one single point moving over time in a continuous manner), numerous studies have been proposed in the current literature, and interestingly, are of interest in several communities: HMM [33], Dynamic Time Warping [35] are well-known methods for computer-vision community, and hand-writing recognition. In the computer graphics community, statistical classification using geometric descriptors has previously been used [31]; in the Human-Computer interaction community, some simple (and easy to implement) methods have been proposed, that provide a very good compromise between technical complexity and practical efficiency [34].
- *mapping to application*, that studies how to link gesture inputs to application. This ranges from transfer function that is classically involved in pointing tasks [28], to the question to know how to link gesture analysis and recognition to the algorithmic of application content, with specific reference examples.

We ground our activity on the topic of numerical algorithm, expertise that has been previously achieved by team members in the physical simulation community (within which we think that aspects such as elastic deformation energies evaluation, simulation of rigid bodies composed of unstructured particles, constraint-based animation... will bring up interesting and novel insights within HCI community).

### 3.3. Design and control of haptic devices

Our scientific approach in the design and control of haptic devices is focused on the interaction forces between the user and the device. We search of controlling them, as precisely as possible. This leads to different designs compared to other systems which control the deformation instead. The research is carried out in three steps:

- *identification*: we measure the forces which occur during the exploration of a real object, for example a surface for tactile purposes. We then analyse the record to deduce the key components – *on user's point of view* – of the interaction forces.
- *design*: we propose new designs of haptic devices, based on our knowledge of the key components of the interaction forces. For example, coupling tactile and kinesthetic feedback is a promising design to achieve a good simulation of actual surfaces. Our goal is to find designs which lead to compact systems, and which can stand close to a computer in a desktop environment.
- *control*: we have to supply the device with the good electrical signals to accurately output the good forces.

## Mjolnir Team

### 3. Research Program

#### 3.1. Introduction

Our research program is organized around three main themes: leveraging human control skills, leveraging human perceptual skills, and leveraging human learning skills.

#### 3.2. Leveraging human control skills

Our group has developed a unique and recognized expertise in *transfer functions*, i.e. the algorithmic transformations of raw user input for system use. Transfer functions define how user actions are taken into account by the system. They can make a task easier or impossible and thus largely condition user performance, no matter the criteria (speed, accuracy, comfort, fatigue, etc). Ideally, the transfer function should be chosen or tuned to match the interaction context. Yet the question of how to design a function to maximize one or more criteria in a given context remains an open one, and on-demand adaptation is difficult because functions are usually implemented at the lowest possible level to avoid latency problems. Latency management and transfer function design are two problems that require cross examination to improve human performance with interactive systems. Both also contribute to the senses of *initiation* and *control*, two crucial component of the sense of *agency* [51]. Our ultimate goal on these topics is to adapt the transfer function to the user and task in order to support stable and appropriate control. To achieve this, we investigate combinations of low-level (embedded) and high-level (application) ways to take user capabilities and task characteristics into account and reduce or compensate for latency in different contexts, e.g. using a mouse or a touchpad, a touch-screen, an optical finger navigation device or a brain-computer interface.

#### 3.3. Leveraging human perceptual skills

Our work under this theme concerns the physicality of human-computer interaction, with a focus on haptic perception and related technologies, and the perception of animated displays.

Vibrators have long been used to provide basic kinesthetic feedback. Other piezoceramic and electro-active polymer technologies make it possible to support programmable friction or emboss a surface, and thin, organic technologies should soon provide transparent and conformable, flexible or stretchable substrates. We want to study the use of these different technologies for static and dynamic haptic feedback from both an engineering and an HCI perspective. We want to develop the tools and knowledge required to facilitate and inform the design of future haptic interactions taking best advantage of the different technologies.

Animations are increasingly common in graphical interfaces. Beyond their compelling nature, they are powerful tools that can be used to depict dynamic data, to help understand time-varying behaviors, to communicate a particular message or to capture attention. Yet despite their popularity, they are still largely under-comprehended as cognitive aids. While best practices provide useful directions, very little empirical research examine different types of animation, and their actual benefits and limitations remain to be determined. We want to increase current knowledge and develop the tools required to best take advantage of them.

#### 3.4. Leveraging human learning skills

By looking at ways to leverage human control and perceptual skills, the research yet proposed mainly aims at improving perception-action coupling to better support transparent use. This third research theme addresses the different and orthogonal topic of skill acquisition and improvement. We want to move away from the usual binary distinction between “novices” and “experts” and explore means to promote and assist digital skill development in a more progressive fashion. We are interested in means to support the analytic use of computing tools. We want to help people become aware of the particular ways they use their tools, the other

ways that exist for the things they do, and the other things they might do. We want to help them increase their performance by adjusting their current ways of doing, by providing new and more efficient ways, and by facilitating transitions from one way to another. We are also interested in means to foster reflection among users and facilitate the dissemination of best practices.

## POTIOC Project-Team

### 3. Research Program

#### 3.1. Introduction

The project of team potioc is oriented along three axes:

- Understanding humans interacting with the digital world
- Creating interactive systems
- Exploring new applications and usages

These axes are depicted in Figure 2 .

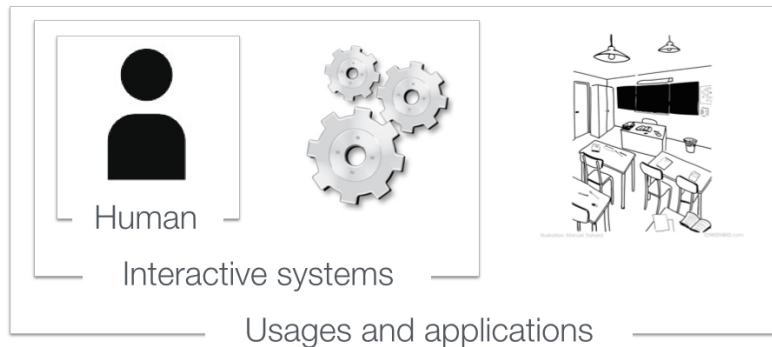


Figure 2. The three axes of the potioc team objectives.

Objective 1 is centered on the human sensori-motor and cognitive abilities, as well as user strategies and preferences, for completing interaction tasks. Our target contribution for this objective are a better understanding of humans interacting with interactive systems. The impact of this objective is mainly at a fundamental level.

In objective 2, our goal is to create interactive systems. This may include hardware parts where new input and output modalities are explored. This also includes software parts, that are strongly linked to the underlying hardware components. Our target contribution in objective 2 is to develop (hardware/software) interaction techniques allowing humans to perform interaction tasks.

Finally, in objective 3, we consider interaction at a higher level, taking into account factors that are linked to specific application domains and usages. Our target contribution in this area is the exploration and the emergence of new applications and usages that take benefit from the results of the project. With this objective, we target mainly a societal impact.

Of course, strong links exist between the three objectives of the project. For example, the results obtained in objective 1 guide the development of objective 2. Conversely, new systems developed in objective 2 may feed research questions of objective 1. There are similar links with objective 3.

## **3.2. Objective 1: Understanding humans interacting with the digital world**

Our first objective is centered on the human side. Our finality is not to enhance the general knowledge about the human being as a research team in psychology would do. Instead, we focus on human skills and behaviors during interaction processes. To this end, we conduct experiments that allow us to better understand what users like, where and why they have difficulties. Thanks to these investigations, we are able to design interaction techniques and systems (described in Objective 2) that are well suited to the targeted users. We believe that this fundamental piece of work is the first step that is required for the design of usable popular interactions. We are particularly interested in 3D interaction tasks for which we design dedicated experiments. We also explore a new approach based on physiological and brain (ElectroEncephaloGraphy - EEG) signals for the evaluation of these interactions.

### ***3.2.1. Interacting with physical and virtual environments***

Interacting with digital content displayed on 2D screens has been extensively studied in HCI. On the other hand, less conventional contexts have been studied less. This is the case of 3D environments, immersive virtual environments, augmented reality, and tangible objects. With the final goal of making interaction in such contexts user-friendly, we conduct experiments to better understand user strategies and performance. This allows us to propose guidelines to help designers creating of tools that are accessible to non-expert users.

### ***3.2.2. Evaluating (3D) interaction with physiological signals***

Recently, physiological computing has been shown to be a promising complement to Human-Computer Interfaces (HCI) in general, and to 3D User Interfaces (3DUI) in particular, in several directions. Within this research area, we are interested in using various physiological signals, and notably EEG signals, as a new tool to assess objectively the ergonomic quality of a given (3D) UI, to identify where and when are the pros and cons of this interface, based on the user's mental state during interaction. For instance, estimating the user's mental workload during interaction can give insights about where and when the interface is cognitively difficult to use. This could be useful for 2D HCI in general, and even more for 3DUI. Indeed, in a 3DUI, the user perception of the 3D scene – part of which could potentially be measured in EEG - is essential. Moreover, the usual need for a mapping between the user inputs and the corresponding actions on 3D objects make 3DUI and interaction techniques more difficult to assess and to design.

### ***3.2.3. Interacting with Brain-Computer Interfaces***

Although very promising for numerous applications, BCIs mostly remain prototypes not used outside laboratories, due to their low reliability. Poor BCI performances are partly due to imperfect EEG signal processing algorithms but also to the user who may not be able to produce reliable EEG patterns. Indeed, BCI use is a skill, requiring the user to be properly trained to achieve BCI control. If he/she cannot perform the desired mental commands, no signal processing algorithm can identify them. Therefore, rather than improving EEG signal processing alone, an interesting research direction is to also guide users to learn BCI control mastery. We aim at addressing this objective. We are notably exploring theoretical models and guidelines from educational sciences to improve BCI training protocols. We also study which users' profiles (personality and cognitive profile) fail or succeed at learning BCI control. Finally, we explore new feedback types and new EEG visualization techniques in order to help users gain BCI control skills more efficiently. These new feedback and visualizations notably aim at providing BCI users with more information about their EEG patterns, in order to identify more easily relevant BCI control strategies, as well as motivating and engaging them in the learning task.

### ***3.2.4. Interaction for people with special needs***

Interaction capabilities and needs largely depend on the target user group. In the Potioc project-team, we work with people having special needs. As an example, we work with children in the context of education, which requires us to design interfaces that are usable, engaging and support learning for this target group. Furthermore, we work with people with cognitive or perceptive disabilities, which requires us to consider accessibility, while at the same time designing interfaces that are learnable and enjoyable to use. In order to meet the needs of the different target groups, we apply participative and user-centred design methods.

### 3.3. Objective 2: Creating interactive systems

Our objective here is to create interactive systems and design interaction techniques dedicated to the completion of interaction tasks. We divide our work into three main categories:

- Interaction techniques based on existing Input/Output (IO) devices.
- New IO and related techniques.
- BCI and physiological computing.

#### 3.3.1. Interaction techniques based on existing Input/Output (IO) devices

When using desktop IOs (i.e., based on mice/keyboards/monitors), a big challenge is to design interaction techniques that allow users to complete 3D interaction tasks. Indeed, the desktop IO space that is mainly dedicated to the completion of 2D interaction task is not well suited to 3D content and, consequently, 3D user interfaces need to be designed with a great care. In the past few years, we have been particularly interested in the problem of interaction when the 3D content is displayed on a touchscreen. Indeed, standard (2D) HCI has evolved from mouse to touch input, and numerous research projects have been conducted. On the contrary, in 3D, very little work has been proposed. We are contributing to moving desktop 3D UIs from the mouse to the touch paradigm; what we used to do with mice in front of a screen does not work well on touch devices anymore. In the future, we will continue designing new interaction techniques that are based on standard IOs (eg. pointing devices and webcams) and that target the main objectives of Potioc which are to enhance the interaction bandwidth for non expert users.

#### 3.3.2. New IO and related techniques

Beyond standard IOs, we are interested in exploring new IO modalities that may make interaction easier, more engaging and motivating. In Potioc, we design new interactive systems that exploit unconventional IO modalities such as stereoscopy, 3D spatial input, augmented reality and so on. In particular, tangible interaction and spatial augmented reality are major subjects of interest for us. Indeed, we believe that manipulating directly physical objects for interacting with the digital world has a great potential, in particular when the general public is targeted. With such approaches, the computer disappears, and the user interacts with the digital content as he or she would do with physical content, which reduces the distance to the manipulated content. As an example, we recently designed Teegi, a new system based on a unique combination of spatial augmented reality, tangible interaction and real-time neurotechnologies. With Teegi, a user can visualize and analyze his or her own brain activity in real-time, on a tangible character that can be easily manipulated, and with which it is possible to interact. Such unconventional user interfaces that are based on rich sensing modalities hold great promises in the field of popular interaction.

We are also interested in designing systems that combine different sensory modalities, such as vision, touch and audition. Concrete examples include the design of tangible user interfaces or interfaces for visually impaired people. It has been shown that multimodality can provide rich interaction that can efficiently support learning, and it is also important in the context of accessibility.

#### 3.3.3. BCI and physiological computing

Although Brain-Computer Interfaces (BCI) have demonstrated their tremendous potential in numerous applications, they are still mostly prototypes, not used outside laboratories. This is mainly due to the following limitations:

- Performances: the poor classification accuracies of BCIs make them inconvenient to use or simply useless compared to available alternatives
- Stability and robustness: the sensibility of ElectroEncephaloGraphic (EEG) signals to noise and their inherent non-stationarity make the already poor initial performances difficult to maintain over time
- Calibration time: the need to tune current BCIs to each user's EEG signals makes their calibration times too long.

As part of our research on EEG-based BCIs, we notably aim at addressing these limitations by designing robust EEG signal processing tools with minimal calibration times, in order to design practical BCI systems, usable and useful outside laboratories. To do so we explore the design of alternative features and robust spatial filtering algorithms to make BCIs more robust to noise and non-stationarities, as well as more accurate. We also explore artificial EEG data generation and user-to-user data transfer to reduce calibration times.

### **3.4. Objective 3: Exploring new applications and usages**

Objective 3 is centered on the applications and usages. Beyond the human sensori-motor and cognitive skills (Objective 1), and the hardware and software components (Objective 2), Objectives 3 takes into account broader criteria for the emergence of new usages and applications in various areas, and in particular in the scope of education, art, popularization of science and entertainment. Our goal here is not to develop full-fledged end-user applications. Instead, our contribution is to stimulate the evolution of current applications with new engaging interactive systems.

#### **3.4.1. Education**

Education is at the core of the motivations of the Potioc group. Indeed, we are convinced that the approaches we investigate—which target motivation, curiosity, pleasure of use and high level of interactivity—may serve education purposes. To this end, we collaborate with experts in Educational Sciences and teachers for exploring new interactive systems that enhance learning processes. We are currently investigating the fields of astronomy, optics, and neurosciences. We are also working with special education centres for the blind on accessible augmented reality prototypes. In the future, we will continue exploring new interactive approaches dedicated to education, in various fields.

#### **3.4.2. Popularization of science**

Popularization of Science is also a key domain for Potioc. Focusing on this subject allows us to get inspiration for the development of new interactive approaches. In particular, we have built a strong partnership with Cap Sciences, which is a center dedicated to the popularization of science in Bordeaux that is visited by thousands of visitors every month. This was initiated with the ANR national project InSTInCT, whose goal was to study the benefits of 3D touch-based interaction in public exhibitions. This project has led to the creation of a Living Lab where several systems developed by Potioc have been tested and will be tested by the visitors. This provides us with very interesting observations that go beyond the feedback we can obtain in our controlled lab-experiments.

#### **3.4.3. Art**

Art, which is strongly linked with emotions and user experiences, is also a target area for Potioc. We believe that the work conducted in Potioc may be beneficial for creation from the artist point of view, and it may open new interactive experiences from the audience point of view. As an example, we are working with colleagues who are specialists in digital music, and with musicians. We are also working with jugglers and mockup builders with the goal of enhancing interactivity and user experience.

#### **3.4.4. Entertainment**

Similarly, entertainment is a domain where our work may have an impact. We notably explored BCI-based gaming and non-medical applications of BCI, as well as mobile Augmented Reality games. Once again, we believe that our approaches that merge the physical and the virtual world may enhance the user experience. Exploring such a domain will raise numerous scientific and technological questions.



## TITANE Project-Team

### 3. Research Program

#### 3.1. Context

Geometric modeling and processing revolve around three main end goals: a computerized shape representation that can be visualized (creating a realistic or artistic depiction), simulated (anticipating the real) or realized (manufacturing a conceptual or engineering design). Aside from the mere editing of geometry, central research themes in geometric modeling involve conversions between physical (real), discrete (digital), and mathematical (abstract) representations. Going from physical to digital is referred to as shape acquisition and reconstruction; going from mathematical to discrete is referred to as shape approximation and mesh generation; going from discrete to physical is referred to as shape rationalization.

Geometric modeling has become an indispensable component for computational and reverse engineering. Simulations are now routinely performed on complex shapes issued not only from computer-aided design but also from an increasing amount of available measurements. The scale of acquired data is quickly growing: we no longer deal exclusively with individual shapes, but with entire *scenes*, possibly at the scale of entire cities, with many objects defined as structured shapes. We are witnessing a rapid evolution of the acquisition paradigms with an increasing variety of sensors and the development of community data, as well as disseminated data.

In recent years, the evolution of acquisition technologies and methods has translated in an increasing overlap of algorithms and data in the computer vision, image processing, and computer graphics communities. Beyond the rapid increase of resolution through technological advances of sensors and methods for mosaicing images, the line between laser scan data and photos is getting thinner. Combining, e.g., laser scanners with panoramic cameras leads to massive 3D point sets with color attributes. In addition, it is now possible to generate dense point sets not just from laser scanners but also from photogrammetry techniques when using a well-designed acquisition protocol. Depth cameras are getting increasingly common, and beyond retrieving depth information we can enrich the main acquisition systems with additional hardware to measure geometric information about the sensor and improve data registration: e.g., accelerometers or GPS for geographic location, and compasses or gyrometers for orientation. Finally, complex scenes can be observed at different scales ranging from satellite to pedestrian through aerial levels.

These evolutions allow practitioners to measure urban scenes at resolutions that were until now possible only at the scale of individual shapes. The related scientific challenge is however more than just dealing with massive data sets coming from increase of resolution, as complex scenes are composed of multiple objects with structural relationships. The latter relate i) to the way the individual shapes are grouped to form objects, object classes or hierarchies, ii) to geometry when dealing with similarity, regularity, parallelism or symmetry, and iii) to domain-specific semantic considerations. Beyond reconstruction and approximation, consolidation and synthesis of complex scenes require rich structural relationships.

The problems arising from these evolutions suggest that the strengths of geometry and images may be combined in the form of new methodological solutions such as photo-consistent reconstruction. In addition, the process of measuring the geometry of sensors (through gyrometers and accelerometers) often requires both geometry process and image analysis for improved accuracy and robustness. Modeling urban scenes from measurements illustrates this growing synergy, and it has become a central concern for a variety of applications ranging from urban planning to simulation through rendering and special effects.

#### 3.2. Analysis

Complex scenes are usually composed of a large number of objects which may significantly differ in terms of complexity, diversity, and density. These objects must be identified and their structural relationships must be recovered in order to model the scenes with improved robustness, low complexity, variable levels of details and ultimately, semantization (automated process of increasing degree of semantic content).

*Object classification* is an ill-posed task in which the objects composing a scene are detected and recognized with respect to predefined classes, the objective going beyond scene segmentation. The high variability in each class may explain the success of the stochastic approach which is able to model widely variable classes. As it requires a priori knowledge this process is often domain-specific such as for urban scenes where we wish to distinguish between instances as ground, vegetation and buildings. Additional challenges arise when each class must be refined, such as roof super-structures for urban reconstruction.

*Structure extraction* consists in recovering structural relationships between objects or parts of object. The structure may be related to adjacencies between objects, hierarchical decomposition, singularities or canonical geometric relationships. It is crucial for effective geometric modeling through levels of details or hierarchical multiresolution modeling. Ideally we wish to learn the structural rules that govern the physical scene manufacturing. Understanding the main canonical geometric relationships between object parts involves detecting regular structures and equivalences under certain transformations such as parallelism, orthogonality and symmetry. Identifying structural and geometric repetitions or symmetries is relevant for dealing with missing data during data consolidation.

*Data consolidation* is a problem of growing interest for practitioners, with the increase of heterogeneous and defect-laden data. To be exploitable, such defect-laden data must be consolidated by improving the data sampling quality and by reinforcing the geometrical and structural relations sub-tending the observed scenes. Enforcing canonical geometric relationships such as local coplanarity or orthogonality is relevant for registration of heterogeneous or redundant data, as well as for improving the robustness of the reconstruction process.

### 3.3. Approximation

Our objective is to explore the approximation of complex shapes and scenes with surface and volume meshes, as well as on surface and domain tiling. A general way to state the shape approximation problem is to say that we search for the shape discretization (possibly with several levels of detail) that realizes the best complexity / distortion trade-off. Such a problem statement requires defining a discretization model, an error metric to measure distortion as well as a way to measure complexity. The latter is most commonly expressed in number of polygon primitives, but other measures closer to information theory lead to measurements such as number of bits or minimum description length.

For surface meshes we intend to conceive methods which provide control and guarantees both over the global approximation error and over the validity of the embedding. In addition, we seek for resilience to heterogeneous data, and robustness to noise and outliers. This would allow repairing and simplifying triangle soups with cracks, self-intersections and gaps. Another exploratory objective is to deal generically with different error metrics such as the symmetric Hausdorff distance, or a Sobolev norm which mixes errors in geometry and normals.

For surface and domain tiling the term meshing is substituted for tiling to stress the fact that tiles may be not just simple elements, but can model complex smooth shapes such as bilinear quadrangles. Quadrangle surface tiling is central for the so-called *resurfacing* problem in reverse engineering: the goal is to tile an input raw surface geometry such that the union of the tiles approximates the input well and such that each tile matches certain properties related to its shape or its size. In addition, we may require parameterization domains with a simple structure. Our goal is to devise surface tiling algorithms that are both reliable and resilient to defect-laden inputs, effective from the shape approximation point of view, and with flexible control upon the structure of the tiling.

### 3.4. Reconstruction

Assuming a geometric dataset made out of points or slices, the process of shape reconstruction amounts to recovering a surface or a solid that matches these samples. This problem is inherently ill-posed as infinitely-many shapes may fit the data. One must thus regularize the problem and add priors such as simplicity or smoothness of the inferred shape.

The concept of geometric simplicity has led to a number of interpolating techniques commonly based upon the Delaunay triangulation. The concept of smoothness has led to a number of approximating techniques that commonly compute an implicit function such that one of its isosurfaces approximates the inferred surface. Reconstruction algorithms can also use an explicit set of prior shapes for inference by assuming that the observed data can be described by these predefined prior shapes. One key lesson learned in the shape problem is that there is probably not a single solution which can solve all cases, each of them coming with its own distinctive features. In addition, some data sets such as point sets acquired on urban scenes are very domain-specific and require a dedicated line of research.

In recent years the *smooth, closed case* (i.e., shapes without sharp features nor boundaries) has received considerable attention. However, the state-of-the-art methods have several shortcomings: in addition to being in general not robust to outliers and not sufficiently robust to noise, they often require additional attributes as input, such as lines of sight or oriented normals. We wish to devise shape reconstruction methods which are both geometrically and topologically accurate without requiring additional attributes, while exhibiting resilience to defect-laden inputs. Resilience formally translates into stability with respect to noise and outliers. Correctness of the reconstruction translates into convergence in geometry and (stable parts of) topology of the reconstruction with respect to the inferred shape known through measurements.

Moving from the smooth, closed case to the *piecewise smooth case* (possibly with boundaries) is considerably harder as the ill-posedness of the problem applies to each sub-feature of the inferred shape. Further, very few approaches tackle the combined issue of robustness (to sampling defects, noise and outliers) and feature reconstruction.

## ALPAGE Project-Team

### 3. Research Program

#### 3.1. From programming languages to linguistic grammars

**Participants:** Éric Villemonte de La Clergerie, Benoît Sagot, Pierre Boullier, Djamé Seddah, Corentin Ribeyre.

Historically, several members of Alpage were originally specialists in the domain of modeling and parsing for programming languages, and have been working for more than 15 years on the generalization and extension of the techniques involved to the domain of natural language. The shift from programming language grammars to NLP grammars seriously increases complexity (e.g., grammar size <sup>0</sup>) and requires ways to handle the ambiguities inherent in every human language. It is well known that these ambiguities are the sources of many badly handled combinatorial explosions.

Furthermore, while most programming languages are expressed by (subclasses) of well-understood context-free grammars (CFGs), no consensual grammatical formalism has yet been accepted by the whole linguistic community for the description of human languages. On the contrary, new formalisms (or variants of older ones) appear constantly. Many of them may be classified into the three following large families:

**Mildly Context-Sensitive (MCS) formalisms** They manipulate possibly complex elementary structures with enough restrictions to ensure the possibility of parsing with polynomial time complexities. They include, for instance, Tree Adjoining Grammars (TAGs) and Multi-component TAGs with trees as elementary structures, Linear Indexed Grammars (LIGs). Although they are strictly more powerful than MCS formalisms, Range Concatenation Grammars (RCGs, introduced and used by Alpage members, such as Pierre Boullier and Benoît Sagot [56], [79], [84]) are also parsable in polynomial time.

**Unification-based formalisms** They combine a context-free backbone with logic arguments as decoration on non-terminals. Most famous representatives are Definite Clause Grammars (DCGs) where PROLOG powerful unification is used to compute and propagate these logic arguments. More recent formalisms, like Lexical Functional Grammars (LFGs) and Head-Driven Phrasal Structure Grammars (HPSGs) rely on more expressive Typed Feature Structures (TFS) or constraints.

**Unification-based formalisms with an MCS backbone** The two above-mentioned characteristics may be combined, for instance by adding logic arguments or constraints to non-terminals in TAGs.

An efficient way to develop large-coverage hand-crafted symbolic grammars is to use adequate tools and adequate levels of representation, and in particular Meta-Grammars, one of Alpage's areas of expertise, especially with the FRMG grammar and parser for French based on the DyALog logic programming environment [92], [91]. Meta-Grammars (MGs) allows the linguist to focus on a modular description of the linguistic aspects of a grammar, rather than focusing on the specific aspects of a given grammatical formalism. Translation from MGs to grammatical formalisms such as TAG or LFG may be automatically handled. Graphical environments can be used to design MGs and their modularity provides a promising way for sharing the description of common linguistic phenomena across human languages.

#### 3.2. Statistical Parsing

**Participants:** Djamé Seddah, Marie-Hélène Candito, Benoît Crabbé, Éric Villemonte de La Clergerie, Benoît Sagot, Corentin Ribeyre, Pierre Boullier, Maximin Coavoux.

---

<sup>0</sup>boullier:2010:inria-00516341:1

Contrary to symbolic approaches to parsing, in statistical parsing, the grammar is extracted from a corpus of syntactic trees : a treebank. The main advantage of the statistical approach is to encode within the same framework the parsing and disambiguating tasks. The extracted grammar rules are associated with probabilities that allow to score and rank the output parse trees of an input sentence. This obvious advantage of probabilistic context-free grammars has long been counterbalanced by two main shortcomings that resulted in poor performance for plain PCFG parsers: (i) the generalization encoded in non terminal symbols that stand for syntagmatic phrases is too coarse (so probabilistic independence between rules is too strong an assertion) and (ii) lexical items are underused. In the last decade though, effective solutions to these shortcomings have been proposed. Symbol annotation, either manual [72] or automatic [75], [76] captures inter-dependence between CFG rules. Lexical information is integrated in frameworks such as head-driven models that allow lexical heads to percolate up the syntagmatic tree [59], or probabilistic models derived from lexicalized Tree Adjoining grammars, such as Stochastic Tree Insertion Grammars [58].

In the same period, totally different parsing architectures have been proposed, to obtain dependency-based syntactic representations. The properties of dependency structures, in which each word is related to exactly one other word, make it possible to define dependency parsing as a sequence of simple actions (such as read buffer and store word on top of a stack, attach read word as dependent of stack top word, attach read word as governor of stack top word ...) [94], [74]. Classifiers can be trained to choose the best action to perform given a partial parsing configuration. In another approach, dependency parsing is cast into the problem of finding the maximum spanning tree within the graph of all possible word-to-word dependencies, and online classification is used to weight the edges [73]. These two kinds of statistical dependency parsing allow to benefit from discriminative learning, and its ability to easily integrate various kinds of features, which is typically needed in a complex task such as parsing.

Statistical parsing is now effective, both for syntagmatic representations and dependency-based syntactic representations. Alpage has obtained state-of-the-art parsing results for French, by adapting various parser learners for French, and works on the current challenges in statistical parsing, namely (1) robustness and portability across domains and (2) the ability to incorporate exogenous data to improve parsing attachment decisions. Alpage is the first French team to have turned the French TreeBank into a resource usable for training statistical parsers, to distribute a dependency version of this treebank, and to make freely available various state-of-the-art statistical POS-taggers and parsers for French. We review below the approaches that Alpage has tested and adapted, and the techniques that we plan to investigate to answer these challenges.

In order to investigate statistical parsers for French, we have first worked how to use the French Treebank [53], [52] and derive the best input for syntagmatic statistical parsing [60]. Benchmarking several PCFG-based learning frameworks [86] has led to state-of-the-art results for French, the best performance being obtained with the split-merge Berkeley parser (PCFG with latent annotations) [76].

In parallel to the work on dependency based representation, presented in the next paragraph, we also conducted a preliminary set of experiments on richer parsing models based on Stochastic Tree Insertion Grammars as used in [58] and which, besides their inferior performance compared to PCFG-LA based parser, raise promising results with respect to dependencies that can be extracted from derivation trees. One variation we explored, that uses a specific TIG grammar instance, a *vertical* grammar called *spinal* grammars, exhibits interesting properties wrt the grammar size typically extracted from treebanks (a few hundred unlexicalized trees, compared to 14 000 CFG rules). These models are currently being investigated in our team [89].

Pursuing our work on PCFG-LA based parsing, we investigated the automatic conversion of the treebank into dependency syntax representations [57], that are easier to use for various NLP applications such as question-answering or information extraction, and that are a better ground for further semantic analysis. This conversion can be applied on the treebank, before training a dependency-based parser, or on PCFG-LA parsed trees. This gives the possibility to evaluate and compare on the same gold data, both syntagmatic- and dependency-based statistical parsing. This also paved the way for studies on the influence of various types of lexical information.

### 3.3. Robust linguistic processing

**Participants:** Djamé Seddah, Benoît Sagot, Éric Villemonte de La Clergerie, Marie-Hélène Candito, Pierre Magistry.

The constitution of resources such as lexica or grammars raises the issues of the evaluation of these resources to assess their quality and coverage. For this reason, Alpage was the leader of the PASSAGE ANR project (ended in June 2010), which is the follow-up of the EASy parsing evaluation campaign held in 2004 and conducted by team LIR at LIMSI.

However, although developing parsing techniques, grammars (symbolic or probabilistic), and lexica constitute the key efforts towards deep large-scale linguistic processing, these components need to be included inside a full and robust processing chain, able to handle any text from any source, especially out-of-domain text genres. Such texts that exhibit properties (e.g., lexical and syntactic properties) that are different or differently distributed than what is found on standard data (e.g., training corpora for statistical parsers). The development of shallow processing chains, such as SXPipe, is not a trivial task [80]. Obviously, they are often used as such, and not only as pre-processing tools before parsing, since they perform the basic tasks that produce immediately usable results for many applications, such as tokenization, sentence segmentation, spelling correction (e.g., for improving the output of OCR systems), named entity detection, disambiguation and resolution, as well as morphosyntactic tagging.

Still, when used as a preliminary step before parsers, the quality of parsers' results strongly depends on the quality of such chains. This is especially the case, beyond the standard out-of-domain corpora mentioned above, for user-generated content. Indeed, until very recently out-of-domain text genres that have been prioritized have not been Web 2.0 sources, but rather biomedical texts, child language and general fiction (Brown corpus). Adaptation to user-generated content is a particularly difficult instance of the domain adaptation problem since Web 2.0 is not really a domain: it consists of utterances that are often ungrammatical. It even shares some similarities with spoken language [90]. The poor overall quality of texts found on such media lead to weak parsing and even POS-tagging results. This is because user-generated content exhibits both the same issues as other out-of-domain data, but also tremendous issues related to tokenization, typographic and spelling issues that go far beyond what statistical tools can learn from standard corpora. Even lexical specificities are often more challenging than on edited out-of-domain text, as neologisms built using productive morphological derivation, for example, are less frequent, contrarily to slang, abbreviations or technical jargon that are harder to analyse and interpret automatically.

In order to fully prepare a shift toward more robustness, we developed a first version of a richly annotated corpus of user-generated French text, the French Social Media Bank [7], which includes not only POS, constituency and functional information, but also a layer of "normalized" text. This corpus is fully available and constitutes the first data set on Facebook data to date and the first instance of user generated content for a morphologically-rich language. Thanks to the support of the Labex EFL through, we are currently finalizing the second release of this data set, extending toward a full treebank of over 4,000 sentences.

Besides delivering a new data set, our main purpose here is to be able to compare two different approaches to user-generated content processing: either training statistical models on the original annotated text, and use them on raw new text; or developing normalization tools that help improving the consistency of the annotations, train statistical models on the normalized annotated text, and use them on normalized texts (before un-normalizing them).

However, this raises issues concerning the normalization step. A good sandbox for working on this challenging task is that of POS-tagging. For this purpose, we did leverage Alpage's work on MEIt, a state-of-the-art POS tagging system [68]. A first round of experiments on English have already led to promising results during the shared task on parsing user-generated content organized by Google in May 2012 [77], as Alpage was ranked second and third [88]. For achieving this result, we brought together a preliminary implementation of a normalization wrapper around the MEIt POS tagger followed by a state-of-the-art statistical parser improved by several domain adaptation techniques we originally developed for parsing edited out-of-domain texts. Those techniques are based on the unsupervised learning of word clusters *a la* Brown and benefit from morphological treatments (such as lemmatization or desinflexion) [87].

One of our objectives is to generalize the use of the normalization wrapper approach to both POS tagging and parsing, for English and French, in order to improve the quality of the output parses. However, this raises several challenges: non-standard contractions and compounds lead to unexpected syntactic structures. A first round of experiments on the French Social Media Bank showed that parsing performance on such data are much lower than expected. This is why, we are actively working to improve on the baselines we established on that matter.

### 3.4. Dynamic wide coverage lexical resources

**Participants:** Benoît Sagot, Laurence Danlos, Éric Villemonte de La Clergerie, Marie-Hélène Candito, Lucie Barque, Marianne Djemaa.

Grammatical formalisms and associated parsing generators are useful only when used together with linguistic resources (lexicons, grammars) so as to build operational parsers, especially when considering modern lexically oriented grammatical formalisms. Hence, linguistic resources are the topic of the following section.

However, wide coverage linguistic resources are scarce and expensive, because they are difficult to build, especially when hand-crafted. This observation motivates us to investigate methods, along to manual development techniques, to automatically or semi-automatically acquire, supplement and correct linguistic resources.

Linguistic expertise remains a very important asset to benefit efficiently from such techniques, including those described below. Moreover, linguistically oriented environments with adequate collaborative interfaces are needed to facilitate the edition, comparison, validation and maintenance of large scale linguistic resources. Just to give some idea of the complexity, a syntactic lexicon, as described below, should provide rich information for several tens of thousands of lemma and several hundreds of thousands of forms.

Successful experiments have been conducted by Alpage members with different languages for the automatic acquisition of morphological knowledge from raw corpora [83]. At the syntactic level, work has been achieved on automatic acquisition of atomic syntactic information and automatic detection of errors in the lexicon [95],[6]. At the semantic level, automatic wordnet development tools have been described [78], [93], [71], [69]. All such techniques need of course to be followed by manual validation, so as to ensure high-quality results.

For French, these techniques, and others, have lead some Alpage members to develop one of the main syntactic resources for French, the *Lefff* [81], [85], developed within the Alexina framework. At the semantic level, Alpage members have developed or are developing various syntactico-semantic or semantic resources, including:

- a wordnet for French, the WOLF [82], [70], the first freely available resource of the kind;
- a French FrameNet lexicon (together with an annotated corpus) within the ASFALDA ANR project;
- and a French VerbNet.

In the last few years, Alpage members have shown how to benefit from other more linguistically-oriented resources, such as the *Lexique-Grammaire* and *DICOVALENCE*, in order to improve the coverage and quality of the *Lefff*, the WOLF, the French FrameNet lexicon and the French VerbNet. This work is a good example of how Inria and Paris 7 members of Alpage fruitful collaborate: this collaboration between NLP computer scientists and NLP linguists have resulted in significant advances which would have not been possible otherwise.

Moreover, an increasing effort has been made towards multilingual aspects. In particular, Alexina lexicons exist for German, Slovak, Polish, English, Spanish, Persian, Latin (verbs only), Kurmanji Kurdish, Maltese (verbs only, restricted to the so-called first *binyan*) and Khaling, not including freely-available lexicons adapted to the Alexina framework.

### 3.5. Discourse structures

**Participants:** Laurence Danlos, Timothée Bernard, Raphaël Salmon.

Until now, the linguistic modeling and automatic processing of sentences has been the main focus of the community. However, many applications would benefit from more large-scale approaches which go beyond the level of sentences. This is not only the case for automatic translation: information extraction/retrieval, summarizing, and other applications do need to resolve anaphora, which in turn can benefit from the availability of hierarchical discourse structures induced by discourse relations (in particular through the notion of right frontier of discourse structures). Moreover, discourse structures are required to extract sequential (chronological, logical,...) or hierarchical representations of events. It is also useful for topic extraction, which in turns can help syntactic and semantic disambiguation.

Although supra-sentential problematics received increasing attention in the last years, there is no satisfying solution to these problems. Among them, anaphora resolution and discourse structures have a far-reaching impact and are domains of expertise of Alpage members. But their formal modeling has now reached a maturity which allows to integrate them, in a near future, inside future Alpage tools, including parsing systems inherited from Atoll.

It is well known that a text is not a random sequence of sentences: sentences are linked the ones to the others by “discourse relations”, which give to the text a hierarchical structure. Traditionally, it is considered that discourse relations are lexicalized by connectors (adverbial connectors like *ensuite*, conjunctions like *parce que*), or are not lexicalized. This vision is however too simple:

- first, some connectors (in particular conjunctions of subordination) introduce pure modifiers and must not be considered as bearing discourse relations,
- second, other elements than connectors can lexicalize discourse relations, in particular verbs like *précéder / to precede* or *causer / to cause*, which have facts or fact eventualities as arguments [62].

There are three main frameworks used to model discourse structures: RST, SDRT, and, more recently, the TAG-based formalism D-LTAG. Inside Alpage, Laurence Danlos has introduced D-STAG (Discourse Synchronous TAGs, [63],[4]), which subsumes in an elegant way both SDRT and RST, to the extent that SDRT and RST structures can be obtained by two different partial projections of D-STAG structures. As done in D-LTAG, D-STAG extends a lexicalized TAG analysis so as to deal with the level of discourse. D-STAG has been fully formalized, and is hence possible to implement (thanks to Synchronous TAG, or even TAG parsers), provided one develops linguistic descriptions in this formalism.



## MULTISPEECH Project-Team

### 3. Research Program

#### 3.1. Explicit Modeling of Speech Production and Perception

Speech signals are the consequence of the deformation of the vocal tract under the effect of the movements of the articulators (jaw, lips, tongue, ...) to modulate the excitation signal produced by the vocal cords or air turbulence. These deformations are visible on the face (lips, cheeks, jaw) through the coordination of different orofacial muscles and skin deformation induced by the latter. These deformations may also express different emotions. We should note that human speech expresses more than just phonetic content, to be able to communicate effectively. In this project, we address the different aspects related to speech production from the modeling of the vocal tract up to the production of expressive audiovisual speech. Phonetic contrasts used by the phonological system of any language result from constraints imposed by the nature of the human speech production apparatus. For a given language these contrasts are organized so as to guarantee that human listeners can identify (categorize) sounds robustly. The study of the categorization of sounds and prosody thus provides a complementary view on speech signals by focusing on the discrimination of sounds by humans, particularly in the context of language learning.

##### 3.1.1. *Articulatory modeling*

Modeling speech production is a major issue in speech sciences. Acoustic simulation makes the link between articulatory and acoustic domains. Unfortunately this link cannot be fully exploited because there is almost always an acoustic mismatch between natural and synthetic speech generated with an articulatory model approximating the vocal tract. However, the respective effects of the geometric approximation, of the fact of neglecting some cavities in the simulation, of the imprecision of some physical constants and of the dimensionality of the acoustic simulation are still unknown. Hence, the first objective is to investigate the origin of the acoustic mismatch by designing more precise articulatory models, developing new methods to acquire tridimensional Magnetic Resonance Imaging (MRI) data of the entire vocal tract together with denoised speech signals, and evaluating several approaches of acoustic simulation. The articulatory data acquisition relies on a head-neck antenna at Nancy Hospital to acquire MRI of the vocal tract, and on the articulograph Carstens AG501 available in the laboratory.

Up to now, acoustic-to-articulatory inversion has been addressed as an instantaneous problem, articulatory gestures being recovered by concatenating local solutions. The second objective is thus to investigate how more elaborated strategies (a syllabus of primitive gestures, articulatory targets. . . ) can be incorporated in the acoustic-to-articulatory inversion algorithms to take into account dynamic aspects.

##### 3.1.2. *Expressive acoustic-visual synthesis*

Speech is considered as a bimodal communication means; the first modality is audio, provided by acoustic speech signals and the second one is visual, provided by the face of the speaker. In our approach, the Acoustic-Visual Text-To-Speech synthesis (AV-TTS) is performed simultaneously with respect to its acoustic and visible components, by considering a bimodal signal comprising both acoustic and visual channels. A first AV-TTS system has been developed resulting in a talking head; the system relied on 3D-visual data and on an extension of our acoustic-unit concatenation text-to-speech synthesis system (SoJA). An important goal is to provide an audiovisual synthesis that is intelligible, both acoustically and visually. Thus, we continue working on adding visible components of the head through a tongue model and a lip model. We will also improve the TTS engine to increase the accuracy of the unit selection simultaneously into the acoustic and visual domains. To acquire the facial data, we consider using a marker-less motion capture system using a kinect-like system with a face tracking software, which constitutes a relatively low-cost alternative to the Vicon system.

Another challenging research goal is to add expressivity in the AV-TTS. The expressivity comes through the acoustic signal (prosody aspects) and also through head and eyebrow movements. One objective is to add a prosodic component in the TTS engine in order to take into account some prosodic entities such as emphasis (to highlight some important key words). One intended approach will be to explore an expressivity measure at sound, syllable and/or sentence levels that describes the degree of perception or realization of an expression/emotion (audio and 3D domain). Such measures will be used as criteria in the selection process of the synthesis system. To tackle the expressivity issue we will also investigate Hidden Markov Model (HMM) based synthesis which allows for easy adaptation of the system to available data and to various conditions.

### **3.1.3. Categorization of sounds and prosody for native and non-native speech**

Discriminating speech sounds and prosodic patterns is the keystone of language learning whether in the mother tongue or in a second language. This issue is associated with the emergence of phonetic categories, i.e., classes of sounds related to phonemes and prosodic patterns. The study of categorization is concerned not only with acoustic modeling but also with speech perception and phonology. Foreign language learning raises the issue of categorizing phonemes of the second language given the phonetic categories of the mother tongue. Thus, studies on the emergence of new categories, whether in the mother tongue (for people with language deficiencies) or in a second language, must rely upon studies on native and non-native acoustic realizations of speech sounds and prosody, and on perceptual experiments. Concerning prosody, studies are focused on native and non-native realizations of modalities (e.g., question, affirmation, command, ...), as well as non-native realizations of lexical accents and focus (emphasis).

For language learning, the analysis of the prosody and of the acoustic realization of the sounds aims at providing automatic feedback to language learners with respect to acquisition of prosody as well as acquisition of a correct pronunciation of the sounds of the foreign language. Concerning the mother tongue we are interested in the monitoring of the process of sound categorization in the long term (mainly at primary school) and its relation with the learning of reading and writing skills [7], especially for children with language deficiencies.

## **3.2. Statistical Modeling of Speech**

Whereas the first research direction deals with the physical aspects of speech and its explicit modeling, this second research direction investigates statistical models for speech data. Acoustic models are used to represent the pronunciation of the sounds or other acoustic events such as noise. Whether they are used for source separation, for speech recognition, for speech transcription, or for speech synthesis, the achieved performance strongly depends on the accuracy of these models. At the linguistic level, MULTISPEECH investigates models for handling the context (beyond the few preceding words currently handled by the  $n$ -gram models) and evolutive lexicons necessary when dealing with diachronic audio documents. Statistical approaches are also useful for generating speech signals. Along this direction, MULTISPEECH considers voice transformation techniques, with their application to pathological voices, and statistical speech synthesis applied to expressive multimodal speech synthesis.

### **3.2.1. Source separation**

Acoustic modeling is a key issue for automatic speech recognition. Despite the progress made for many years, current speech recognition applications rely on strong constraints (close-talk microphone, limited vocabulary, or restricted syntax) to achieve acceptable performance. The quality of the input speech signals is particularly important and performance degrades quickly with noisy signals. Accurate signal enhancement techniques are therefore essential to increase the robustness of both automatic speech recognition and speech-text alignment systems to noise and non-speech events.

In MULTISPEECH, focus is set on source separation techniques using multiple microphones and/or models of non-speech events. Some of the challenges include getting the most of the new modeling frameworks based on alpha-stable distributions and deep neural networks, combining them with established spatial filtering approaches, modeling more complex properties of speech and audio sources (phase, inter-frame and inter-frequency properties), and exploiting large data sets of speech, noise, and acoustic impulse responses to

automatically discover new models. Beyond the definition of such models, the difficulty will be to design scalable estimation algorithms robust to overfitting, integrate them into the recently developed FASST [6] and KAM software frameworks if relevant, and develop new software frameworks otherwise.

### **3.2.2. Linguistic modeling**

MULTISPEECH investigates lexical and language models in speech recognition with a focus on improving the processing of proper names and of spontaneous speech. Proper names are relevant keys in information indexing, but are a real problem in transcribing many diachronic spoken documents which refer to data, especially proper names, that evolve over time. This leads to the challenge of dynamically adjusting lexicons and language models through the use of the context of the documents or of some relevant external information. We also investigate language models defined on a continuous space (through neural network based approaches) in order to achieve a better generalization on unseen data, and to model long-term dependencies. We also want to introduce into these models additional relevant information such as linguistic features, semantic relation, topic or user-dependent information.

Other topics are spontaneous speech and pronunciation lexicons. Spontaneous speech utterances are often ill-formed and frequently contain disfluencies (hesitations, repetitions, ...) that degrade speech recognition performance. Hence the objective of improving the modeling of disfluencies and of spontaneous speech pronunciation variants. Attention will also be set on pronunciation lexicons with respect to non-native speech and foreign names. Non-native pronunciation variants have to take into account frequent mis-pronunciations due to differences between mother tongue and target language phoneme inventories. Proper name pronunciation variants are a similar problem where difficulties are mainly observed for names of foreign origin that can be pronounced either in a French way or kept close to foreign origin native pronunciation.

### **3.2.3. Speech generation by statistical methods**

Over the last few years statistical speech synthesis has emerged as an alternative to corpus-based speech synthesis. The announced advantages of the statistical speech synthesis are the possibility to deal with small amounts of speech resources and the flexibility for adapting models (for new emotions or new speakers), however, the quality is not as good as that of the concatenation-based speech synthesis. MULTISPEECH will focus on a hybrid approach, combining corpus-based synthesis, for its high-quality speech signal output, and HMM-based speech synthesis for its flexibility to drive selection, and the main challenge will be on its application to producing expressive audio-visual speech.

Moreover, in the context of acoustic feedback in foreign language learning, voice modification approaches are investigated to modify the learner's (or teacher's) voice in order to emphasize the difference between the learner's acoustic realization and the expected realization.

## **3.3. Uncertainty Estimation and Exploitation in Speech Processing**

This axis focuses on the uncertainty associated with some processing steps. Uncertainty stems from the high variability of speech signals and from imperfect models. For example, enhanced speech signals resulting from source separation are not exactly the clean original speech signals. Words or phonemes resulting from automatic speech recognition contain errors, and the phone boundaries resulting from an automatic speech-text alignment are not always correct, especially in acoustically degraded conditions. Hence it is important to know the reliability of the results and/or to estimate the uncertainty of the results.

### **3.3.1. Uncertainty and acoustic modeling**

Because small distortions in the separated source signals can translate into large distortions in the cepstral features used for speech recognition, this limits the recognition performance on noisy data. One way to address this issue is to estimate the uncertainty of the separated sources in the form of their posterior distribution and to propagate this distribution, instead of a point estimate, through the subsequent feature extraction and speech decoding stages. Although major improvements have been demonstrated in proof-of-concept experiments using knowledge of the true uncertainty, accurate uncertainty estimation and propagation remains an open issue.

MULTISPEECH seeks to provide more accurate estimates of the posterior distribution of the separated source signals accounting for, e.g., posterior correlations over time and frequency which have not been considered so far. The framework of variational Bayesian (VB) inference appears to be a promising direction. Mappings learned on training data and fusion of multiple uncertainty estimators are also explored. The estimated uncertainties are then exploited for acoustic modeling in speech recognition and, in the future, also for speech-text alignment. This approach may later be extended to the estimation of the resulting uncertainty of the acoustic model parameters and of the acoustic scores themselves.

### ***3.3.2. Uncertainty and phonetic segmentation***

The accuracy of the phonetic segmentation is important in several cases, as for example for the computation of prosodic features, for avoiding incorrect feedback to the learner in computer assisted foreign language learning, or for the post-synchronization of speech with face/lip images. Currently the phonetic boundaries obtained are quite correct on good quality speech, but the precision degrades significantly on noisy and non-native speech. Phonetic segmentation aspects will be investigated, both in speech recognition (i.e., spoken text unknown) and in forced alignment (i.e., when the spoken text is known).

In the same way that combining several speech recognition outputs leads to improved speech recognition performance, MULTISPEECH will investigate the combination of several speech-text alignments as a way of improving the quality of speech-text alignment and of determining which phonetic boundaries are reliable and which ones are not, and also for estimating the uncertainty of the boundaries. Knowing the reliability of the boundaries will also be useful when segmenting speech corpora; this will help deciding which parts of the corpora need to be manually checked and corrected without an exhaustive checking of the whole corpus.

### ***3.3.3. Uncertainty and prosody***

Prosody information is also investigated as a means for structuring speech data (determining sentence boundaries, punctuation. . .) possibly in addition to syntactic dependencies. Structuring automatic transcription output is important for further exploitation of the transcription results such as easier reading after the addition of punctuation, or exploitation of full sentences in automatic translation. Prosody information is also necessary for determining the modality of the utterance (question or not), as well as determining accented words.

Prosody information comes from the fundamental frequency, the duration of the sounds and their energy. Any error in estimating these parameters may lead to a wrong decision. MULTISPEECH will investigate estimating the uncertainty of the duration of the phones (see uncertainty of phonetic boundaries above) and on the fundamental frequency, as well as how this uncertainty shall be propagated in the detection of prosodic phenomena such as accented words, utterance modality, or determination of the structure of the utterance.

## PANAMA Project-Team

### 3. Research Program

#### 3.1. Axis 1: Sparse Models and Representations

##### 3.1.1. *Efficient Sparse Models and Dictionary Design for Large-scale Data*

Sparse models are at the core of many research domains where the large amount and high-dimensionality of digital data requires concise data descriptions for efficient information processing. Recent breakthroughs have demonstrated the ability of these models to provide concise descriptions of complex data collections, together with algorithms of provable performance and bounded complexity.

A crucial prerequisite for the success of today's methods is the knowledge of a "dictionary" characterizing how to concisely describe the data of interest. Choosing a dictionary is currently something of an "art", relying on expert knowledge and heuristics.

Pre-chosen dictionaries such as wavelets, curvelets or Gabor dictionaries, are based upon stylized signal models and benefit from fast transform algorithms, but they fail to fully describe the content of natural signals and their variability. They do not address the huge diversity underlying modern data much beyond time series and images: data defined on graphs (social networks, internet routing, brain connectivity), vector valued data (diffusion tensor imaging of the brain), multichannel or multi-stream data (audiovisual streams, surveillance networks, multimodal biomedical monitoring).

The alternative to a pre-chosen dictionary is a trained dictionary learned from signal instances. While such representations exhibit good performance on small-scale problems, they are currently limited to low-dimensional signal processing due to the necessary training data, memory requirements and computational complexity. Whether designed or learned from a training corpus, dictionary-based sparse models and the associated methodology fail to scale up to the volume and resolution of modern digital data, for they intrinsically involve difficult linear inverse problems. To overcome this bottleneck, a new generation of efficient sparse models is needed, beyond dictionaries, encompassing the ability to provide sparse and structured data representations as well as computational efficiency. For example, while dictionaries describe low-dimensional signal models in terms of their "synthesis" using few elementary building blocks called atoms, in "analysis" alternatives the low-dimensional structure of the signal is rather "carved out" by a set of equations satisfied by the signal. Linear as well as nonlinear models can be envisioned.

##### 3.1.2. *Compressive Learning*

A flagship emerging application of sparsity is the paradigm of compressive sensing, which exploits sparse models at the analog and digital levels for the acquisition, compression and transmission of data using limited resources (fewer/less expensive sensors, limited energy consumption and transmission bandwidth, etc.). Besides sparsity, a key pillar of compressive sensing is the use of random low-dimensional projections. Through compressive sensing, random projections have shown their potential to allow drastic dimension reduction with controlled information loss, provided that the projected signal vector admits a sparse representation in some transformed domain. A related scientific domain, where sparsity has been recognized as a key enabling factor, is Machine Learning, where the overall goal is to design statistically founded principles and efficient algorithms in order to infer general properties of large data collections through the observation of a limited number of representative examples. Marrying sparsity and random low-dimensional projections with machine learning shall allow the development of techniques able to efficiently capture and process the information content of large data collections. The expected outcome is a dramatic increase of the impact of sparse models in machine learning, as well as an integrated framework from the signal level (signals and their acquisition) to the semantic level (information and its manipulation), and applications to data sizes and volumes of collections that cannot be handled by current technologies.

## 3.2. Axis 2: Robust Acoustic Scene Analysis

### 3.2.1. Compressive Acquisition and Processing of Acoustic Scenes

Acoustic imaging and scene analysis involve acquiring the information content from acoustic fields with a limited number of acoustic sensors. A full 3D+t field at CD quality and Nyquist spatial sampling represents roughly  $10^6$  microphones/ $m^3$ . Dealing with such high-dimensional data requires to drastically reduce the data flow by positioning appropriate sensors, and selecting from all spatial locations the few spots where acoustic sources are active. The main goal is to develop a theoretical and practical understanding of the conditions under which compressive acoustic sensing is both feasible and robust to inaccurate modeling, noisy measures, and partially failing or uncalibrated sensing devices, in various acoustic sensing scenarios. This requires the development of adequate algorithmic tools, numerical simulations, and experimental data in simple settings where hardware prototypes can be implemented.

### 3.2.2. Robust Audio Source Separation

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the combination of (i) complex source characteristics, (ii) sophisticated underlying mixing model and (iii) adverse recording environments. Moreover, as opposed to the “academic” blind source separation task, most applicative contexts and new interaction paradigms offer a variety of situations in which prior knowledge and adequate interfaces enable the design and the use of informed and/or manually assisted source separation methods.

The former METISS team has developed a generic and flexible probabilistic audio source separation framework that has the ability to combine various acoustic models such as spatial and spectral source models. Building on this existing framework, a first objective of PANAMA is to instantiate and validate specific instances of this framework targeted to real-world industrial applications, such as 5.1 movie re-mastering, interactive music soloist control and outdoor speech enhancement. Extensions of the framework are needed to achieve real-time online processing, and advanced constraints or probabilistic priors for the sources at hand need to be designed, while paying attention to computational scalability issues.

In parallel to these efforts, expected progress in sparse modeling for inverse problems shall bring new approaches to source separation and modeling, as well as to source localization, which is often an important first step in a source separation workflow.

### 3.2.3. Robust Audio Source Localization

Audio source localization consists in estimating the position of one or several sound sources given the signals received by a microphone array. Knowing the geometry of an audio scene is often a pre-requisite to perform higher-level tasks such as speaker identification and tracking, speech enhancement and recognition or audio source separation. It can be decomposed into two sub-tasks : (i) compute spatial auditory features from raw audio input and (ii) map these features to the desired spatial information. Robustly addressing both these aspects with a limited number of microphones, in the presence of noise, reverberation, multiple and possibly moving sources remains a key challenge in audio signal processing. The first aspect will be tackled by both advanced statistical and acoustical modeling of spatial auditory features. The second one will be addressed by two complementary approaches. *Physics-driven* approaches cast sound source localization as an inverse problem given the known physics of sound propagation within the considered system. *Data-driven* approaches aim at learning the desired feature-to-source-position mapping using real-world or synthetic training datasets adapted to the problem at hand. Combining these approaches should allow a widening of the notion of source localization, considering problems such as the identification of the directivity or diffuseness of the source as well as some of the boundary conditions of the room. A general perspective is to investigate the relations between the physical structure of the source and the particular structures that can be discovered or enforced in the representations and models used for characterization, localization and separation.

### **3.3. Axis 3: Large-scale Audio Content Processing and Self-organization**

#### ***3.3.1. Motif Discovery in Audio Data***

Facing the ever-growing quantity of multimedia content, the topic of motif discovery and mining has become an emerging trend in multimedia data processing with the ultimate goal of developing weakly supervised paradigms for content-based analysis and indexing. In this context, speech, audio and music content, offers a particularly relevant information stream from which meaningful information can be extracted to create some form of “audio icons” (key-sounds, jingles, recurrent locutions, musical choruses, etc ...) without resorting to comprehensive inventories of expected patterns.

This challenge raises several fundamental questions that will be among our core preoccupations over the next few years. The first question is the deployment of motif discovery on a large scale, a task that requires extending audio motif discovery approaches to incorporate efficient time series pattern matching methods (fingerprinting, similarity search indexing algorithms, stochastic modeling, etc.). The second question is that of the use and interpretation of the motifs discovered. Linking motif discovery and symbolic learning techniques, exploiting motif discovery in machine learning are key research directions to enable the interpretation of recurring motifs.

On the application side, several use cases can be envisioned which will benefit from motif discovery deployed on a large scale. For example, in spoken content, word-like repeating fragments can be used for several spoken document-processing tasks such as language-independent topic segmentation or summarization. Recurring motifs can also be used for audio summarization of audio content. More fundamentally, motif discovery paves the way for a shift from supervised learning approaches for content description to unsupervised paradigms where concepts emerge from the data.

#### ***3.3.2. Structure Modeling and Inference in Audio and Musical Contents***

Structuring information is a key step for the efficient description and learning of all types of contents, and in particular audio and musical contents. Indeed, structure modeling and inference can be understood as the task of detecting dependencies (and thus establishing relationships) between different fragments, parts or sections of information content.

A stake of structure modeling is to enable more robust descriptions of the properties of the content and better model generalization abilities that can be inferred from a particular content, for instance via cache models, trigger models or more general graphical models designed to render the information gained from structural inference. Moreover, the structure itself can become a robust descriptor of the content, which is likely to be more resistant than surface information to a number of operations such as transmission, transduction, copyright infringement or illegal use.

In this context, information theory concepts need to be investigated to provide criteria and paradigms for detecting and modeling structural properties of audio contents, covering potentially a wide range of application domains in speech content mining, music modeling or audio scene monitoring.

## SEMAGRAMME Project-Team

### 3. Research Program

#### 3.1. Overview

The Sémagramme project relies on deep mathematical foundations. We intend to develop models based on well-established mathematics. We seek two main advantages from this approach. On the one hand, by relying on mature theories, we have at our disposal sets of mathematical tools that we can use to study our models. On the other hand, developing various models on a common mathematical background will make them easier to integrate, and will ease the search for unifying principles.

The main mathematical domains on which we rely are formal language theory, symbolic logic, and type theory.

#### 3.2. Formal language theory

Formal language theory studies the purely syntactic and combinatorial aspects of languages, seen as sets of strings (or possibly trees or graphs). Formal language theory has been especially fruitful for the development of parsing algorithms for context-free languages. We use it, in a similar way, to develop parsing algorithms for formalisms that go beyond context-freeness. Language theory also appears to be very useful in formally studying the expressive power and the complexity of the models we develop.

#### 3.3. Symbolic logic

Symbolic logic (and, more particularly, proof-theory) is concerned with the study of the expressive and deductive power of formal systems. In a rule-based approach to computational linguistics, the use of symbolic logic is ubiquitous. As we previously said, at the level of syntax, several kinds of grammars (generative, categorial...) may be seen as basic deductive systems. At the level of semantics, the meaning of an utterance is captured by computing (intermediate) semantic representations that are expressed as logical forms. Finally, using symbolic logics allows one to formalize notions of inference and entailment that are needed at the level of pragmatics.

#### 3.4. Type theory and typed $\lambda$ -calculus

Among the various possible logics that may be used, Church's simply typed  $\lambda$ -calculus and simple theory of types (a.k.a. higher-order logic) play a central part. On the one hand, Montague semantics is based on the simply typed  $\lambda$ -calculus, and so is our syntax-semantics interface model. On the other hand, as shown by Gallin [39], the target logic used by Montague for expressing meanings (i.e., his intensional logic) is essentially a variant of higher-order logic featuring three atomic types (the third atomic type standing for the set of possible worlds).



## CHROMA Team

### 3. Research Program

#### 3.1. Introduction

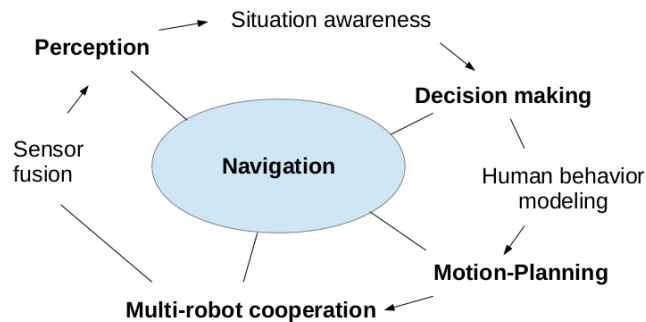


Figure 1. Research themes of the team and their relation

The Chroma team aims to deal with different issues of autonomous mobile robotics : perception, decision-making and cooperation. Figure 1 schemes the different themes and sub-themes investigated by Chroma.

We present here after our approaches to address these different theme of research, and how they combine to contribute to the general problem of robotic navigation. Chroma pays particular attention to current challenges that are autonomous navigation in highly dynamic environments populated by humans and cooperation in (large) multi-robot systems. These challenges are common with other major robotic laboratories/teams in the world, such as Autonomous Systems Lab at ETH Zurich, Robotic Embedded Systems Laboratory at USC, KIT<sup>0</sup> (Prof Christoph Stiller lab and Prof Ruediger Dillmann lab), UC Berkeley, Vislab Parma (Prof. Alberto Broggi), iCeiRA<sup>0</sup> laboratory in Taipei. Chroma is collaborating at various levels (visits, postdocs, research projects, common publications, etc.) with most of these laboratories, see Sections 9.3 and 9.4 .

#### 3.2. Perception and Situation Awareness

**Participants:** Christian Laugier, Agostino Martinelli, Jilles S. Dibangoye, Anne Spalanzani, Olivier Simonin.

Robust perception in open and dynamic environments populated by human beings is an open and challenging scientific problem. Traditional perception techniques do not provide an adequate solution for these problems, mainly because such environments are uncontrolled<sup>0</sup> and exhibit strong constraints to be satisfied (in particular high dynamicity and uncertainty). This means that **the proposed solutions have to simultaneously take into account characteristics such as real time processing, temporary occultations, dynamic changes or motion predictions.**

<sup>0</sup>Karlsruhe Institut für Technologie

<sup>0</sup>International Center of Excellence in Intelligent Robotics and Automation Research.

<sup>0</sup>partially unknown and open



Figure 2. Illustrations of the HSBOF model and experiment with the Zoe car

### 3.2.1. Bayesian perception

**Context and previous work.** Perception is known to be one of the main bottleneck for robot motion autonomy, in particular when navigating in open and dynamic environments is subject to strong real-time and uncertainty constraints. Traditional object-level solutions<sup>0</sup> still exhibit a lack of efficiency and of robustness when operating in such complex environments. In order to overcome this difficulty, we have proposed in the scope of the former e-Motion team, a new paradigm in robotics called “Bayesian Perception”. The foundation of this approach relies on the concept of “Bayesian Occupancy Filter (BOF)” initially proposed in the PhD thesis of Christophe Coué [42] and further developed in the team [58]<sup>0</sup>.

The basic idea is to combine a Bayesian filter with a probabilistic grid representation of both the space and the motions, see illustration Fig. 2. This new approach can be seen as an extension for uncertain dynamic scenes, of the initial concept of “Occupancy Grid” proposed in 1989 by Elfes<sup>0</sup>. It allows the filtering and the fusion of heterogeneous and uncertain sensors data, by taking into account the history of the sensors measurements, a probabilistic model of the sensors and of the uncertainty, and a dynamic model of the observed objects motions.

In the scope of the Chroma team and of several academic and industrial projects, we went on with the development and the extension under strong embedded implementation constraints, of our Bayesian Perception concept. This work has already led to the development of more powerful models and more efficient implementations, e.g. the *HSBOF*<sup>0</sup> approach [76] and the *CMCDOT*<sup>0</sup> framework [84] which is still under development.

Current and future work address the extension of this model and its software implementation.

**Objective — Extending the Bayesian Perception paradigm to the object level —** We aim at defining a complete framework extending the Bayesian Perception paradigm to the object level. The main objective is to be simultaneously more robust, more efficient for embedded implementations, and more informative for the subsequent scene interpretation step.

We propose to integrate in a robust way higher level functions such as multiple objects detection and tracking or objects classification. The idea is to avoid well known object level detection errors and data association problems, by simultaneously reasoning at the *grid level* and at the *object level* by extracting / identifying / tracking / classifying clusters of dynamic cells (first work has been published in [84]).

<sup>0</sup>object recognition based on image processing

<sup>0</sup>The *Bayesian programming formalism* developed in e-Motion, pioneered (together with the contemporary work of Thrun, Burgard and Fox [94]) a systematic effort to formalize robotics problems under Probability theory –an approach that is now pervasive in Robotics.

<sup>0</sup>A. Elfes. “Occupancy grids: A probabilistic framework for robot perception and navigation”, Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, USA, 1989.

<sup>0</sup>Hybrid Sampling Bayesian Occupancy Filter

<sup>0</sup>Conditional Monte Carlo Dense Occupancy Tracker

**Software development : new approaches for software / hardware integration** The objective is to improve the efficiency of the approach (by exploiting the highly parallel characteristic of our approach), while drastically reducing important factors such as the required memory size, the size of the hardware component, its price and the required energy consumption. This work is absolutely necessary for studying *embedded solutions* for the future generation of mobile robots and autonomous vehicles.

### 3.2.2. Situation Awareness and Prediction

**Context.** Prediction of the evolution of the perceived actors is an important ability required for navigation in dynamic and uncertain environments, in particular to allow on-line safe decisions. We have recently shown that an interesting property of the Bayesian Perception approach is to generate short-term conservative<sup>0</sup> predictions on the likely future evolution of the observed scene, even if the sensing information is temporary incomplete or not available [76]. But in human populated environments, estimating more abstract properties (e.g. object classes, affordances, agents intentions) is also crucial to understand the future evolution of the scene.

**Objective** We aim to develop an integrated approach for “Situation Awareness & Risk Assessment” in complex dynamic scenes involving multiples moving agents (e.g vehicles, cyclists, pedestrians ...), whose behaviors are most of the time unknown but predictable.

Our approach relies on combining machine learning to build a model of the agent behaviors and generic motion prediction techniques (Kalman-based, GHMM [98], Gaussian Processes [92]). A strong challenge of prediction in multiple moving agents environments is to take into consideration the interactions between the different agents (traffic participants). Existing interaction-aware approaches estimate exhaustively the intent of all road users [59], assume a cooperative behavior [88], or learn the policy model of drivers using supervised learning [50]. In contrast, we adopt a *planning-based motion prediction approach*, which is a framework to predict human behavior [102], [56][27]. Planning-based approaches assume that humans, when they perform a task, they do so by minimizing a cost function that depends on their preferences and the context. Such a cost function can be obtained, for example, from demonstrations using *Inverse Reinforcement Learning* [75], [36]. This constitutes an intuitive approach and, more importantly, enables us to overcome the limitations of other approaches, namely, high complexity [59], unrealistic assumptions [88], and overfitting [50]. We have recently demonstrated the predictive potential of our approach in [26].

**Towards an On-line Bayesian Decision-Making framework.** The team aims at building a general framework for perception and decision-making in multi-robot/vehicle environments. The navigation will be performed under both dynamic and uncertainty constraints, with contextual information and a continuous analysis of the evolution of the probabilistic collision risk (see above). Results have recently been obtained in cooperation with Renault and Berkeley, by using the “Intention / Expectation” paradigm and Dynamic Bayesian Networks; these results have been published in [60], [61] and patented.

We are currently working on the generalization of this approach, in order to take into account the dynamics of the vehicles and multiple traffic participants. The objective is to design a new framework, allowing to overcome the shortcomings of rules-based reasoning approaches usually showing good results [64] [49], but leading to a lack of scalability and long terms predictions. Our research work is carried out through several cooperative projects (Toyota, Renault, project Prefect of IRT Naoelec, European project ECSEL Enable-S3) and related PhD theses.

### 3.2.3. Robust state estimation (Sensor fusion)

**Context.** In order to safely and autonomously navigate in an unknown environment, a mobile robot is required to estimate in real time several physical quantities (e.g., position, orientation, speed). These physical quantities are often included in a common state vector and their simultaneous estimation is usually achieved by fusing the information coming from several sensors (e.g., camera, laser range finder, inertial sensors). The problem of fusing the information coming from different sensors is known as the *Sensor Fusion* problem and it is a fundamental problem which plays a major role in robotics.

<sup>0</sup>i.e. when motion parameters are supposed to be stable during a small amount of time

**Objective.** A fundamental issue to be investigated in any sensor fusion problem is to understand whether the state is observable or not. Roughly speaking, we need to understand if the information contained in the measurements provided by all the sensors allows us to carry out the estimation of the state. If the state is not observable, we need to detect a new observable state. This is a fundamental step in order to properly define the state to be estimated. To achieve this goal, we apply standard analytic tools developed in control theory together with some new theoretical concepts we introduced in [68] (concept of continuous symmetry). Additionally, we want to account the presence of disturbances in the observability analysis.

Our approach is to introduce general analytic tools able to derive the observability properties in the nonlinear case when some of the system inputs are unknown (and act as disturbances). We recently obtained a simple analytic tool able to account the presence of unknown inputs [71], which extends a heuristic solution derived by the team of Prof. Antonio Bicchi [40] with whom we collaborate (Centro Piaggio at the University of Pisa).

**Fusing visual and inertial data.** A special attention is devoted to the fusion of inertial and monocular vision sensors (which have strong application for instance in UAV navigation). The problem of fusing visual and inertial data has been extensively investigated in the past. However, most of the proposed methods require a state initialization. Because of the system nonlinearities, lack of precise initialization can irreparably damage the entire estimation process. In literature, this initialization is often guessed or assumed to be known [38], [63], [46]. Recently, this sensor fusion problem has been successfully addressed by enforcing observability constraints [51], [52] and by using optimization-based approaches [62], [45], [66], [53], [74]. These optimization methods outperform filter-based algorithms in terms of accuracy due to their capability of relinearizing past states. On the other hand, the optimization process can be affected by the presence of local minima. We are therefore interested in a deterministic solution that analytically expresses the state in terms of the measurements provided by the sensors during a short time-interval.

For some years we explore deterministic solutions as presented in [69] and [70]. Our objective is to improve the approach by taking into account the biases that affect low-cost inertial sensors (both gyroscopes and accelerometers) and to exploit the power of this solution for real applications. This work is currently supported by the ANR project VIMAD<sup>0</sup> and experimented with a quadrotor UAV. We have a collaboration with Prof. Stergios Roumeliotis (the leader of the MARS lab at the University of Minnesota) and with Prof. Anastasios Mourikis from the University of California Riverside. Regarding the usage of our solution for real applications we have a collaboration with Prof. Davide Scaramuzza (the leader of the Robotics and Perception group at the University of Zurich) and with Prof. Roland Siegwart from the ETHZ.

### 3.3. Navigation and cooperation in dynamic environments

**Participants:** Olivier Simonin, Anne Spalanzani, Jilles S. Dibangoye, Christian Laugier, Laetitia Matignon, Fabrice Jumel, Jacques Saraydaryan.

In his reference book *Planning algorithms*<sup>0</sup> S. LaValle discusses the different dimensions that made the motion-planning problem complex, which are the number of robots, the obstacle region, the uncertainty of perception and action, and the allowable velocities. In particular, it is emphasized that complete algorithms require at least exponential time to deal with multiple robot planning in complex environments, preventing them to be scalable in practice (p. 320). Moreover, dynamic and uncertain environments, as human-populated ones, expand this complexity.

In this context, we aim at **scale up decision-making in human-populated environments and in multi-robot systems, while dealing with the intrinsic limits of the robots (computation capacity, limited communication).**

#### 3.3.1. Motion-planning in human-populated environment

<sup>0</sup>Navigation autonome des drones aériens avec la fusion des données visuelles et inertielles, lead by A. Martinelli, Chroma.

<sup>0</sup>Steven M. LaValle, *Planning Algorithms*, Cambridge University Press, 2006.

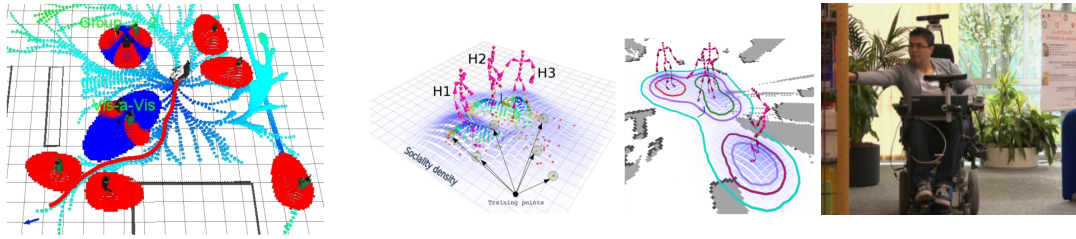


Figure 3. Illustrations of a. the Risk-RRT planning b. the human interaction space model c. experiment with the wheelchair.

**Context.** Motion planning in dynamic and human-populated environments is a current challenge of robotics. Many research teams work on this topic. We can cite the Institut of robotic in Barcelone [44], the MIT [37], the Autonomous Intelligent Systems lab in Freiburg [41], or the LAAS [77]. In Chroma, we explore different issues : **integrating the risk (uncertainty) in planning processes, modeling and taking into account human behaviors and flows.**

**Objective** We aim to give the robot some socially compliant behaviors by anticipating the near future (trajectories of mobile obstacle in the robot's surroundings) and by integrating knowledge from psychology, sociology and urban planning. In this context, we will focus on the following 3 topics.

**Risk-based planning.** Unlike static or controlled environments<sup>0</sup> where global path planning approaches are suitable, dealing with highly dynamic and uncertain environments requires to integrate the notion of risk (risk of collision, risk of disturbance). This risk can be computed by methods proposed in the section 3.2.2 . Then, we examine how motion planning approaches can integrate this risk in the generation and selection of the paths. An algorithm called RiskRRT was proposed in the eMotion team. This algorithm plans goal oriented trajectories that minimize the risk estimated at each instant. It fits environments that are highly dynamic and adapts to a representation of uncertainty [90] (see Figure 3 .a for illustration). Now, we aim to extend this principle to be adapted to various risk evaluation methods (proposed in 3.2 ) and various situation (highways, urban environments, even in dense traffic or in crowds).

**Sharing the physical space with humans.** Robots are expected to share their physical space with humans. Hence, robots need to take into account the presence of humans and to behave in a socially acceptable way. Their trajectories must be safe but also predictable, that is why they must follow social conventions, respecting proximity constraints, avoiding people interacting or joining a group engaged in conversation without disturbing. For this purpose, we proposed earlier to integrate some knowledge from the psychology domain (i.e. proxemics theory), see figure 3 .b. We aim now to integrate semantic knowledge<sup>0</sup> and psycho-social theories of human behavior<sup>00</sup> in the navigation framework we have developed for a few years (i.e. the Risk-based navigation algorithms [48], [90], [97]). These concepts were tested on our automated wheelchair (see figure 3 .c) but they have and will be adapted to autonomous cars, telepresence robots and companion robots. This work is currently supported by the ANR Valet, the TENSIVE project and the Associated team Sampen (with Iceira Lab, Taipei).

<sup>0</sup>known environment without uncertainty

<sup>0</sup>B. Kuipers, The Spatial Semantic Hierarchy, Artificial Intelligence, Volume 119, Issues 1–2, May 2000, Pages 191-233

<sup>0</sup>Gibson, J. (1977). The theory of affordances, in Perceiving, Acting, and Knowing. Towards an Ecological Psychology. Number eds

Shaw R., Bransford J. Hoboken,NJ: John Wiley & Sons Inc.

<sup>0</sup>Hall, E. (1966). The hidden dimension. Doubleday Anchor Books.

**Mapping human flows.** We investigate the problem of modeling recurring human displacements to improve robots navigation in such dense populated environments. It has been shown that such recurring behaviours can be mapped from spatial-temporal observations, as in [95]. In this context we address the problem of mapping human flows from robot(s) perception. We started to propose counting-based mapping models [30] that contain motion probabilities in the grid cells. Then such a grid can be exploited to define path-planning functions (eg. A\* based) that take into account the probability to encounter humans in opposite direction. We also aim at demonstrating the efficiency of the approach which real robots evolving in dense human-populated environments.

### 3.3.2. Decision Making in Multi-robot systems

**Context.** A central challenge in Chroma is to define **decision-making algorithms that scale up to large multi-robot systems**. This work takes place in the general framework of Multi-Agent Systems (MAS). The objective is to compute/define agent behaviors that provide cooperation and adaptation abilities. Solutions must also take into account the agent/robot computational limits.

We can abstract the challenge in three objectives :

- i) mastering the complexity of large fleet of robots/vehicles (scalability),
- ii) dealing with limited computational/memory capacity
- iii) building adaptive solutions (robustness).

#### Combining Decision-theoretic models and Swarm intelligence.

Over the past few years, our attempts to address multi-robot decision-making are mainly due to Multi-Agent Sequential Decision Making (MA-SDM) and Swarm Intelligence (SI). MA-SDM builds upon well-known decision-theoretic models (e.g., Markov decision processes and games) and related algorithms, that come with strong theoretical guarantees. In contrast, the expressiveness of MA-SDM models has limited scalability in face of realistic multi-robot systems<sup>0</sup>, resulting in computational overload. On their side, SI methods, which rely on local rules – generally bio-inspired – and relating to Self-Organized Systems<sup>0</sup>, can scale up to multiple robots and provide robustness to disturbances, but with poor theoretical guarantees<sup>0</sup>. Swarm models can also answer to the need of designing tractable solutions [89], but they remain not geared to express complex realistic tasks or to handle (point-to-point) communication between robots. This motivates our work to go beyond these two approaches and to combine them.

First, we plan to investigate **incremental expansion mechanisms in anytime decision-theoretic planning**, starting from local rules (from SI) to complex strategies with performance guarantees (from MA-SDM) [43]. This methodology is grounded into our research on anytime algorithms, that are guaranteed to stop at anytime while still providing a reliable solution to the original problem. It further relies on decision theoretical models and tools including: Decentralized and Partially Observable Markov Decision Processes and Games, Dynamic Programming, Distributed Reinforcement Learning and Statistical Machine Learning.

Second, we plan to extend the SI approach by considering **the integration of optimization techniques at the local level**. The purpose is to force the system to explore solutions around the current stabilized state – potentially a local optimum – of the system. We aim at keeping scalability and self-organization properties by not compromising the decentralized nature of such systems. Introducing optimization in this way requires to measure locally the performances, which is generally possible from local perception of robots (or using learning techniques). The main optimization methods we will consider are Local Search (Gradient Descent), Distributed Stochastic Algorithm and Reinforcement Learning. We have shown in [96] the interest of such an approach for driverless vehicle traffic optimization. In 2016, we started a new PHD in collaboration with the VOLVO Group to deal with global-local optimization for goods distribution using a fleet of autonomous vehicles.

<sup>0</sup>Martin L. Puterman, Markov Decision Processes; Stuart Russell and Peter Norvig, Artificial Intelligence - A Modern Approach

<sup>0</sup>D. Floreano and C. Mattiussi, Bio-Inspired Artificial Intelligence - Theories, Methods, and Technologies, MIT Press, 2008.

<sup>0</sup>S. A. Brueckner, G. Di Marzo Serugendo, A. Karageorgos, R. Nagpal (2005). Engineering Self-Organising Systems, Methodologies and Applications. LNAI 3464 State-of-the-Art Survey, Springer book.

Both approaches must lead to **master the complexity** inherent to large and open multi-robot systems. Such systems are prone to combinatorial problems, in term of state space and communication, when the number of robots grows. To cope with this complexity we started to develop a methodology which relies on incrementally refining the environment representation while the robots perform their tasks.

Mastering the computational cost involved in cooperative decision-making relies also on building heuristics. We explore how exact (global) solutions can be decentralized in local computation allocated to group of robots or to each robot. We started to apply this methodology to dynamic problems such as the patrolling of moving persons (see [87]).

Beyond this methodological work, we aim to evaluate our models on benchmarks from the literature, by using simulation tools as a complement of robotic experiments. This will lead us to develop simulators, allowing to deploy thousands of humans and robots in constrained environments.

### **Towards adaptive connected robots.**

Mobile robots and autonomous vehicles are becoming more connected to one another and to other devices in the environment (concept of cloud of robots<sup>0</sup> and V2V/V2I connectivity in transportation systems). Such robotic systems are open systems as the number of connected entities is varying dynamically. Network of robots brought with them new problems, as the need of (online) adaption to changes in the system and to the variability of the communication.

In Chroma, we address the problem of adaptation by considering machine learning techniques and local mechanisms as discussed above (SI models). More specifically we investigate the problem of maintaining the connectivity between robots which perform dynamic version of tasks such as patrolling, exploration or transportation, i.e. where the setting of the problem is continuously changing and growing (see [25]).

Robot fleets should be able to adapt their behavior and organisation to communication limits and variation. It has been shown that wireless communication are very changing in time and space [65]. So we explore how robots can optimize their behaviors by perceiving and learning the quality of their communication in the environment. In Lyon, the CITI Lab. conducts research in many aspects of telecommunication, from signal theory to distributed computation. In this context, Chroma develops cooperations with the Inria team Urbanet [25] (wireless communication protocols) and with the Dynamid team [19] (middleware and cloud aspects), that we wish to reinforce in the next years.

---

<sup>0</sup>see for instance the first International Workshop on Cloud and Robotics, 2016.

## DEFROST Team

### 3. Research Program

#### 3.1. Introduction

Our research crosses different disciplines: numerical mechanics, control design, robotics, optimisation methods, clinical applications. Our organisation aims at facilitating the team work and cross-fertilisation of research results in the group. We have three objectives (1, 2 and 3) that correspond to the main scientific challenges. In addition, we have two transversal objectives that are also highly challenging: the development of a high performance software support for the project (objective 4) and the validation tools and protocols for the models and methods (objective 5).

#### 3.2. Objective 1: Accurate model of soft robot deformation computed in finite time

The objective is to find concrete numerical solutions to the challenge of modelling soft robots with strong real-time constraints. To solve continuum mechanics equations, we will start our research with real-time FEM or equivalent methods that were developed for soft-tissue simulation. We will extend the functionalities to account for the needs of a soft-robotic system:

- Coupling with other physical phenomena that govern the activity of sensors and actuators (hydraulic, pneumatic, electro-active polymers, shape-memory alloys...).
- Fulfill the new computational time constraints (harder than surgical simulation for training) and find better tradeoff between cost and precision of numerical solvers using reduced-order modelling techniques with error control.
- Exploring interactive and semi-automatic optimisation methods for design based on obtained solution for fast computation on soft robot models.

#### 3.3. Objective 2: Model based control of soft robot behavior

The focus of this objective is on obtaining a generic methodology for soft robot feedback control. Several steps are needed to design a model based control from FEM approach:

- The fundamental question of the kinematic link between actuators, sensors, effectors and contacts using the most reduced mathematical space must be carefully addressed. We need to find efficient algorithms for real-time projection of non-linear FEM models in order to pose the control problem using the only relevant parameters of the motion control.
- Intuitive remote control is obtained when the user directly controls the effector motion. To add this functionality, we need to obtain real-time inverse models of the soft robots by optimisation. Several criteria will be combined in this optimisation: effector motion control, structural stiffness of the robot, reduce intensity of the contact with the environment...
- Investigating closed-loop approaches using sensor feedback: as sensors cannot monitor all points of the deformable structure, the information provided will only be partial. We will need additional algorithms based on the FEM model to obtain the best possible treatment of the information. The final objective of these models and algorithms is to have robust and efficient feedback control strategies for soft robots. One of the main challenge here is to ensure / prove stability in closed-loop.



### 3.4. Objective 3: Modeling the interaction with a complex environment

Even if the inherent mechanical compliance of soft robots makes them more safe, robust and particularly adapted to interaction with fragile environments, the contact forces need to be controlled by:

- Setting up real-time modelling and the control methods needed to pilot the forces that the robot imposes on its environment and to control the robot deformations imposed by its environment. Note that if an operative task requires to apply forces on the surrounding structures, the robot must be anchored to other structures or structurally rigidified.
- Providing mechanics models of the environment that include the uncertainties on the geometry and on the mechanical properties, and are capable of being readjusted in real-time.
- Using the visual feedback of the robot behavior to adapt dynamically the models. The observation provided in the image coupled with an inverse accurate model of the robot could transform the soft robot into sensor: as the robot deforms with the contact of the surroundings, we could retrieve some missing parameters of the environment by a smart monitoring of the robot deformations.

### 3.5. Objective 4: Soft Robotic Software

Expected research results of this project are numerical methods and algorithms that require high-performance computing and suitability with robotic applications. There is no existing software support for such development. We propose to develop our own software, in a suite split into three applications:

- The first one will facilitate the design of deformable robots by an easy passage from CAD software (for the design of the robot) to the FEM based simulation
- The second one is an anticipative clinical simulator. The aim is to co-design the robotic assistance with the physicians, thanks to a realistic simulation of the procedure or the robotic assistance. This will facilitate the work of reflection on new clinical approaches prior any manufacturing
- The third one is the control design software. It will provide the real-time solutions for soft robot control developed in the project.

### 3.6. Objective 5: Validation and application demonstrations

The implementation of experimental validation is a key challenge for the project. On one side, we need to validate the model and control algorithms using concrete test case example in order to improve the modelling and to demonstrate the concrete feasibility of our methods. On the other side, concrete applications will also feed the reflexions on the objectives of the scientific program.

We will build our own experimental soft robots for the validation of objective 2 and 3 when there is no existing « turn-key » solution. Designing and making our own soft robots, even if only for validation, will help the setting-up of adequate models.

For the validation of objective 4, we will develop « anatomical soft robot »: soft robot with the shape of organs, equipped with sensors (to measure the contact forces) and actuators (to be able to stiffen the walls and recreate natural motion of soft-tissues). We will progressively increase the level of realism of this novel validation set-up to come closer to the anatomical properties.

## FLOWERS Project-Team

### 3. Research Program

#### 3.1. Research Program

Research in artificial intelligence, machine learning and pattern recognition has produced a tremendous amount of results and concepts in the last decades. A blooming number of learning paradigms - supervised, unsupervised, reinforcement, active, associative, symbolic, connectionist, situated, hybrid, distributed learning... - nourished the elaboration of highly sophisticated algorithms for tasks such as visual object recognition, speech recognition, robot walking, grasping or navigation, the prediction of stock prices, the evaluation of risk for insurances, adaptive data routing on the internet, etc... Yet, we are still very far from being able to build machines capable of adapting to the physical and social environment with the flexibility, robustness, and versatility of a one-year-old human child.

Indeed, one striking characteristic of human children is the nearly open-ended diversity of the skills they learn. They not only can improve existing skills, but also continuously learn new ones. If evolution certainly provided them with specific pre-wiring for certain activities such as feeding or visual object tracking, evidence shows that there are also numerous skills that they learn smoothly but could not be “anticipated” by biological evolution, for example learning to drive a tricycle, using an electronic piano toy or using a video game joystick. On the contrary, existing learning machines, and robots in particular, are typically only able to learn a single pre-specified task or a single kind of skill. Once this task is learnt, for example walking with two legs, learning is over. If one wants the robot to learn a second task, for example grasping objects in its visual field, then an engineer needs to re-program manually its learning structures: traditional approaches to task-specific machine/robot learning typically include engineer choices of the relevant sensorimotor channels, specific design of the reward function, choices about when learning begins and ends, and what learning algorithms and associated parameters shall be optimized.

As can be seen, this requires a lot of important choices from the engineer, and one could hardly use the term “autonomous” learning. On the contrary, human children do not learn following anything looking like that process, at least during their very first years. Babies develop and explore the world by themselves, focusing their interest on various activities driven both by internal motives and social guidance from adults who only have a folk understanding of their brains. Adults provide learning opportunities and scaffolding, but eventually young babies always decide for themselves what activity to practice or not. Specific tasks are rarely imposed to them. Yet, they steadily discover and learn how to use their body as well as its relationships with the physical and social environment. Also, the spectrum of skills that they learn continuously expands in an organized manner: they undergo a developmental trajectory in which simple skills are learnt first, and skills of progressively increasing complexity are subsequently learnt.

A link can be made to educational systems where research in several domains have tried to study how to provide a good learning experience to learners. This includes the experiences that allow better learning, and in which sequence they must be experienced. This problem is complementary to that of the learner that tries to learn efficiently, and the teacher here has to use as efficiently the limited time and motivational resources of the learner. Several results from psychology [112] and neuroscience [22] have argued that the human brain feels intrinsic pleasure in practicing activities of optimal difficulty or challenge. A teacher must exploit such activities to create positive psychological states of flow [124].

A grand challenge is thus to be able to build robotic machines that possess this capability to discover, adapt and develop continuously new know-how and new knowledge in unknown and changing environments, like human children. In 1950, Turing wrote that the child’s brain would show us the way to intelligence: “Instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulates the child’s” [174]. Maybe, in opposition to work in the field of Artificial Intelligence who has focused on mechanisms trying to match the capabilities of “intelligent” human adults such as chess playing or natural

language dialogue [134], it is time to take the advice of Turing seriously. This is what a new field, called developmental (or epigenetic) robotics, is trying to achieve [145] [178]. The approach of developmental robotics consists in importing and implementing concepts and mechanisms from developmental psychology [148], cognitive linguistics [123], and developmental cognitive neuroscience [139] where there has been a considerable amount of research and theories to understand and explain how children learn and develop. A number of general principles are underlying this research agenda: embodiment [116] [156], grounding [132], situatedness [105], self-organization [172] [158], enaction [176], and incremental learning [119].

Among the many issues and challenges of developmental robotics, two of them are of paramount importance: exploration mechanisms and mechanisms for abstracting and making sense of initially unknown sensorimotor channels. Indeed, the typical space of sensorimotor skills that can be encountered and learnt by a developmental robot, as those encountered by human infants, is immensely vast and inhomogeneous. With a sufficiently rich environment and multimodal set of sensors and effectors, the space of possible sensorimotor activities is simply too large to be explored exhaustively in any robot's life time: it is impossible to learn all possible skills and represent all conceivable sensory percepts. Moreover, some skills are very basic to learn, some other very complicated, and many of them require the mastery of others in order to be learnt. For example, learning to manipulate a piano toy requires first to know how to move one's hand to reach the piano and how to touch specific parts of the toy with the fingers. And knowing how to move the hand might require to know how to track it visually.

Exploring such a space of skills randomly is bound to fail or result at best on very inefficient learning [153]. Thus, exploration needs to be organized and guided. The approach of epigenetic robotics is to take inspiration from the mechanisms that allow human infants to be progressively guided, i.e. to develop. There are two broad classes of guiding mechanisms which control exploration:

1. **internal guiding mechanisms**, and in particular intrinsic motivation, responsible of spontaneous exploration and curiosity in humans, which is one of the central mechanisms investigated in FLOWERS, and technically amounts to achieve online active self-regulation of the growth of complexity in learning situations;
2. **social learning and guidance**, a learning mechanisms that exploits the knowledge of other agents in the environment and/or that is guided by those same agents. These mechanisms exist in many different forms like emotional reinforcement, stimulus enhancement, social motivation, guidance, feedback or imitation, some of which being also investigated in FLOWERS;

### 3.1.1. Internal guiding mechanisms

In infant development, one observes a progressive increase of the complexity of activities with an associated progressive increase of capabilities [148], children do not learn everything at one time: for example, they first learn to roll over, then to crawl and sit, and only when these skills are operational, they begin to learn how to stand. The perceptual system also gradually develops, increasing children perceptual capabilities other time while they engage in activities like throwing or manipulating objects. This make it possible to learn to identify objects in more and more complex situations and to learn more and more of their physical characteristics.

Development is therefore progressive and incremental, and this might be a crucial feature explaining the efficiency with which children explore and learn so fast. Taking inspiration from these observations, some roboticists and researchers in machine learning have argued that learning a given task could be made much easier for a robot if it followed a developmental sequence and "started simple" [108] [127]. However, in these experiments, the developmental sequence was crafted by hand: roboticists manually build simpler versions of a complex task and put the robot successively in versions of the task of increasing complexity. And when they wanted the robot to learn a new task, they had to design a novel reward function.

Thus, there is a need for mechanisms that allow the autonomous control and generation of the developmental trajectory. Psychologists have proposed that intrinsic motivations play a crucial role. Intrinsic motivations are mechanisms that push humans to explore activities or situations that have intermediate/optimal levels of novelty, cognitive dissonance, or challenge [112] [124] [126]. The role and structure of intrinsic motivation in humans have been made more precise thanks to recent discoveries in neuroscience showing the implication of

dopaminergic circuits and in exploration behaviours and curiosity [125] [136] [166]. Based on this, a number of researchers have begun in the past few years to build computational implementation of intrinsic motivation [153] [154] [164] [111] [137] [146] [165]. While initial models were developed for simple simulated worlds, a current challenge is to manage to build intrinsic motivation systems that can efficiently drive exploratory behaviour in high-dimensional unprepared real world robotic sensorimotor spaces [154], [153], [155], [163]. Specific and complex problems are posed by real sensorimotor spaces, in particular due to the fact that they are both high-dimensional as well as (usually) deeply inhomogeneous. As an example for the latter issue, some regions of real sensorimotor spaces are often unlearnable due to inherent stochasticity or difficulty, in which case heuristics based on the incentive to explore zones of maximal unpredictability or uncertainty, which are often used in the field of active learning [120] [133] typically lead to catastrophic results. The issue of high dimensionality does not only concern motor spaces, but also sensory spaces, leading to the problem of correctly identifying, among typically thousands of quantities, those latent variables that have links to behavioral choices. In FLOWERS, we aim at developing intrinsically motivated exploration mechanisms that scale in those spaces, by studying suitable abstraction processes in conjunction with exploration strategies.

### **3.1.2. Socially Guided and Interactive Learning**

Social guidance is as important as intrinsic motivation in the cognitive development of human babies [148]. There is a vast literature on learning by demonstration in robots where the actions of humans in the environment are recognized and transferred to robots [107]. Most such approaches are completely passive: the human executes actions and the robot learns from the acquired data. Recently, the notion of interactive learning has been introduced in [173], [113], motivated by the various mechanisms that allow humans to socially guide a robot [160]. In an interactive context the steps of self-exploration and social guidances are not separated and a robot learns by self exploration and by receiving extra feedback from the social context [173], [142] [147].

Social guidance is also particularly important for learning to segment and categorize the perceptual space. Indeed, parents interact a lot with infants, for example teaching them to recognize and name objects or characteristics of these objects. Their role is particularly important in directing the infant attention towards objects of interest that will make it possible to simplify at first the perceptual space by pointing out a segment of the environment that can be isolated, named and acted upon. These interactions will then be complemented by the children own experiments on the objects chosen according to intrinsic motivation in order to improve the knowledge of the object, its physical properties and the actions that could be performed with it.

In FLOWERS, we are aiming at including intrinsic motivation system in the self-exploration part thus combining efficient self-learning with social guidance [150], [151]. We also work on developing perceptual capabilities by gradually segmenting the perceptual space and identifying objects and their characteristics through interaction with the user [32] and robots experiments [138]. Another challenge is to allow for more flexible interaction protocols with the user in terms of what type of feedback is provided and how it is provided [144].

Exploration mechanisms are combined with research in the following directions:

### **3.1.3. Cumulative learning, reinforcement learning and optimization of autonomous skill learning**

FLOWERS develops machine learning algorithms that can allow embodied machines to acquire cumulatively sensorimotor skills. In particular, we develop optimization and reinforcement learning systems which allow robots to discover and learn dictionaries of motor primitives, and then combine them to form higher-level sensorimotor skills.

### **3.1.4. Autonomous perceptual and representation learning**

In order to harness the complexity of perceptual and motor spaces, as well as to pave the way to higher-level cognitive skills, developmental learning requires abstraction mechanisms that can infer structural information out of sets of sensorimotor channels whose semantics is unknown, discovering for example the topology of

the body or the sensorimotor contingencies (proprioceptive, visual and acoustic). This process is meant to be open-ended, progressing in continuous operation from initially simple representations towards abstract concepts and categories similar to those used by humans. Our work focuses on the study of various techniques for:

- autonomous multimodal dimensionality reduction and concept discovery;
- incremental discovery and learning of objects using vision and active exploration, as well as of auditory speech invariants;
- learning of dictionaries of motion primitives with combinatorial structures, in combination with linguistic description;
- active learning of visual descriptors useful for action (e.g. grasping);

### ***3.1.5. Embodiment and maturational constraints***

FLOWERS studies how adequate morphologies and materials (i.e. morphological computation), associated to relevant dynamical motor primitives, can importantly simplify the acquisition of apparently very complex skills such as full-body dynamic walking in biped. FLOWERS also studies maturational constraints, which are mechanisms that allow for the progressive and controlled release of new degrees of freedoms in the sensorimotor space of robots.

### ***3.1.6. Discovering and abstracting the structure of sets of uninterpreted sensors and motors***

FLOWERS studies mechanisms that allow a robot to infer structural information out of sets of sensorimotor channels whose semantics is unknown, for example the topology of the body and the sensorimotor contingencies (proprioceptive, visual and acoustic). This process is meant to be open-ended, progressing in continuous operation from initially simple representations to abstract concepts and categories similar to those used by humans.

## HEPHAISTOS Project-Team

### 3. Research Program

#### 3.1. Interval analysis

We are interested in real-valued system solving ( $f(X) = 0$ ,  $f(X) \leq 0$ ), in optimization problems, and in the proof of the existence of properties (for example, it exists  $X$  such that  $f(X) = 0$  or it exist two values  $X_1, X_2$  such that  $f(X_1) > 0$  and  $f(X_2) < 0$ ). There are few restrictions on the function  $f$  as we are able to manage explicit functions using classical mathematical operators (e.g.  $\sin(x + y) + \log(\cos(e^x) + y^2)$ ) as well as implicit functions (e.g. determining if there are parameter values of a parametrized matrix such that the determinant of the matrix is negative, without calculating the analytical form of the determinant).

Solutions are searched within a finite domain (called a *box*) which may be either continuous or mixed (i.e. for which some variables must belong to a continuous range while other variables may only have values within a discrete set). An important point is that we aim at finding all the solutions within the domain whenever the computer arithmetic will allow it: in other words we are looking for *certified* solutions. For example, for 0-dimensional system solving, we will provide a box that contains one, and only one, solution together with a numerical approximation of this solution. This solution may further be refined at will using multi-precision.

The core of our methods is the use of *interval analysis* that allows one to manipulate mathematical expressions whose unknowns have interval values. A basic component of interval analysis is the *interval evaluation* of an expression. Given an analytical expression  $F$  in the unknowns  $\{x_1, x_2, \dots, x_n\}$  and ranges  $\{X_1, X_2, \dots, X_n\}$  for these unknowns we are able to compute a range  $[A, B]$ , called the interval evaluation, such that

$$\forall \{x_1, x_2, \dots, x_n\} \in \{X_1, X_2, \dots, X_n\}, A \leq F(x_1, x_2, \dots, x_n) \leq B \quad (1)$$

In other words the interval evaluation provides a lower bound of the minimum of  $F$  and an upper bound of its maximum over the box.

For example if  $F = x \sin(x + x^2)$  and  $x \in [0.5, 1.6]$ , then  $F([0.5, 1.6]) = [-1.362037441, 1.6]$ , meaning that for any  $x$  in  $[0.5, 1.6]$  we guarantee that  $-1.362037441 \leq f(x) \leq 1.6$ .

The interval evaluation of an expression has interesting properties:

- it can be implemented in such a way that the results are guaranteed with respect to round-off errors i.e. property 1 is still valid in spite of numerical errors induced by the use of floating point numbers
- if  $A > 0$  or  $B < 0$ , then no values of the unknowns in their respective ranges can cancel  $F$
- if  $A > 0$  ( $B < 0$ ), then  $F$  is positive (negative) for any value of the unknowns in their respective ranges

A major drawback of the interval evaluation is that  $A(B)$  may be overestimated i.e. values of  $x_1, x_2, \dots, x_n$  such that  $F(x_1, x_2, \dots, x_n) = A(B)$  may not exist. This overestimation occurs because in our calculation each occurrence of a variable is considered as an independent variable. Hence if a variable has multiple occurrences, then an overestimation may occur. Such phenomena can be observed in the previous example where  $B = 1.6$  while the real maximum of  $F$  is approximately 0.9144. The value of  $B$  is obtained because we are using in our calculation the formula  $F = x \sin(y + z^2)$  with  $y, z$  having the same interval value than  $x$ .

Fortunately there are methods that allow one to reduce the overestimation and the overestimation amount decreases with the width of the ranges. The latter remark leads to the use of a branch-and-bound strategy in which for a given box a variable range will be bisected, thereby creating two new boxes that are stored in a list and processed later on. The algorithm is complete if all boxes in the list have been processed, or if during the process a box generates an answer to the problem at hand (e.g. if we want to prove that  $F(X) < 0$ , then the algorithm stops as soon as  $F(\mathcal{B}) \geq 0$  for a certain box  $\mathcal{B}$ ).

A generic interval analysis algorithm involves the following steps on the current box [1], [8], [5]:

1. *exclusion operators*: these operators determine that there is no solution to the problem within a given box. An important issue here is the extensive and smart use of the monotonicity of the functions
2. *filters*: these operators may reduce the size of the box i.e. decrease the width of the allowed ranges for the variables
3. *existence operators*: they allow one to determine the existence of a unique solution within a given box and are usually associated with a numerical scheme that allows for the computation of this solution in a safe way
4. *bisection*: choose one of the variable and bisect its range for creating two new boxes
5. *storage*: store the new boxes in the list

The scope of the HEPHAISTOS project is to address all these steps in order to find the most efficient procedures. Our efforts focus on mathematical developments (adapting classical theorems to interval analysis, proving interval analysis theorems), the use of symbolic computation and formal proofs (a symbolic pre-processing allows one to automatically adapt the solver to the structure of the problem), software implementation and experimental tests (for validation purposes).

**Important note:** We have insisted on interval analysis because this is a **major component** of our robotics activity. Our theoretical work in robotics is an analysis of the robotic environment in order to exhibit proofs on the behavior of the system that may be qualitative (e.g. the proof that a cable-driven parallel robot with more than 6 non-deformable cables will have at most 6 cables under tension simultaneously) or quantitative. In the quantitative case as we are dealing with realistic and not toy examples (including our own prototypes that are developed whenever no equivalent hardware is available or to verify our assumptions) we have to manage problems that are so complex that analytical solutions are probably out of reach (e.g. the direct kinematics of parallel robots) and we have to resort to algorithms and numerical analysis. We are aware of different approaches in numerical analysis (e.g. some team members were previously involved in teams devoted to computational geometry and algebraic geometry) but interval analysis provides us another approach with high flexibility, the possibility of managing non algebraic problems (e.g. the kinematics of cable-driven parallel robots with sagging cables, that involves inverse hyperbolic functions) and to address various types of issues (system solving, optimization, proof of existence ...).

## 3.2. Robotics

HEPHAISTOS, as a follow-up of COPRIN, has a long-standing tradition of robotics studies, especially for closed-loop robots [4], especially cable-driven parallel robots. We address theoretical issues with the purpose of obtaining analytical and theoretical solutions, but in many cases only numerical solutions can be obtained due to the complexity of the problem. This approach has motivated the use of interval analysis for two reasons:

1. the versatility of interval analysis allows us to address issues (e.g. singularity analysis) that cannot be tackled by any other method due to the size of the problem
2. uncertainties (which are inherent to a robotic device) have to be taken into account so that the *real* robot is guaranteed to have the same properties as the *theoretical* one, even in the worst case [15]. This is a crucial issue for many applications in robotics (e.g. medical or assistance robot)

Our field of study in robotics focuses on *kinematic* issues such as workspace and singularity analysis, positioning accuracy, trajectory planning, reliability, calibration, modularity management and, prominently, *appropriate design*, i.e. determining the dimensioning of a robot mechanical architecture that guarantees that the real robot satisfies a given set of requirements. The methods that we develop can be used for other robotic problems, see for example the management of uncertainties in aircraft design [6].

Our theoretical work must be validated through experiments that are essential for the sake of credibility. A contrario, experiments will feed theoretical work. Hence HEPHAISTOS works with partners on the development of real robots but also develops its own prototypes. In the last years we have developed a large number of prototypes and we have extended our development to devices that are not strictly robots but are part of an overall environment for assistance. We benefit here from the development of new miniature, low energy computers with an interface for analog and logical sensors such as the Arduino or the Phidgets. The web pages <http://www-sop.inria.fr/hephaistos/mediatheque/index.html> presents all of our prototypes and experimental work.



## LAGADIC Project-Team

### 3. Research Program

#### 3.1. Visual servoing

Basically, visual servoing techniques consist in using the data provided by one or several cameras in order to control the motions of a dynamic system [1]. Such systems are usually robot arms, or mobile robots, but can also be virtual robots, or even a virtual camera. A large variety of positioning tasks, or mobile target tracking, can be implemented by controlling from one to all the degrees of freedom of the system. Whatever the sensor configuration, which can vary from one on-board camera on the robot end-effector to several free-standing cameras, a set of visual features has to be selected at best from the image measurements available, allowing to control the desired degrees of freedom. A control law has also to be designed so that these visual features  $\mathbf{s}(t)$  reach a desired value  $\mathbf{s}^*$ , defining a correct realization of the task. A desired planned trajectory  $\mathbf{s}^*(t)$  can also be tracked. The control principle is thus to regulate the error vector  $\mathbf{s}(t) - \mathbf{s}^*(t)$  to zero. With a vision sensor providing 2D measurements, potential visual features are numerous, since 2D data (coordinates of feature points in the image, moments, ...) as well as 3D data provided by a localization algorithm exploiting the extracted 2D features can be considered. It is also possible to combine 2D and 3D visual features to take the advantages of each approach while avoiding their respective drawbacks.

More precisely, a set  $\mathbf{s}$  of  $k$  visual features can be taken into account in a visual servoing scheme if it can be written:

$$\mathbf{s} = \mathbf{s}(\mathbf{x}(\mathbf{p}(t)), \mathbf{a}) \quad (2)$$

where  $\mathbf{p}(t)$  describes the pose at the instant  $t$  between the camera frame and the target frame,  $\mathbf{x}$  the image measurements, and  $\mathbf{a}$  a set of parameters encoding a potential additional knowledge, if available (such as for instance a coarse approximation of the camera calibration parameters, or the 3D model of the target in some cases).

The time variation of  $\mathbf{s}$  can be linked to the relative instantaneous velocity  $\mathbf{v}$  between the camera and the scene:

$$\dot{\mathbf{s}} = \frac{\partial \mathbf{s}}{\partial \mathbf{p}} \dot{\mathbf{p}} = \mathbf{L}_s \mathbf{v} \quad (3)$$

where  $\mathbf{L}_s$  is the interaction matrix related to  $\mathbf{s}$ . This interaction matrix plays an essential role. Indeed, if we consider for instance an eye-in-hand system and the camera velocity as input of the robot controller, we obtain when the control law is designed to try to obtain an exponential decoupled decrease of the error:

$$\mathbf{v}_c = -\lambda \widehat{\mathbf{L}}_s^+ (\mathbf{s} - \mathbf{s}^*) - \widehat{\mathbf{L}}_s^+ \frac{\partial \mathbf{s}}{\partial t} \quad (4)$$

where  $\lambda$  is a proportional gain that has to be tuned to minimize the time-to-convergence,  $\widehat{\mathbf{L}}_s^+$  is the pseudo-inverse of a model or an approximation of the interaction matrix, and  $\frac{\partial \mathbf{s}}{\partial t}$  an estimation of the features velocity due to a possible own object motion.

From the selected visual features and the corresponding interaction matrix, the behavior of the system will have particular properties as for stability, robustness with respect to noise or to calibration errors, robot 3D trajectory, etc. Usually, the interaction matrix is composed of highly non linear terms and does not present any decoupling properties. This is generally the case when  $s$  is directly chosen as  $x$ . In some cases, it may lead to inadequate robot trajectories or even motions impossible to realize, local minimum, tasks singularities, etc. It is thus extremely important to design adequate visual features for each robot task or application, the ideal case (very difficult to obtain) being when the corresponding interaction matrix is constant, leading to a simple linear control system. To conclude in a few words, **visual servoing is basically a non linear control problem. Our Holy Grail quest is to transform it into a linear control problem.**

Furthermore, embedding visual servoing in the task function approach allows solving efficiently the redundancy problems that appear when the visual task does not constrain all the degrees of freedom of the system. It is then possible to realize simultaneously the visual task and secondary tasks such as visual inspection, or joint limits or singularities avoidance. This formalism can also be used for tasks sequencing purposes in order to deal with high level complex applications.

## 3.2. Visual tracking

Elaboration of object tracking algorithms in image sequences is an important issue for researches and applications related to visual servoing and more generally for robot vision. A robust extraction and real time spatio-temporal tracking process of visual cues is indeed one of the keys to success of a visual servoing task. If fiducial markers may still be useful to validate theoretical aspects in modeling and control, natural scenes with non-cooperative objects and subject to various illumination conditions have to be considered for addressing large scale realistic applications.

Most of the available tracking methods can be divided into two main classes: feature-based and model-based. The former approach focuses on tracking 2D features such as geometrical primitives (points, segments, circles,...), object contours, regions of interest, etc. The latter explicitly uses a model of the tracked objects. This can be either a 3D model or a 2D template of the object. This second class of methods usually provides a more robust solution. Indeed, the main advantage of the model-based methods is that the knowledge about the scene allows improving tracking robustness and performance, by being able to predict hidden movements of the object, detect partial occlusions and acts to reduce the effects of outliers. The challenge is to build algorithms that are fast and robust enough to meet our application requirements. Therefore, even if we still consider 2D feature tracking in some cases, our researches mainly focus on real-time 3D model-based tracking, since these approaches are very accurate, robust, and well adapted to any class of visual servoing schemes. Furthermore, they also meet the requirements of other classes of application, such as augmented reality.

## 3.3. Slam

Most of the applications involving mobile robotic systems (ground vehicles, aerial robots, automated submarines,...) require a reliable localization of the robot in its environment. A challenging problem is when neither the robot localization nor the map is known. Localization and mapping must then be considered concurrently. This problem is known as Simultaneous Localization And Mapping (Slam). In this case, the robot moves from an unknown location in an unknown environment and proceeds to incrementally build up a navigation map of the environment, while simultaneously using this map to update its estimated position.

Nevertheless, solving the Slam problem is not sufficient for guaranteeing an autonomous and safe navigation. The choice of the representation of the map is, of course, essential. The representation has to support the different levels of the navigation process: motion planning, motion execution and collision avoidance and, at the global level, the definition of an optimal strategy of displacement. The original formulation of the Slam problem is purely metric (since it basically consists in estimating the Cartesian situations of the robot and a set of landmarks), and it does not involve complex representations of the environment. However, it is now well recognized that **several complementary representations are needed to perform exploration, navigation, mapping, and control tasks successfully. We propose to use composite models of the environment that**

**mix topological, metric, and grid-based representations.** Each type of representation is well adapted to a particular aspect of autonomous navigation [7]: the metric model allows one to locate the robot precisely and plan Cartesian paths, the topological model captures the accessibility of different sites in the environment and allows a coarse localization, and finally the grid representation is useful to characterize the free space and design potential functions used for reactive obstacle avoidance. However, ensuring the consistency of these various representations during the robot exploration, and merging observations acquired from different viewpoints by several cooperative robots, are difficult problems. This is particularly true when different sensing modalities are involved. New studies to derive efficient algorithms for manipulating the hybrid representations (merging, updating, filtering...) while preserving their consistency are needed.

### **3.4. Scene modeling and understanding**

Long-term mapping has received an increasing amount of attention during last years, largely motivated by the growing need to integrate robots into the real world wherein dynamic objects constantly change the appearance of the scene. A mobile robot evolving in such a dynamic world should not only be able to build a map of the observed environment at a specific moment, but also to maintain this map consistent over a long period of time. It has to deal with dynamic changes that can cause the navigation process to fail. However updating the map is particularly challenging in large-scale environments. To identify changes, robots have to keep a memory of the previous states of the environment and the more dynamic it is, the higher will be the number of states to manage and the more computationally intensive will be the updating process. Mapping large-scale dynamic environments is then particularly difficult as the map size can be arbitrary large. Additionally, mapping many times the whole environment is not always possible or convenient and it is useful to take advantages of methods using only a small number of observations.

A recent trend in robotic mapping is to augment low-level maps with semantic interpretation of their content, which allows to improve the robot's environmental awareness through the use of high-level concepts. In mobile robot navigation, the so-called semantic maps have already been used to improve path planning methods, mainly by providing the robot with the ability to deal with human-understandable targets.

## LARSEN Team

### 3. Research Program

#### 3.1. Lifelong Autonomy

##### 3.1.1. Scientific Context

So far, only a few autonomous robots have been deployed for a long time (weeks, months, or years) outside of factories and laboratories. They are mostly mobile robots that simply “move around” (e.g., vacuum cleaners or museum “guides”) and data collecting robots (e.g., boats or underwater “gliders” that collect data about the water of the ocean).

A large part of the long-term autonomy community is focused on simultaneous localization and mapping (SLAM), with a recent emphasis on changing and outdoor environments [39], [50]. A more recent theme is life-long learning: during long-term deployment, we cannot hope to equip robots with everything they need to know, therefore some things will have to be learned along the way. Most of the work on this topic leverages machine learning and/or evolutionary algorithms to improve the ability of robots to react to unforeseen changes [39], [48].

##### 3.1.2. Main Challenges

**The first major challenge is to endow robots with a stable situation awareness in open and dynamic environments.** This covers both the state estimation of the robot itself as well as the perception/representation of the environment. Both problems have been claimed to be solved but it is only the case for static environments [47].

In the LARSEN team, we aim at deployment in environments shared with humans which imply dynamic objects that degrade both the mapping and localization of a robot, especially in cluttered spaces. Moreover, when robots stay longer in the environment than for the acquisition of a snapshot map, they have to face structural changes, such as the displacement of a piece of furniture or the opening or closing of a door. The current approach is to simply update an implicitly static map with all observations with no attempt at distinguishing the suitable changes. For localization in not-too-cluttered or not-too-empty environments, this is generally sufficient as a significant fraction of the environment should remain stable. But for life-long autonomy, and in particular navigation, the quality of the map, and especially the knowledge of the stable parts, is primordial.

**A second major obstacle to move robots outside of labs and factories is their fragility:** current robots often break in a few hours, if not a few minutes. This fragility mainly stems from the overall complexity of robotic systems, which involve many actuators, many sensors, and complex decisions, and from the diversity of situations that robots can encounter. Low-cost robots exacerbate this issue because they can be broken in many ways (high-quality material is expensive), because they have low self-sensing abilities (sensors are expensive and increase the overall complexity), and because they are typically targeted towards non-controlled environments (e.g., houses rather than factories, in which robots are protected from most unexpected events). More generally, this fragility is a symptom of the lack of adaptive abilities in current robots.

##### 3.1.3. Angle of Attack

To solve the state estimation problem, our approach is to combine classical estimation filters (Extended Kalman Filters, Unscented Kalman Filters, or particle filters) with a Bayesian reasoning model in order to internally simulate various configurations of the robot in its environment. This should allow for adaptive estimation that can be used as one aspect of long-term adaptation. To handle dynamic and structural changes in an environment, we aim at assessing, for each piece of observation, whether it is static or not.

We also plan to address active sensing to improve the situation awareness of robots. Literally, active sensing is the ability of an interacting agent to act so as to control what it senses from its environment with the typical objective of acquiring information about this environment. A formalism for representing and solving active sensing problems has already been proposed by members of the team [38] and we aim to use this to formalize decision making problems of improving situation awareness.

Situation awareness of robots can also be tackled by cooperation, whether it be between robots or between robots and sensors in the environment (led out intelligent spaces) or between robots and humans. This is in rupture with classical robotics, in which robots are conceived as self-contained. But, in order to cope with as diverse environments as possible, these classical robots use precise, expensive, and specialized sensors, whose cost prohibits their use in large-scale deployments for service or assistance applications. Furthermore, when all sensors are on the robot, they share the same point of view on the environment, which is a limit for perception. Therefore, we propose to complement a cheaper robot with sensors distributed in a target environment. This is an emerging research direction that shares some of the problematics of multi-robot operation and we are therefore collaborating with other teams at Inria that address the issue of communication and interoperability.

To address the fragility problem, the traditional approach is to first diagnose the situation, then use a planning algorithm to create/select a contingency plan. But, again, this calls for both expensive sensors on the robot for the diagnosis and extensive work to predict and plan for all the possible faults that, in an open and dynamic environment, are almost infinite. An alternative approach is then to skip the diagnosis and let the robot discover by trial and error a behavior that works in spite of the damage with a reinforcement learning algorithm [57], [48]. However, current reinforcement learning algorithms require hundreds of trials/episodes to learn a single, often simplified, task [48], which makes them impossible to use for real robots and more ambitious tasks.

**We therefore need to design new trial-and-error algorithms that will allow robots to learn with a much smaller number of trials (typically, a dozen).** We think the key idea is to guide online learning on the physical robot with dynamic simulations. For instance, in our recent work, we successfully mixed evolutionary search in simulation, physical tests on the robot, and machine learning to allow a robot to recover from physical damage [49], [2].

A final approach to address fragility is to deploy several robots or a swarm of robots or to make robots evolve in an active environment. We will consider several paradigms such as (1) those inspired from collective natural phenomena in which the environment plays an active role for coordinating the activity of a huge number of biological entities such as ants and (2) those based on online learning [46]. We envision to transfer our knowledge of such phenomenon to engineer new artificial devices such as an intelligent floor (which is in fact a spatially distributed network in which each node can sense, compute and communicate with contiguous nodes and can interact with moving entities on top of it) in order to assist people and robots (see the principle in [55], [46], [37]).

## 3.2. Natural Interaction with Robotic Systems

### 3.2.1. Scientific Context

Interaction with the environment is a primordial requirement for an autonomous robot. When the environment is sensorized, the interaction can include localizing, tracking, and recognizing the behavior of robots and humans. One specific issue lies in the lack of predictive models for human behavior and a critical constraint arises from the incomplete knowledge of the environment and the other agents.

On the other hand, when working in the proximity of or directly with humans, robots must be capable of safely interacting with them, which calls upon a mixture of physical and social skills. Currently, robot operators are usually trained and specialized but potential end-users of robots for service or personal assistance are not skilled robotics experts, which means that the robot needs to be accepted as reliable, trustworthy and efficient [61]. Most Human-Robot Interaction (HRI) studies focus on verbal communication [56] but applications such as assistance robotics require a deeper knowledge of the intertwined exchange of social and physical signals to provide suitable robot controllers.

### 3.2.2. Main Challenges

We are here interested in building the bricks for a situated Human-Robot Interaction (HRI) addressing both the physical and social dimension of the close interaction, and the cognitive aspects related to the analysis and interpretation of human movement and activity.

The combination of physical and social signals into robot control is a crucial investigation for assistance robots [58] and robotic co-workers [53]. A major obstacle is the control of physical interaction (precisely, the control of contact forces) between the robot and the human while both partners are moving. In mobile robots, this problem is usually addressed by planning the robot movement taking into account the human as an obstacle or as a target, then delegating the execution of this “high-level” motion to whole-body controllers, where a mixture of weighted tasks is used to account for the robot balance, constraints, and desired end-effector trajectories [43].

**The first challenge is to make these controllers easier to deploy in real robotics systems**, as currently they require a lot of tuning and can become very complex to handle the interaction with unknown dynamical systems such as humans. Here, the key is to combine machine learning techniques with such controllers.

**The second challenge is to make the robot react and adapt online to the human feedback**, exploiting the whole set of measurable verbal and non-verbal signals that humans naturally produce during a physical or social interaction. Technically, this means finding the optimal policy that adapts the robot controllers online, taking into account feedback from the human. Here, we need to carefully identify the significant feedback signals or some metrics of human feedback. In real-world conditions (i.e., outside the research laboratory environment) the set of signals is technologically limited by the robot’s and environmental sensors and the onboard processing capabilities.

**The third challenge is for a robot to be able to identify and track people on board.** The motivation is to be able to estimate online either the position, the posture, or even moods and intentions of persons surrounding the robot. The main challenge is to be able to do that online, in real-time and in cluttered environments.

### 3.2.3. Angle of Attack

Our key idea is to exploit the physical and social signals produced by the human during the interaction with the robot and the environment in controlled conditions, to learn simple models of human behavior and consequently to use these models to optimize the robot movements and actions. In a first phase, we will exploit human physical signals (e.g., posture and force measurements) to identify the elementary posture tasks during balance and physical interaction. The identified model will be used to optimize the robot whole-body control as prior knowledge to improve both the robot balance and the control of the interaction forces. Technically, we will combine weighted and prioritized controllers with stochastic optimization techniques. To adapt online the control of physical interaction and make it possible with human partners that are not robotics experts, we will exploit verbal and non-verbal signals (e.g., gaze, touch, prosody). The idea here is to estimate online from these signals the human intent along with some inter-individual factors that the robot can exploit to adapt its behavior, maximizing the engagement and acceptability during the interaction.

Another promising approach already investigated in the LARSEN team is the capability for a robot and/or an intelligent space to localize humans in its surrounding environment and to understand their activities. This is an important issue to handle both for safe and efficient human-robot interaction.

Simultaneous Tracking and Activity Recognition (STAR) [60] is an approach we want to develop. The activity of a person is highly correlated with his position, and this approach aims at combining tracking and activity recognition to benefit one from another. By tracking the individual, the system may help infer its possible activity, while by estimating the activity of the individual, the system may make a better prediction of his possible future positions (which can be very effective in case of occlusion). This direction has been tested with simulator and particle filters [45], and one promising direction would be to couple STAR with decision making formalisms like partially observable Markov decision processes, POMDPs). This would allow to formalize problems such as deciding which action to take given an estimate of the human location and activity. This could also formalize other problems linked to the active sensing direction of the team: how the robotic system

might choose its actions in order to have a better estimate of the human location and activity (for instance by moving in the environment or by changing the orientation of its cameras)?

Another issue we want to address is robotic human body pose estimation. Human body pose estimation consists of tracking body parts by analyzing a sequence of input images from single or multiple cameras.

Human posture analysis is of high value for human robot interaction and activity recognition. However, even if the arrival of new sensors like RGB-D cameras has simplified the problem, it still poses a great challenge, especially if we want to do it online, on a robot and in realistic world conditions (cluttered environment). This is even more difficult for a robot to bring together different capabilities both at the perception and navigation level [44]. This will be tackled through different techniques, going from Bayesian state estimation (particle filtering), to learning, active and distributed sensing.

## **PERSVASIVE INTERACTION Team**

### **3. Research Program**

#### **3.1. Situation Models**

Situation Modelling, Situation Awareness, Probabilistic Description Logistics

The objectives of this research area are to develop and refine new computational techniques that improve the reliability and performance of situation models, extend the range of possible application domains, and reduce the cost of developing and maintaining situation models. Important research challenges include developing machine-learning techniques to automatically acquire and adapt situation models through interaction, development of techniques to reason and learn about appropriate behaviors, and the development of new algorithms and data structures for representing situation models.

Over the next four years we will address the following research challenges:

**Techniques for learning and adapting situation models:** Hand crafting of situation models is currently an expensive process requiring extensive trial and error. We will investigate combination of interactive design tools coupled with supervised and semi-supervised learning techniques for constructing initial, simplified prototype situation models in the laboratory. One possible approach is to explore developmental learning to enrich and adapt the range of situations and behaviors through interaction with users.

**Reasoning about actions and behaviors:** Constructing systems for reasoning about actions and their consequences is an important open challenge. We will explore integration of planning techniques for operationalizing actions sequences within behaviors, and for constructing new action sequences when faced with unexpected difficulties. We will also investigate reasoning techniques within the situation modeling process for anticipating the consequences of actions, events and phenomena.

**Algorithms and data structures for situation models:** In recent years, we have experimented with an architecture for situated interaction inspired by work in human factors. This model organises perception and interaction as a cyclic process in which directed perception is used to detect and track entities, verify relations between entities, detect trends, anticipate consequences and plan actions. Each phase of this process raises interesting challenges questions algorithms and programming techniques. We will experiment alternative programming techniques representing and reasoning about situation models both in terms of difficulty of specification and development and in terms of efficiency of the resulting implementation. We will also investigate the use of probabilistic graph models as a means to better accommodate uncertain and unreliable information. In particular, we will experiment with using probabilistic predicates for defining situations, and maintaining likelihood scores over multiple situations within a context. Finally, we will investigate the use of simulation as technique for reasoning about consequences of actions and phenomena.

**Probabilistic Description Logics:** In our work, we will explore the use of probabilistic predicates for representing relations within situation models. As with our earlier work, entities and roles will be recognized using multi-modal perceptual processes constructed with supervised and semi-supervised learning [Brdiczka 07], [Barraquand 12]. However, relations will be expressed with probabilistic predicates. We will explore learning based techniques to probabilistic values for elementary predicates, and propagate these through probabilistic representation for axioms using Probabilistic Graphical Models and/or Bayesian Networks.

The challenges in this research area will be addressed through three specific research actions covering situation modelling in homes, learning on mobile devices, and reasoning in critical situations.



### ***3.1.1. Learning Routine patterns of activity in the home.***

The objective of this research action is to develop a scalable approach to learning routine patterns of activity in a home using situation models. Information about user actions is used to construct situation models in which key elements are semantic time, place, social role and actions. Activities are encoded as sequences of situations. Recurrent activities are detected as sequences of activities that occur at a specific time and place each day. Recurrent activities provide routines what can be used to predict future actions and anticipate needs and services. An early demonstration has been to construct an intelligent assistant that can respond to and filter communications.

This research action is carried out as part of the doctoral research of Julian Cumin in cooperation with researchers at Orange labs, Meylan. Results are to be published at Ubicomp, Ambient intelligence, Intelligent Environments and IEEE Transactions on System Man and Cybernetics. Julien Cumin will complete and defend his doctoral thesis in 2018.

### ***3.1.2. Learning Patterns of Activity with Mobile Devices***

The objective of this research action is to develop techniques to observe and learn recurrent patterns of activity using the full suite of sensors available on mobile devices such as tablets and smart phones. Most mobile devices include seven or more sensors organized in 4 groups: Positioning Sensors, Environmental Sensors, Communications Subsystems, and Sensors for Human-Computer Interaction. Taken together, these sensors can provide a very rich source of information about individual activity.

In this area we explore techniques to observe activity with mobiles devices in order to learn daily patterns of activity. We will explore supervised and semi-supervised learning to construct systems to recognize places and relevant activities. Location and place information, semantic time of day, communication activities, inter-personal interactions, and travel activities (walking, driving, riding public transportation, etc.) are recognized as probabilistic predicates and used to construct situation models. Recurrent sequences of situations will be detected and recorded to provide an ability to predict upcoming situations and anticipate needs for information and services.

Our goal is to develop a theory for building context aware services that can be deployed as part of the mobile applications that companies such as SNCF and RATP use to interact with clients. For example, a current project concerns systems that observe daily travel routines for the Paris region RATP metro and SNCF commuter trains. This system learns individual travel routines on the mobile device without the need to divulge information about personal travel to a cloud based system. The resulting service will consult train and metro schedules to assure that planned travel is feasible and to suggest alternatives in the case of travel disruptions. Similar applications are under discussion for the SNCF inter-city travel and Air France for air travel.

This research action is conducted in collaboration with the Inria Startup Situ8ed. The current objective is to deploy and evaluate a first prototype App during 2017. Techniques will be used commercially by Situ8ed for products to be deployed as early as 2019.

### ***3.1.3. Observing and Modelling Competence and Awareness in Critical Situations***

The aim of this research action is to experimentally evaluate and compare current theories for mental modelling for problem solving and attention in stressful situations, as well as to refine theories and techniques for observing visual fixation, attention and emotion. We are currently investigating differences in visual attention, emotional response and mental states of chess experts and chess novices solving chess problems and participating in chess matches. We observe physiological responses, mental states and visual attention using eye-tracking, long term and instantaneous face-expressions (micro-expressions), skin conductivity, blood flow (BVP), posture and other information extracted from audio-visual recordings of players.

We expect that a high degree of expertise in chess should be reflected in patterns of eye movement and emotional reaction in accordance with the game situation. Information from visual attention will be used to determine and model the degree to which a player understands the game situation in terms of abstract configurations of chess pieces rather than the positions of individual pieces. Information about the emotional reactions of players will be expressed as trajectories in the physiological space of pleasure, arousal and

dominance to determine if a player's understanding of the game situation can be observed from emotional reaction to game play.

This work is supported by the ANR project CEEGE in cooperation with the department of NeuroCognition of Univ. Bielefeld, as well as the LIG internal project AirBorne in cooperation with the French Air Force training center at ISTRE. Work in this area includes the Doctoral research of Thomas Guntz to be defended in 2019.

### **3.1.4. Bibliography**

[Brdiczka 07] O. Brdiczka, "Learning Situation Models for Context-Aware Services", Doctoral Thesis of the INPG, 25 May 2007.

[Barraquand 12] R. Barraquand, "Design of Sociable Technologies", Doctoral Thesis of the University Grenoble Alps, 2 Feb 2012.

## **3.2. Perception of People, Activities and Emotions**

Machine perception is fundamental for situated behavior. Work in this area will concern construction of perceptual components using computer vision, acoustic perception, accelerometers and other embedded sensors. These include low-cost accelerometers [Bao 04], gyroscopic sensors and magnetometers, vibration sensors, electromagnetic spectrum and signal strength (wifi, bluetooth, GSM), infrared presence detectors, and bolometric imagers, as well as microphones and cameras. With electrical usage monitoring, every power switch can be used as a sensor [Fogarty 06], [Coutaz 16]. We will develop perceptual components for integrated vision systems that combine a low-cost imaging sensor with on-board image processing and wireless communications in a small, low-cost package. Such devices are increasingly available, with the enabling manufacturing technologies driven by the market for integrated imaging sensors on mobile devices. Such technology enables the use of embedded computer vision as a practical sensor for smart objects.

Research challenges to be addressed in this area include development of practical techniques that can be deployed on smart objects for perception of people and their activities in real world environments, integration and fusion of information from a variety of sensor modalities with different response times and levels of abstraction, and perception of human attention, engagement, and emotion using visual and acoustic sensors.

Work in this research area will focus on three specific Research Actions

### **3.2.1. Multi-modal perception and modeling of activities**

The objective of this research action is to develop techniques for observing and scripting activities for common household tasks such as cooking and cleaning. An important part of this project involves acquiring annotated multi-modal datasets of activity using an extensive suite of visual, acoustic and other sensors. We are interested in real-time on-line techniques that capture and model full body movements, head motion and manipulation actions as 3D articulated motion sequences decorated with semantic labels for individual actions and activities with multiple RGB and RGB-D cameras.

We will explore the integration of 3D articulated models with appearance based recognition approaches and statistical learning for modeling behaviors. Such techniques provide an important enabling technology for context aware services in smart environments [Coutaz 05], [Crowley 15], investigated by Pervasive Interaction team, as well as research on automatic cinematography and film editing investigated by the Imagine team [Gandhi 13] [Gandhi 14] [Ronfard 14] [Galvane 15]. An important challenge is to determine which techniques are most appropriate for detecting, modeling and recognizing a large vocabulary of actions and activities under different observational conditions.

We will explore representations of behavior that encodes both temporal-spatial structure and motion at multiple levels of abstraction. We will further propose parameters to encode temporal constraints between actions in the activity classification model using a combination of higher-level action grammars [Pirsiavash 14] and episodic reasoning [Santofimia 14] [Edwards 14].

Our method will be evaluated using long-term recorded dataset that contains recordings of activities in home environments. This work will be reported in the IEEE Conference on Face and Gesture Recognition, IEEE transactions on Pattern Analysis and Machine Intelligence, (PAMI) et IEEE Transactions on Systems man and Cybernetics. This work is carried out in the doctoral research of Nachwa Abubakr in cooperation with Remi Ronfard of the Imagine Team of Inria.

### **3.2.2. Perception with low-cost integrated sensors**

In this research action, we will continue work on low-cost integrated sensors using visible light, infrared, and acoustic perception. We will continue development of integrated visual sensors that combine micro-cameras and embedded image processing for detecting and recognizing objects in storage areas. We will combine visual and acoustic sensors to monitor activity at work-surfaces. Low cost real-time image analysis procedures will be designed that acquire and process images directly as they are acquired by the sensor.

Bolometric image sensors measure the Far Infrared emissions of surfaces in order to provide an image in which each pixel is an estimate of surface temperature. Within the European MIRTIC project, Grenoble startup, ULIS has created a relatively low-cost Bolometric image sensor (Retina) that provides small images of 80 by 80 pixels taken from the Far-infrared spectrum. Each pixel provides an estimate of surface temperature. Working with Schneider Electric, engineers in the Pervasive Interaction team had developed a small, integrated sensor that combines the MIRTIC Bolometric imager with a microprocessor for on-board image processing. The package has been equipped with a fish-eye lens so that an overhead sensor mounted at a height of 3 meters has a field of view of approximately 5 by 5 meters. Real-time algorithms have been demonstrated for detecting, tracking and counting people, estimating their trajectories and work areas, and estimating posture.

Many of the applications scenarios for Bolometric sensors proposed by Schneider Electric assume a scene model that assigns pixels to surfaces of the floor, walls, windows, desks or other items of furniture. The high cost of providing such models for each installation of the sensor would prohibit most practical applications. We have recently developed a novel automatic calibration algorithm that determines the nature of the surface under each pixel of the sensor.

Work in this area will continue to develop low-cost real time infrared image sensing, as well as explore combinations of far-infrared images with RGB and RGBD images.

### **3.2.3. Observation of emotion from physiological responses in critical situations**

Recent research in Cognitive Science indicates that the human emotions result in physiological manifestations in the heart rate, skin conductance, skin color, body movements and facial expressions. It has been proposed that these manifestations can be measured by observation of skin color, body motions, and facial expressions and modeled as activation levels in three dimensions known as Valence, Arousal and Dominance. The goal of this project is to evaluate the effectiveness of visual and acoustic perception technique for measuring these physiological manifestations.

Experimental data will be collected by observing subjects engaged in playing chess. A special apparatus has been constructed that allows synchronized recording from a color camera, Kinect2 3D camera, and Tobi Eye Tracker of a player seated before a computer generated display of a chess board. The masters student will participate in the definition and recording of scenarios for recording test data, apply recently proposed techniques from the scientific literature for measuring emotions, and provide a comparative performance evaluation of various techniques. The project is expected to reveal the relative effectiveness of computer vision and other techniques for observing human emotions.

### **3.2.4. Bibliography**

[Bao 04] L. Bao, and S. S. Intille. "Activity recognition from user-annotated acceleration data.", IEEE Pervasive computing. Springer Berlin Heidelberg, pp1-17, 2004.

[Fogarty 06] J. Fogarty, C. Au and S. E. Hudson. "Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition." In Proceedings of the 19th annual ACM symposium on User interface software and technology, UIST 2006, pp. 91-100. ACM, 2006.

- [Coutaz 16] J. Coutaz and J.L. Crowley, A First-Person Experience with End-User Development for Smart Homes. *IEEE Pervasive Computing*, 15(2), pp.26-39, 2016.
- [Coutaz 05] J. Coutaz, J.L. Crowley, S. Dobson, D. Garlan, "Context is key", *Communications of the ACM*, 48 (3), 49-53, 2005.
- [Crowley 15] J. L. Crowley and J. Coutaz, "An Ecological View of Smart Home Technologies", 2015 European Conference on Ambient Intelligence, Athens, Greece, Nov. 2015.
- [Gandhi 13] Vineet Gandhi, Remi Ronfard. "Detecting and Naming Actors in Movies using Generative Appearance Models", *Computer Vision and Pattern Recognition*, 2013.
- [Gandhi 14] Vineet Gandhi, R√©mi Ronfard, Michael Gleicher. "Multi-Clip Video Editing from a Single Viewpoint", *European Conference on Visual Media Production*, 2014
- [Ronfard 14] R. Ronfard, N. Szilas. "Where story and media meet: computer generation of narrative discourse". *Computational Models of Narrative*, 2014.
- [Galvane 15] Quentin Galvane, R√©mi Ronfard, Christophe Lino, Marc Christie. "Continuity Editing for 3D Animation". *AAAI Conference on Artificial Intelligence*, Jan 2015.
- [Pirsiavash 14] Hamed Pirsiavash , Deva Ramanan, "Parsing Videos of Actions with Segmental Grammars", *Computer Vision and Pattern Recognition*, p.612-619, 2014.
- [Edwards 14] C. Edwards. 2014, "Decoding the language of human movement". *Commun. ACM* 57, 12, 12-14, November 2014.

### 3.3. Sociable Interaction with Smart Objects

Reeves and Nass argue that a social interface may be the truly universal interface [Reeves 98]. Current systems lack ability for social interaction because they are unable to perceive and understand humans or to learn from interaction with humans. One of the goals of the research to be performed in Pervasive Interaction is to provide such abilities.

Work in research area RA3 will demonstrate the use of situation models for sociable interaction with smart objects and companion robots. We will explore the use of situation models as a representation for sociable interaction. Our goal in this research is to develop methods to endow an artificial agent with the ability to acquire social common sense using the implicit feedback obtained from interaction with people. We believe that such methods can provide a foundation for socially polite man-machine interaction, and ultimately for other forms of cognitive abilities. We propose to capture social common sense by training the appropriateness of behaviors in social situations. A key challenge is to employ an adequate representation for social situations.

Knowledge for sociable interaction will be encoded as a network of situations that capture both linguistic and non-verbal interaction cues and proper behavioral responses. Stereotypical social interactions will be represented as trajectories through the situation graph. We will explore methods that start from simple stereotypical situation models and extending a situation graph through the addition of new situations and the splitting of existing situations. An important aspect of social common sense is the ability to act appropriately in social situations. We propose to learn the association between behaviors and social situation using reinforcement learning. Situation models will be used as a structure for learning appropriateness of actions and behaviors that may be chosen in each situation, using reinforcement learning to determine a score for appropriateness based on feedback obtained by observing partners during interaction.

Work in this research area will focus on four specific Research Actions

#### 3.3.1. *Moving with people*

Our objective in this area is to establish the foundations for robot motions that are aware of human social situation that move in a manner that complies with the social context, social expectations, social conventions and cognitive abilities of humans. Appropriate and socially compliant interactions require the ability for real time perception of the identity, social role, actions, activities and intents of humans. Such perception can

be used to dynamically model the current situation in order to understand the situation and to compute the appropriate course of action for the robot depending on the task at hand.

To reach this objective, we propose to investigate three interacting research areas:

- Modeling the context and situation of human activities for motion planning
- Planning and acting in a social context.
- Identifying and modeling interaction behaviors.

In particular, we will investigate techniques that allow a tele-presence robot, such as the BEAM system, to autonomously navigate in crowds of people as may be found at the entry to a conference room, or in the hallway of a scientific meeting. We will also continue experiments on autonomous motion for personal assistance robots (project PRAMAD). Work in this area includes the doctoral work of Jos,ã© da Silva, to be defended in 2019.

### ***3.3.2. Understanding and communicating intentions from motion***

This research area concerns the communication through motion. When two or more people move as a group, their motion is regulated by implicit rules that signal a shared sense of social conventions and social roles. For example, moving towards someone while looking directly at them signals an intention for engagement. In certain cultures, subtle rules dictate who passes through a door first or last. When humans move in groups, they implicitly communicate intentions with motion. In this research area, we will explore the scientific literature on proxemics and the social sciences on such movements, in order to encode and evaluate techniques for socially appropriate motion by robots.

### ***3.3.3. Socially aware interaction***

This research area concerns socially aware man-machine interaction. Appropriate and socially compliant interaction requires the ability for real time perception of the identity, social role, actions, activities and intents of humans. Such perception can be used to dynamically model the current situation in order to understand the context and to compute the appropriate course of action for the task at hand. Performing such interactions in manner that respects and complies with human social norms and conventions requires models for social roles and norms of behavior as well as the ability to adapt to local social conventions and individual user preferences. In this research area, we will complement research area 3.2 with other forms of communication and interaction, including expression with stylistic face expressions rendered on a tablet, facial gestures, body motions and speech synthesis. We will experiment with use of commercially available tool for spoken language interaction in conjunction with expressive gestures.

### ***3.3.4. Stimulating affection and persuasion with affective devices.***

This research area concerns technologies that can stimulate affection and engagement, as well as induce changes in behavior. When acting as a coach or cooking advisor, smart objects must be credible and persuasive. One way to achieve this goal is to express affective feedbacks while interacting. This can be done using sound, light and/or complex moves when the system is composed of actuators.

Research in this area will address 3 questions:

1. How do human perceive affective signals expressed by smart objects (including robots)?
2. How does physical embodiment effect perception of affect by humans?
3. What are the most effective models and tools for animation of affective expression?

Both the physical form and the range of motion have important impact on the ability of a system to inspire affection. We will create new models to propose a generic animation model, and explore the effectiveness of different forms of motion in stimulating affect.

### ***3.3.5. Bibliography***

[Reeves 98] B. Reeves and C. Nass, The Media Equation: how People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, 1998.

### 3.4. Interaction with Pervasive Smart Objects and Displays

Currently, the most effective technologies for new media for sensing, perception and experience are provided by virtual and augmented realities [Van Krevelen 2010]. At the same time, the most effective means to augment human cognitive abilities are provided by access to information spaces such as the world-wide-web using graphical user interfaces. A current challenge is to bring these two media together.

Display technologies continue to decrease exponentially, driven largely by investment in consumer electronics as well as the overall decrease in cost of microelectronics. A consequence has been an increasing deployment of digital displays in both public and private spaces. This trend is likely to accelerate, as new technologies and growth in available communications bandwidth enable ubiquitous low-cost access to information and communications.

The arrival of pervasive displays raises a number of interesting challenges for situated multi-modal interaction. For example:

1. Can we use perception to detect user engagement and identify users in public spaces?
2. Can we replace traditional pointing hardware with gaze and gesture based interaction?
3. Can we tailor information and interaction for truly situated interaction, providing the right information at the right time using the right interaction modality?
4. How can we avoid information overload and unnecessary distraction with pervasive displays?

It is increasingly possible to embed sensors and displays in clothing and ordinary devices, leading to new forms of tangible and wearable interaction with information. This raises challenges such as

1. What are the tradeoffs between large-scale environmental displays and wearable displays using technologies such as e-textiles and pico-projector?
2. How can we manage the tradeoffs between implicit and explicit interaction with both tangible and wearable interaction?
3. How can we determine the appropriate modalities for interaction?
4. How can we make users aware of interaction possibilities without creating distraction?

In addition to display and communications, the continued decrease in microelectronics has also driven an exponential decrease in cost of sensors, actuators, and computing resulting in an exponential growth in the number of smart objects in human environments. Current models for systems organization are based on centralized control, in which a controller or local hub, orchestrates smart objects, generally in connection with cloud computing. This model creates problems with privacy and ownership of information. An alternative is to organize local collections of smart objects to provide distributed services without the use of a centralized controller. The science of ecology can provide an architectural model for such organization.

This approach raises a number of interesting research challenges for pervasive interaction:

1. Can we devise distributed models for multi-modal fusion and interaction with information on heterogeneous devices?
2. Can we devise models for distributed interaction that migrates over available devices as the user changes location and task?
3. Can we manage migration of interaction over devices in a manner that provides seamless immersive interaction with information, services and media?
4. Can we provide models of distributed interaction that conserve the interaction context as services migrate?

Research Actions for Interaction with Pervasive Smart Objects for the period 2017 - 2020 include

### ***3.4.1. Situated interaction with pervasive displays***

The emergence of low-cost interactive displays will enable a confluence of virtual and physical environments. Our goal in this area is to go beyond simple graphical user interfaces in such environments to provide immersive multi-sensorial interaction and communication. A primary concern will be interaction technologies that blend visual with haptic/tactile feedback and 3D interaction and computer vision. We will investigate the use of visual-tactile feedback as well as vibratory signals to augment multi-sensorial interaction and communication. The focus will be on the phenomena of immersive interaction in real worlds that can be made possible by the blending of physical and virtual in ordinary environments.

### ***3.4.2. Wearable and tangible interaction with smart textiles and wearable projectors***

Opportunities in this area result from the emergence of new forms of interactive media using smart objects. We will explore the use of smart objects as tangible interfaces that make it possible to experience and interact with information and services by grasping and manipulating objects. We will explore the use of sensors and actuators in clothing and wearable devices such as gloves, hats and wrist bands both as a means of unobtrusively sensing human intentions and emotional states and as a means of stimulating human senses through vibration and sound. We will explore the new forms of interaction and immersion made possible by deploying interactive displays over large areas of an environment.

### ***3.4.3. Pervasive interaction with ecologies of smart objects in the home***

In this research area, we will explore and evaluate interaction with ecologies of smart objects in home environments. We will explore development of a range of smart objects that provide information services, such as devices for Episodic Memory for work surfaces and storage areas, devices to provide energy efficient control of environmental conditions, and interactive media that collect and display information. We propose to develop a new class of socially aware managers that coordinate smart objects and manage logistics in functional areas such as the kitchen, living rooms, closets, bedrooms, bathroom or office.

### ***3.4.4. Bibliography***

[Van Krevelen 10] D. W. F. Van Krevelen and R. Poelman, A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2), 1, 2010

## RITS Project-Team

### 3. Research Program

#### 3.1. Vehicle guidance and autonomous navigation

**Participants:** Zayed Alsayed, Pierre de Beaucorps, Raoul de Charette, Rafael Colmenares Prieto, Aitor Gomez Torres, Fernando Garrido Carpio, David González Bautista, Pierre Merdrignac, Alexis Meyer, Vicente Milanés, Francisco Navas, Fawzi Nashashibi, Carlos Flores, Dinh-Van Nguyen, Danut-Ovidiu Pop, Oyunchimeg Shagdar, Thomas Streubel, Guillaume Trehard, Anne Verroust-Blondet, Itheri Yahiaoui.

There are three basic ways to improve the safety of road vehicles and these ways are all of interest to the project-team. The first way is to assist the driver by giving him better information and warning. The second way is to take over the control of the vehicle in case of mistakes such as inattention or wrong command. The third way is to completely remove the driver from the control loop.

All three approaches rely on information processing. Only the last two involve the control of the vehicle with actions on the actuators, which are the engine power, the brakes and the steering. The research proposed by the project-team is focused on the following elements:

- perception of the environment,
- planning of the actions,
- real-time control.

##### 3.1.1. Perception of the road environment

**Participants:** Zayed Alsayed, Raoul de Charette, Rafael Colmenares Prieto, Aitor Gomez Torres, Pierre Merdrignac, Alexis Meyer, Fawzi Nashashibi, Dinh-Van Nguyen, Danut-Ovidiu Pop, Guillaume Trehard, Anne Verroust-Blondet, Itheri Yahiaoui.

Either for driver assistance or for fully automated guided vehicle purposes, the first step of any robotic system is to perceive the environment in order to assess the situation around itself. Proprioceptive sensors (accelerometer, gyrometer,...) provide information about the vehicle by itself such as its velocity or lateral acceleration. On the other hand, exteroceptive sensors, such as video camera, laser or GPS devices, provide information about the environment surrounding the vehicle or its localization. Obviously, fusion of data with various other sensors is also a focus of the research.

The following topics are already validated or under development in our team:

- relative ego-localization with respect to the infrastructure, i.e. lateral positioning on the road can be obtained by mean of vision (lane markings) and the fusion with other devices (e.g. GPS);
- global ego-localization by considering GPS measurement and proprioceptive information, even in case of GPS outage;
- road detection by using lane marking detection and navigable free space;
- detection and localization of the surrounding obstacles (vehicles, pedestrians, animals, objects on roads, etc.) and determination of their behavior can be obtained by the fusion of vision, laser or radar based data processing;
- simultaneous localization and mapping as well as mobile object tracking using laser-based and stereovision-based (SLAMMOT) algorithms.

Scene understanding is a large perception problem. In this research axis we have decided to use only computer vision as cameras have evolved very quickly and can now provide much more precise sensing of the scene, and even depth information. Two types of hardware setups were used, namely: monocular vision or stereo vision to retrieve depth information which allow extracting geometry information.



We have initiated several works:

- estimation of the ego motion using monocular scene flow. Although in the state of the art most of the algorithms use a stereo setup, researches were conducted to estimate the ego-motion using a novel approach with a strong assumption.
- bad weather conditions evaluations. Most often all computer vision algorithms work under a transparent atmosphere assumption which assumption is incorrect in the case of bad weather (rain, snow, hail, fog, etc.). In these situations the light ray are disrupted by the particles in suspension, producing light attenuation, reflection, refraction that alter the image processing.
- deep learning for object recognition. New works are being initiated in our team to develop deep learning recognition in the context of heterogeneous data.

### 3.1.2. Cooperative Multi-sensor data fusion

**Participants:** Pierre Merdrignac, Fawzi Nashashibi, Oyunchimeg Shagdar.

Since data are noisy, inaccurate and can also be unreliable or unsynchronized, the use of data fusion techniques is required in order to provide the most accurate situation assessment as possible to perform the perception task. RITS team worked a lot on this problem in the past, but is now focusing on collaborative perception approach. Indeed, the use of vehicle-to-vehicle or vehicle-to-infrastructure communications allows an improved on-board reasoning since the decision is made based on an extended perception.

As a direct consequence of the electronics broadly used for vehicular applications, communication technologies are now being adopted as well. In order to limit injuries and to share safety information, research in driving assistance system is now orientating toward the cooperative domain. Advanced Driver Assistance System (ADAS) and Cybercars applications are moving towards vehicle-infrastructure cooperation. In such scenario, information from vehicle based sensors, roadside based sensors and a priori knowledge is generally combined thanks to wireless communications to build a probabilistic spatio-temporal model of the environment. Depending on the accuracy of such model, very useful applications from driver warning to fully autonomous driving can be performed.

The Collaborative Perception Framework (CPF) is a combined hardware/software approach that permits to see remote information as its own information. Using this approach, a communicant entity can see another remote entity software objects as if it was local, and a sensor object, can see sensor data of others entities as its own sensor data. Last year we developed the basic hardware modules that ensure the well functioning of the embedded architecture including perception sensors, communication devices and processing tools.

Finally, since vehicle localization (ground vehicles) is an important task for intelligent vehicle systems, vehicle cooperation may bring benefits for this task. A new cooperative multi-vehicle localization method using split covariance intersection filter was developed during the year 2012, as well as a cooperative GPS data sharing method.

In the first method, each vehicle estimates its own position using a SLAM (Simultaneous Localization And Mapping) approach. In parallel, it estimates a decomposed group state, which is shared with neighboring vehicles; the estimate of the decomposed group state is updated with both the sensor data of the ego-vehicle and the estimates sent from other vehicles; the covariance intersection filter which yields consistent estimates even facing unknown degree of inter-estimate correlation has been used for data fusion.

In the second GPS data sharing method, a new collaborative localization method is proposed. On the assumption that the distance between two communicative vehicles can be calculated with a good precision, cooperative vehicle are considered as additional satellites into the user position calculation by using iterative methods. In order to limit divergence, some filtering process is proposed: Interacting Multiple Model (IMM) is used to guarantee a greater robustness in the user position estimation.

Accidents between vehicles and pedestrians (including cyclists) often result in fatality or at least serious injury for pedestrians, showing the need of technology to protect vulnerable road users. Vehicles are now equipped with many sensors in order to model their environment, to localize themselves, detect and classify obstacles, etc. They are also equipped with communication devices in order to share the information with other road users and the environment. The goal of this work is to develop a cooperative perception and communication system, which merges information coming from the communications device and obstacle detection module to improve the pedestrian detection, tracking, and hazard alarming.

Pedestrian detection is performed by using a perception architecture made of two sensors: a laser scanner and a CCD camera. The laser scanner provides a first hypothesis on the presence of a pedestrian-like obstacle while the camera performs the real classification of the obstacle in order to identify the pedestrian(s). This is a learning-based technique exploiting adaptive boosting (AdaBoost). Several classifiers were tested and learned in order to determine the best compromise between the nature and the number of classifiers and the accuracy of the classification.

### 3.1.3. Planning and executing vehicle actions

**Participants:** Fernando Garrido Carpio, David González Bautista, Vicente Milanés, Fawzi Nashashibi, Francisco Navas, Carlos Flores.

From the understanding of the environment, thanks to augmented perception, we have either to warn the driver to help him in the control of his vehicle, or to take control in case of a driverless vehicle. In simple situations, the planning might also be quite simple, but in the most complex situations we want to explore, the planning must involve complex algorithms dealing with the trajectories of the vehicle and its surroundings (which might involve other vehicles and/or fixed or moving obstacles). In the case of fully automated vehicles, the perception will involve some map building of the environment and obstacles, and the planning will involve partial planning with periodical recomputation to reach the long term goal. In this case, with vehicle to vehicle communications, what we want to explore is the possibility to establish a negotiation protocol in order to coordinate nearby vehicles (what humans usually do by using driving rules, common sense and/or non verbal communication). Until now, we have been focusing on the generation of geometric trajectories as a result of a maneuver selection process using grid-based rating technique or fuzzy technique. For high speed vehicles, Partial Motion Planning techniques we tested, revealed their limitations because of the computational cost. The use of quintic polynomials we designed, allowed us to elaborate trajectories with different dynamics adapted to the driver profile. These trajectories have been implemented and validated in the JointSystem demonstrator of the German Aerospace Center (DLR) used in the European project HAVEit, as well as in RITS's electrical vehicle prototype used in the French project ABV. HAVEit was also the opportunity for RITS to take in charge the implementation of the Co-Pilot system which processes perception data in order to elaborate the high level command for the actuators. These trajectories were also validated on RITS's cybercars. However, for the low speed cybercars that have pre-defined itineraries and basic maneuvers, it was necessary to develop a more adapted planning and control system. Therefore, we have developed a nonlinear adaptive control for automated overtaking maneuver using quadratic polynomials and Lyapunov function candidate and taking into account the vehicles kinematics. For the global mobility systems we are developing, the control of the vehicles includes also advanced platooning, automated parking, automated docking, etc. For each functionality a dedicated control algorithm was designed (see publication of previous years). Today, RITS is also investigating the opportunity of fuzzy-based control for specific maneuvers. First results have been recently obtained for reference trajectories following in roundabouts and normal straight roads.

## 3.2. V2V and V2I Communications for ITS

**Participants:** Thierry Ernst, Oyunchimeg Shagdar, Gérard Le Lann, Pierre Merdrignac, Mohammad Abualhoul, Fawzi Nashashibi.

Wireless communications are expected to play an important role for road safety, road efficiency, and comfort of road users. Road safety applications often require highly responsive and reliable information exchange between neighboring vehicles in any road density condition. Because the performance of the existing radio communications technology largely degrades with the increase of the node density, the challenge of designing wireless communications for safety applications is enabling reliable communications in highly dense scenarios. Targeting this issue, RITS has been working on medium access control design and visible light communications, especially for highly dense scenarios. The works have been carried out considering the vehicle behavior such as vehicle merging and vehicle platooning.

Unlike many of the road safety applications, the applications regarding road efficiency and comfort of road users, on the other hand, often require connectivity to the Internet. Based on our expertise in both Internet-based communications in the mobility context and in ITS, we are now investigating the use of IPv6 (Internet Protocol version 6 which is going to replace the current version, IPv4, in a few years from now) for vehicular communications, in a combined architecture allowing both V2V and V2I.

The wireless channel and the topology dynamics need to be studied when understanding the dynamics and designing efficient communications mechanisms. Targeting this issue, we have been working on channel modeling for both radio and visible light communications, and design of communications mechanisms especially for security, service discovery, multicast and geocast message delivery, and access point selection.

Below follows a more detailed description of the related research issues.

### **3.2.1. Geographic multicast addressing and routing**

**Participants:** Oyunchimeg Shagdar, Thierry Ernst.

Many ITS applications such as fleet management require multicast data delivery. Existing work on this subject tackles mainly the problems of IP multicasting inside the Internet or geocasting in the VANETs. To enable Internet-based multicast services for VANETs, we introduced a framework that:

- i) defines a distributed and efficient geographic multicast auto-addressing mechanism to ensure vehicular multicast group reachability through the infrastructure network,
- ii) introduces a simplified approach that locally manages the group membership and distributes the packets among them to allow simple and efficient data delivery.

### **3.2.2. Platooning control using visible light communications**

**Participants:** Mohammad Abualhoul, Oyunchimeg Shagdar, Fawzi Nashashibi.

The main purpose of our research is to propose and test new successful supportive communication technology, which can provide stable and reliable communication between vehicles, especially for the platooning scenario. Although VLC technology has a short history in comparison with other communication technologies, the infrastructure availability and the presence of the congestion in wireless communication channels lead to propose VLC technology as a reliable and supportive technology which can takeoff some loads of the wireless radio communication. The first objective of this work is to develop an analytical model of VLC to understand its characteristics and limitations. The second objective is to design vehicle platooning control using VLC. In platooning control, a cooperation between control and communication is strongly required in order to guarantee the platoon's stability (e.g. string stability problem). For this purpose we work on VLC model platooning scenario, to permit for each vehicle the trajectory tracking of the vehicle ahead, altogether with a prescribed inter-vehicle distance and considering all the VLC channel model limitations. The integrated channel model of the main Simulink platooning model will be responsible for deciding the availability of the Line-of-Sight for different trajectory's curvatures, which means the capability of using light communication between each couple of vehicles in the platooning queue. At the same time the model will compute all the required parameters acquired from each vehicle controller.

### **3.2.3. V2X radio communications for road safety applications**

**Participants:** Mohammad Abualhoul, Pierre Merdrignac, Oyunchimeg Shagdar, Fawzi Nashashibi.

While 5.9 GHz radio frequency band is dedicated to ITS applications, the channel and network behaviors in mobile scenarios are not very well known. In this work we theoretically and experimentally study the radio channel characteristics in vehicular networks, especially the radio quality and bandwidth availability. Based on our study, we develop mechanisms for efficient and reliable V2X communications, channel allocation, congestion control, and access point selection, which are especially dedicated to road safety and autonomous driving applications.

### 3.2.4. Safety-critical communications in intelligent vehicular networks

**Participant:** Gérard Le Lann.

Intelligent vehicular networks (IVNs) are constituents of ITS. IVNs range from platoons with a lead vehicle piloted by a human driver to fully ad-hoc vehicular networks, a.k.a. VANETs, comprising autonomous/automated vehicles. Safety issues in IVNs appear to be the least studied in the ITS domain. The focus of our work is on safety-critical (SC) scenarios, where accidents and fatalities inevitably occur when such scenarios are not handled correctly. In addition to on-board robotics, inter-vehicular radio communications have been considered for achieving safety properties. Since both technologies have known intrinsic limitations (in addition to possibly experiencing temporary or permanent failures), using them redundantly is mandatory for meeting safety regulations. Redundancy is a fundamental design principle in every SC cyber-physical domain, such as, e.g., air transportation. (Optics-based inter-vehicular communications may also be part of such redundant constructs.) The focus of our on-going work is on safety-critical (SC) communications. We consider IVNs on main roads and highways, which are settings where velocities can be very high, thus exacerbating safety problems acceptable delays in the cyber space, and response times in the physical space, shall be very small. Human lives being at stake, such delays and response times must have strict (non-stochastic) upper bounds under worst-case conditions (vehicular density, concurrency and failures). Consequently, we are led to look for deterministic solutions.

#### Rationale

In the current ITS literature, the term *safety* is used without being given a precise definition. That must be corrected. In our case, a fundamental open question is: what is the exact meaning of *SC communications*? We have devised a definition, referred to as space-time bounds acceptability (STBA) requirements. For any given problem related to SC communications, those STBA requirements serve as yardsticks for distinguishing acceptable solutions from unacceptable ones with respect to safety. In conformance with the above, STBA requirements rest on the following worst-case upper bounds:  $\lambda$  for channel access delays, and  $\Delta$  for distributed inter-vehicular coordination (message dissemination, distributed agreement).

Via discussions with foreign colleagues, notably those active in the IEEE 802 Committee, we have comforted our early diagnosis regarding existing standards for V2V/V2I/V2X communications, such as IEEE 802.11p and ETSI ITS-G5: they are totally inappropriate regarding SC communications. A major flaw is the choice of CSMA/CA as the MAC-level protocol. Obviously, there cannot be such bounds as  $\lambda$  and  $\Delta$  with CSMA/CA. Another flaw is the choice of medium-range omnidirectional communications, radio range in the order of 250 m, and interference range in the order of 400 m. Stochastic delays achievable with existing standards are just unacceptable in moderate/worst-case contention conditions. Consider the following setting, not uncommon in many countries: a highway, 3 lanes each direction, dense traffic, i.e. 1 vehicle per 12.5 m. A simple calculation leads to the following result: any vehicle may experience (destructive) interferences from up to 384 vehicles. Even if one assumes some reasonable communications activity ratio, say 25%, one finds that up to 96 vehicles may be contending for channel access. Under such conditions, MAC-level delays and string-wide dissemination/agreement delays achieved by current standards fail to meet the STBA requirements by huge margins.

Reliance on V2I communications via terrestrial infrastructures and nodes, such as road-side units or WiFi hotspots, rather than direct V2V communications, can only lead to poorer results. First, reachability is not guaranteed: hazardous conditions may develop anywhere anytime, far away from a terrestrial node. Second, mixing SC communications and ordinary communications within terrestrial nodes is a violation of the very fundamental segregation principle: SC communications and processing shall be isolated from

ordinary communications and processing. Third, security: it is very easy to jam or to spy on a terrestrial node; moreover, terrestrial nodes may be used for launching all sorts of attacks, man-in-the-middle attacks for example. Fourth, delays can only get worse than with direct V2V communications, since transiting via a node inevitably introduces additional latencies. Fifth, the delivery of every SC message must be acknowledged, which exacerbates the latency problems. Sixth, availability: what happens when a terrestrial node fails?

Trying to tweak existing standards for achieving SC communications is vain. That is also unjustified. Clearly, medium-range omnidirectional communications are unjustified for the handling of SC scenarios. By definition, accidents can only involve vehicles that are very close to each other. Therefore, short-range directional communications suffice. The obvious conclusion is that novel protocols and inter-vehicular coordination algorithms based on short-range direct V2V communications are needed. It is mandatory to check whether these novel solutions meet the STBA requirements. Future standards specifically aimed at SC communications in IVNs may emerge from such solutions.

### Naming and privacy

Additionally, we are exploring the (re)naming problem as it arises in IVNs. Source and destination names appear in messages exchanged among vehicles. Most often, names are IP addresses or MAC addresses (plate numbers shall not be used for privacy reasons). A vehicle which intends to communicate with some vehicle, denoted  $V$  here, must know which name  $name(V)$  to use in order to reach/designate  $V$ . Existing solutions are based on multicasting/broadcasting existential messages, whereby every vehicle publicizes its existence (name and geolocation), either upon request (replying to a Geocast) or spontaneously (periodic beaconing). These solutions have severe drawbacks. First, they contribute to overloading communication channels (leading to unacceptably high worst-case delays). Second, they amount to breaching privacy voluntarily. Why should vehicles reveal their existence and their time dependent geolocations, making tracing and spying much easier? Novel solutions are needed. They shall be such that:

- At any time, a vehicle can assign itself a name that is unique within a geographical zone centered on that vehicle (no third-party involved),
- No linkage may exist between a name and those identifiers (plate numbers, IP/MAC addresses, etc.) proper to a vehicle,
- Different (unique) names can be computed at different times by a vehicle (names can be short-lived or long-lived),
- $name(V)$  at UTC time  $t$  is revealed only to those vehicles sufficiently close to  $V$  at time  $t$ , notably those which may collide with  $V$ .

We have solved the (re)naming problem in string/cohort formations [48]. Ranks (unique integers in any given string/cohort) are privacy-preserving names, easily computed by every member of a string, in the presence of string membership changes (new vehicles join in, members leave). That problem is open when considering arbitrary clusters of vehicles/strings encompassing multiple lanes.

## 3.3. Probabilistic modeling for large transportation systems

**Participants:** Guy Fayolle, Jean-Marc Lasgouttes.

This activity concerns the modeling of random systems related to ITS, through the identification and development of solutions based on probabilistic methods and more specifically through the exploration of links between large random systems and statistical physics. Traffic modeling is a very fertile area of application for this approach, both for macroscopic (fleet management [46], traffic prediction) and for microscopic (movement of each vehicle, formation of traffic jams) analysis. When the size or volume of structures grows (leading to the so-called “thermodynamic limit”), we study the quantitative and qualitative (performance, speed, stability, phase transitions, complexity, etc.) features of the system.

In the recent years, several directions have been explored.

### 3.3.1. Traffic reconstruction

Large random systems are a natural part of macroscopic studies of traffic, where several models from statistical physics can be fruitfully employed. One example is fleet management, where one main issue is to find optimal ways of reallocating unused vehicles: it has been shown that Coulombian potentials might be an efficient tool to drive the flow of vehicles. Another case deals with the prediction of traffic conditions, when the data comes from probe vehicles instead of static sensors.

While the widely-used macroscopic traffic flow models are well adapted to highway traffic, where the distance between junction is long (see for example the work done by the NeCS team in Grenoble), our focus is on a more urban situation, where the graphs are much denser. The approach we are advocating here is model-less, and based on statistical inference rather than fundamental diagrams of road segments. Using the Ising model or even a Gaussian Random Markov Field, together with the very popular Belief Propagation (BP) algorithm, we have been able to show how real-time data can be used for traffic prediction and reconstruction (in the space-time domain).

This new use of BP algorithm raises some theoretical questions about the ways the make the belief propagation algorithm more efficient:

- find the best way to inject real-valued data in an Ising model with binary variables [50];
- build macroscopic variables that measure the overall state of the underlying graph, in order to improve the local propagation of information [47];
- make the underlying model as sparse as possible, in order to improve BP convergence and quality [49].

### 3.3.2. Exclusion processes for road traffic modeling

The focus here is on road traffic modeled as a granular flow, in order to analyze the features that can be explained by its random nature. This approach is complementary to macroscopic models of traffic flow (as done for example in the Opale team at Inria), which rely mainly on ODEs and PDEs to describe the traffic as a fluid.

One particular feature of road traffic that is of interest to us is the spontaneous formation of traffic jams. It is known that systems as simple as the Nagel-Schreckenberg model are able to describe traffic jams as an emergent phenomenon due to interaction between vehicles. However, even this simple model cannot be explicitly analyzed and therefore one has to resort to simulation.

One of the simplest solvable (but non trivial) probabilistic models for road traffic is the exclusion process. It lends itself to a number of extensions allowing to tackle some particular features of traffic flows: variable speed of particles, synchronized move of consecutive particles (platooning), use of geometries more complex than plain 1D (cross roads or even fully connected networks), formation and stability of vehicle clusters (vehicles that are close enough to establish an ad-hoc communication system), two-lane roads with overtaking.

The aspect that we have particularly studied is the possibility to let the speed of vehicle evolve with time. To this end, we consider models equivalent to a series of queues where the pair (service rate, number of customers) forms a random walk in the quarter plane  $\mathbb{Z}_+^2$ .

Having in mind a global project concerning the analysis of complex systems, we also focus on the interplay between discrete and continuous description: in some cases, this recurrent question can be addressed quite rigorously via probabilistic methods.

We have considered in [43] some classes of models dealing with the dynamics of discrete curves subjected to stochastic deformations. It turns out that the problems of interest can be set in terms of interacting exclusion processes, the ultimate goal being to derive hydrodynamic limits after proper scaling. A seemingly new method is proposed, which relies on the analysis of specific partial differential operators, involving variational calculus and functional integration. Starting from a detailed analysis of the Asymmetric Simple Exclusion Process (ASEP) system on the torus  $\mathbb{Z}/n\mathbb{Z}$ , the arguments a priori work in higher dimensions (ABC, multi-type exclusion processes, etc), leading to systems of coupled partial differential equations of Burgers' type.

### 3.3.3. Random walks in the quarter plane $\mathbb{Z}_+^2$

This field remains one of the important "violon d'Ingres" in our research activities in stochastic processes, both from theoretical and applied points of view. In particular, it is a building block for models of many communication and transportation systems.

One essential question concerns the computation of stationary measures (when they exist). As for the answer, it has been given by original methods formerly developed in the team (see books and related bibliography). For instance, in the case of small steps (jumps of size one in the interior of  $\mathbb{Z}_+^2$ ), the invariant measure  $\{\pi_{i,j}, i, j \geq 0\}$  does satisfy the fundamental functional equation (see [45]):

$$Q(x, y)\pi(x, y) = q(x, y)\pi(x) + \tilde{q}(x, y)\tilde{\pi}(y) + \pi_0(x, y). \quad (5)$$

where the unknown generating functions  $\pi(x, y), \pi(x), \tilde{\pi}(y), \pi_0(x, y)$  are sought to be analytic in the region  $\{(x, y) \in \mathbb{C}^2 : |x| < 1, |y| < 1\}$ , and continuous on their respective boundaries.

The given function  $Q(x, y) = \sum_{i,j} p_{i,j} x^i y^j - 1$ , where the sum runs over the possible jumps of the walk inside  $\mathbb{Z}_+^2$ , is often referred to as the *kernel*. Then it has been shown that equation (1) can be solved by reduction to a boundary-value problem of Riemann-Hilbert type. This method has been the source of numerous and fruitful developments. Some recent and ongoing works have been dealing with the following matters.

- *Group of the random walk.* In several studies, 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*, which is always even. In the case of small jumps, the algebraic curve  $R$  defined by  $\{Q(x, y) = 0\}$  is either of *genus* 0 (the sphere) or 1 (the torus). In [Fayolle-2011a], when the drift of the random walk is equal to 0 (and then so is the genus), an effective criterion gives the *order* of the group. More generally, it is also proved that whenever the genus is 0, this order is infinite, except precisely for the zero drift case, where finiteness is quite possible. When the *genus* is 1, the situation is more difficult. Recently [44], a criterion has been found in terms of a determinant of order 3 or 4, depending on the arity of the group.
- *Nature of the counting generating functions.* Enumeration of planar lattice walks is a classical topic in combinatorics. For a given set of allowed 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. A first basic and natural question arises: how many such paths exist? A second question concerns the nature of the associated counting generating functions (CGF): are they rational, algebraic, holonomic (or D-finite, i.e. solution of a linear differential equation with polynomial coefficients)?

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  $k$ . Then the corresponding CGF

$$F(x, y, z) = \sum_{i,j,k \geq 0} f(i, j, k) x^i y^j z^k \quad (6)$$

satisfies the functional equation

$$K(x, y)F(x, y, z) = c(x)F(x, 0, z) + \tilde{c}(y)F(0, y, z) + c_0(x, y), \quad (7)$$

where  $z$  is considered as a time-parameter. Clearly, equations (2) and (1) are of the same nature, and answers to the above questions have been given in [Fayolle-2010].

- *Some exact asymptotics in the counting of walks in  $\mathbb{Z}_+^2$ .* A new and uniform approach has been proposed about the following problem: *What is the asymptotic behavior, as their length goes to infinity, of the number of walks ending at some given point or domain (for instance one axis)?* The method in [Fayolle-2012] works for *both* finite or infinite groups, and for walks not necessarily restricted to excursions.

#### **3.3.4. Discrete-event simulation for urban mobility**

We have developed two simulation tools to study and evaluate the performance of different transportation modes covering an entire urban area.

- one for collective taxis, a public transportation system with a service quality provided will be comparable with that of conventional taxis (system operating with or without reservations, door-to-door services, well adapted itineraries following the current demand, controlling detours and waits, etc.), and with fares set at rates affordable by almost everyone, simply by utilizing previously wasted vehicle capacity;
- the second for a system of self-service cars that can reconfigure themselves into shuttles, therefore creating a multimodal public transportation system; this second simulator is intended to become a generic tool for multimodal transportation.

These two programs use a technique allowing to run simulations in batch mode and analyze the dynamics of the system afterward.



## AYIN Team

### 3. Research Program

#### 3.1. Geometric and shape modeling

One of the grand challenges of computer vision and image processing is the expression and use of prior geometric information via the construction of appropriate models. For very high resolution imagery, this problem becomes critically important, as the increasing resolution of the data results in the appearance of a great deal of complex geometric structure hitherto invisible. AYIN studies various approaches to the construction of models of geometry and shape.

##### *3.1.1. Stochastic geometry*

One of the most promising approaches to the inclusion of this type of information is stochastic geometry, which is an important research direction in the AYIN team. Instead of defining probabilities for different types of image, probabilities are defined for configurations of an indeterminate number of interacting, parameterized objects located in the image. Such probability distributions are called ‘marked point processes’. New models are being developed both for remote sensing applications, and for skin care problems, such as wrinkle and acne detection.

##### *3.1.2. Contours, phase fields, and MRFs with long-range interactions*

An alternative approach to shape modeling starts with generic ‘regions’ in the image, and adds constraints in order to model specific shapes and objects. AYIN investigates contour, phase field, and binary field representations of regions, incorporating shape information via highly-structured long-range interactions that constrain the set of high-probability regions to those with specific geometric properties. This class of models can represent infinite-dimensional families of shapes and families with unbounded topology, as well as families consisting of an arbitrary number of object instances, at no extra computational cost. Key sub-problems include the development of models of more complex shapes and shape configurations; the development of models in more than two spatial dimensions; and understanding the equivalences between models in different representations and approaches.

##### *3.1.3. Shapes in time*

AYIN is concerned with spectral and spatio-temporal structures. To deal with the latter, the above scene modeling approaches are extended into the time dimension, either by modeling time dependence directly, or, in the field-based approaches, by modeling spacetime structures, or, in the stochastic geometry approach, by including the time  $t$  in the mark. An example is a spatio-temporal graph-cut-based method that introduces directed infinite links connecting pixels in successive image frames in order to impose constraints on shape change.

#### 3.2. Image modeling

The key issue that arises in modeling the high-resolution image data generated in AYIN’s applications, is how to include large-scale spatial, temporal, and spectral dependencies. AYIN investigates approaches to the construction of image models including such dependencies. A central question in the use of such models is how to deal with the large data volumes arising both from the large size of the images involved, and the existence of large image collections. Fortunately, high dimensionality typically implies data redundancy, and so AYIN investigates methods for reducing the dimensionality of the data and describing the spatial, temporal, and spectral dependencies in ways that allow efficient data processing.

### **3.2.1. Markov random fields with long-range and higher-order interactions**

One way to achieve large-scale dependencies is via explicit long-range interactions. MRFs with long-range interactions are also used in AYIN to model geometric spatial and temporal structure, and the techniques and algorithms developed there will also be applied to image modeling. In modeling image structures, however, other important properties, such as control of the relative phase of Fourier components, and spontaneous symmetry breaking, may also be required. These properties can only be achieved by higher-order interactions. These require specific techniques and algorithms, which are developed in parallel with the models.

### **3.2.2. Hierarchical models**

Another way to achieve long-range dependencies is via shorter range interactions in a hierarchical structure. AYIN works on the development of models defined as a set of hierarchical image partitions represented by a binary forest structure. Key sub-problems include the development of multi-feature models of image regions as an ensemble of spectral, texture, geometrical, and classification features, where we search to optimize the ratio between discrimination capacity of the feature space and dimensionality of this space; and the development of similarity criteria between image regions, which would compute distances between regions in the designed feature space and would be data-driven and scale-independent. One way to proceed in the latter case consists in developing a composite kernel method, which would seek to project multi-feature data into a new space, where regions from different thematic categories become linearly or almost linearly separable. This involves developing kernel functions as a combination of basis kernels, and estimating kernel-based support vector machine parameters.

## **3.3. Algorithms**

Computational techniques are necessary in order to extract the information of interest from the models. In addition, most models contain ‘nuisance parameters’, including the structure of the models themselves, that must be dealt with in some way. AYIN is interested in adapting and developing methods for solving these problems in cases where existing methods are inadequate.

### **3.3.1. Nuisance parameters and parameter estimation**

In order to render the models operational, it is crucial to find some way to deal with nuisance parameters. In a Bayesian framework, the parameters must be integrated or marginalized out. Unfortunately, this is usually very difficult. Fortunately, Laplace’s method often provides a good approximation, in many cases being equivalent to classical maximum likelihood parameter estimation. Even these problems are not easy to solve, however, when dealing with complex, structured models. This is particularly true when it is necessary to estimate simultaneously both the information of interest and the parameters. AYIN is developing a number of different methods for dealing with nuisance parameters, corresponding to the diversity of modeling approaches.

### **3.3.2. Information extraction**

Extracting the information of interest from any model involves making estimates based on various criteria, for example MAP, MPM, or MMSE. Computing these estimates often requires the solution of hard optimization problems. The complexity of many of the models to be developed within AYIN means that off-the-shelf algorithms and current techniques are often not capable of solving these problems. AYIN develops a diversity of algorithmic approaches adapted to the particular models developed.

## LINKMEDIA Project-Team

### 3. Research Program

#### 3.1. Scientific background

LINKMEDIA is a multidisciplinary research team, with multimedia data as the main object of study. We are guided by the data and their specificity—semantically interpretable, heterogeneous and multimodal, available in large amounts, unstructured and disconnected—, as well as by the related problems and applications.

With multimedia data at the center, orienting our choices of methods and algorithms and serving as a basis for experimental validation, the team is directly contributing to the following scientific fields:

- multimedia: content-based analysis; multimodal processing and fusion; multimedia applications;
- computer vision: compact description of images; object and event detection;
- natural language processing: topic segmentation; information extraction;
- information retrieval: high-dimensional indexing; approximate k-nn search; efficient set comparison.

LINKMEDIA also takes advantage of advances in the following fields, adapting recent developments to the multimedia area:

- signal processing: image processing; compression;
- machine learning: deep architectures; structured learning; adversarial learning;
- security: data encryption; differential privacy;
- data mining: time series mining and alignment; pattern discovery; knowledge extraction.

#### 3.2. Workplan

Research activities in LINKMEDIA are organized along three major lines of research which build upon the scientific domains already mentioned.

##### 3.2.1. *Unsupervised motif discovery*

As an alternative to supervised learning techniques, unsupervised approaches have emerged recently in multimedia with the goal of discovering directly patterns and events of interest from the data, in a totally unsupervised manner. In the absence of prior knowledge on what we are interested in, meaningfulness can be judged based on one of three main criteria: unexpectedness, saliency and recurrence. This last case posits that repeating patterns, known as motifs, are potentially meaningful, leading to recent work on the unsupervised discovery of motifs in multimedia data [54], [52], [53].

LINKMEDIA seeks to *develop unsupervised motif discovery approaches which are both accurate and scalable*. In particular, we consider the discovery of repeating objects in image collections and the discovery of repeated sequences in video and audio streams. Research activities are organized along the following lines:

- developing the scientific basis for scalable motif discovery: sparse histogram representations; efficient co-occurrence counting; geometry and time aware indexing schemes;
- designing and evaluating accurate and scalable motif discovery algorithms applied to a variety of multimedia content: exploiting efficient geometry or time aware matching functions; fast approximate dynamic time warping; symbolic representations of multimedia data, in conjunction with existing symbolic data mining approaches;
- developing methodology for the interpretation, exploitation and evaluation of motif discovery algorithms in various use-cases: image classification; video stream monitoring; transcript-free natural language processing (NLP) for spoken document.

### 3.2.2. *Description and structuring*

Content-based analysis has received a lot of attention from the early days of multimedia, with an extensive use of supervised machine learning for all modalities [56], [48]. Progress in large scale entity and event recognition in multimedia content has made available general purpose approaches able to learn from very large data sets and performing fairly decently in a large number of cases. Current solutions are however limited to simple, homogeneous, information and can hardly handle structured information such as hierarchical descriptions, tree-structured or nested concepts.

LINKMEDIA aims at *expanding techniques for multimedia content modeling, event detection and structure analysis*. The main transverse research lines that LINKMEDIA will develop are as follows:

- context-aware content description targeting (homogeneous) collections of multimedia data: latent variable discovery; deep feature learning; motif discovery;
- secure description to enable privacy and security aware multimedia content processing: leveraging encryption and obfuscation; exploring adversarial machine learning in a multimedia context; privacy-oriented image processing;
- multilevel modeling with a focus on probabilistic modeling of structured multimodal data: multiple kernels; structured machine learning; conditional random fields.

### 3.2.3. *Linking and collection data model*

Creating explicit links between media content items has been considered on different occasions, with the goal of seeking and discovering information by browsing, as opposed to information retrieval via ranked lists of relevant documents. Content-based link creation has been initially addressed in the hypertext community for well-structured texts [47] and was recently extended to multimedia content [57], [51], [50]. The problem of organizing collections with links remains mainly unsolved for large heterogeneous collections of unstructured documents, with many issues deserving attention: linking at a fine semantic grain; selecting relevant links; characterizing links; evaluating links; etc.

LINKMEDIA targets pioneering research on media linking by *developing scientific ground, methodology and technology for content-based media linking* directed to applications exploiting rich linked content such as navigation or recommendation. Contributions are concentrated along the following lines:

- algorithmic of linked media for content-based link authoring in multimedia collections: time-aware graph construction; multimodal hypergraphs; large scale k-nn graphs;
- link interpretation and characterization to provide links semantics for interpretability: text alignment; entity linking; intention vs. extension;
- linked media usage and evaluation: information retrieval; summarization; data models for navigation; link prediction.

## MAGRIT Project-Team

### 3. Research Program

#### 3.1. Matching and 3D tracking

One of the most basic problems currently limiting AR applications is the registration problem. The objects in the real and virtual worlds must be properly aligned with respect to each other, or the illusion that the two worlds coexist will be compromised.

As a large number of potential AR applications are interactive, real time pose computation is required. Although the registration problem has received a lot of attention in the computer vision community, the problem of real-time registration is still far from being a solved problem, especially for unstructured environments. Ideally, an AR system should work in all environments, without the need to prepare the scene ahead of time, independently of the variations in experimental conditions (lighting, weather condition,...) which may exist between the application and the time the model of the scene was acquired.

For several years, the MAGRIT project has been aiming at developing on-line and marker-less methods for camera pose computation. The main difficulty with on-line tracking is to ensure robustness of the process over time. For off-line processes, robustness is achieved by using spatial and temporal coherence of the considered sequence through move-matching techniques. To get robustness for open-loop systems, we have investigated various methods, ranging from statistical methods to the use of hybrid camera/sensor systems. Many of these methods are dedicated to piecewise-planar scenes and combine the advantage of move-matching methods and model-based methods. In order to reduce statistical fluctuations in viewpoint computation, which lead to unpleasant jittering or sliding effects, we have also developed model selection techniques which allow us to noticeably improve the visual impression and to reduce drift over time. Another line of research which has been considered in the team to improve the reliability and the robustness of pose algorithms is to combine the camera with another form of sensor in order to compensate for the shortcomings of each technology.

The success of pose computation over time largely depends on the quality of the matching at the initialization stage. Indeed, the current image may be very different from the appearances described in the model both on the geometrical and the photometric sides. Research is thus conducted in the team on the use of probabilistic methods to establish robust correspondences of features. The use of *a contrario* methods has been investigated to achieve this aim [8]. We especially addressed the complex case of matching in scenes with repeated patterns which are common in urban scenes. We are also investigating the problem of matching images taken from very different viewpoints which is central for the re-localization issue in AR. Within the context of a scene model acquired with structure from motion techniques, we are currently investigating the use of viewpoint simulation in order to allow successful pose computation even if the considered image is far from the positions used to build the model [4].

Recently, the issue of tracking deformable objects has gained importance in the team. This topic is mainly addressed in the context of medical applications through the design of bio-mechanical models guided by visual features [1]. We have successfully investigated the use of such models in laparoscopy, with a vascularized model of the liver and with a hyper-elastic model for tongue tracking in ultrasound images. However, these results have been obtained so far in relatively controlled environments, with non pathological cases. When clinical routine applications are to be considered, many parameters and considerations need to be taken into account. Among the problems that need to be addressed are more realistic model representations, the specification of the range of physical parameters and the need to enforce the robustness of the tracking with respect to outliers, which are common in the interventional context.

## 3.2. Image-based Modeling

Modeling the scene is a fundamental issue in AR for many reasons. First, pose computation algorithms often use a model of the scene or at least some 3D knowledge on the scene. Second, effective AR systems require a model of the scene to support interactions between the virtual and the real objects such as occlusions, lighting reflections, contacts...in real-time. Unlike pose computation which has to be computed in a sequential way, scene modeling can be considered as an off-line or an on-line problem depending on the requirements of the targeted application. Interactive in-situ modeling techniques have thus been developed with the aim to enable the user to define what is relevant at the time the model is being built during the application. On the other hand, we also proposed off-line multimodal techniques, mainly dedicated to AR medical applications, with the aim to obtain realistic and possibly dynamic models of organs suitable for real-time simulation.

### **In-situ modeling**

In-situ modeling allows a user to directly build a 3D model of his/her surrounding environment and verify the geometry against the physical world in real-time. This is of particular interest when using AR in unprepared environments or building scenes that either have an ephemeral existence (e.g., a film set) or cannot be accessed frequently (e.g., a nuclear power plant). We have especially investigated two systems, one based on the image content only and the other based on multiple data coming from different sensors (camera, inertial measurement unit, laser rangefinder). Both systems use the camera-mouse principle [6] (i.e., interactions are performed by aiming at the scene through a video camera) and both systems have been designed to acquire polygonal textured models, which are particularly useful for camera tracking and object insertion in AR.

### **Multimodal modeling for real-time simulation**

With respect to classical AR applications, AR in medical context differs in the nature and the size of the data which are available: a large amount of multimodal data is acquired on the patient or possibly on the operating room through sensing technologies or various image acquisitions [3]. The challenge is to analyze these data, to extract interesting features, to fuse and to visualize this information in a proper way. Within the MAGRIT team, we address several key problems related to medical augmented environments. Being able to acquire multimodal data which are temporally synchronized and spatially registered is the first difficulty we face when considering medical AR. Another key requirement of AR medical systems is the availability of 3D (+t) models of the organ/patient built from images, to be overlaid onto the users' view of the environment.

Methods for multimodal modeling are strongly dependent on the image modalities and the organ specificities. We thus only address a restricted number of medical applications –interventional neuro-radiology, laparoscopic surgery– for which we have a strong expertise and close relationships with motivated clinicians. In these applications, our aim is to produce realistic models and then realistic simulations of the patient to be used for surgeon's training or patient's re-education/learning.

One of our main applications is about neuroradiology. For the last 20 years, we have been working in close collaboration with the neuroradiology laboratory (CHU-University Hospital of Nancy) and GE Healthcare. As several imaging modalities are now available in an intraoperative context (2D and 3D angiography, MRI, ...), our aim is to develop a multi-modality framework to help therapeutic decision and treatment.

We have mainly been interested in the effective use of a multimodality framework in the treatment of arteriovenous malformations (AVM) and aneurysms in the context of interventional neuroradiology. The goal of interventional gestures is to guide endoscopic tools towards the pathology with the aim to perform embolization of the AVM or to fill the aneurysmal cavity by placing coils. We have proposed and developed multimodality and augmented reality tools which make various image modalities (2D and 3D angiography, fluoroscopic images, MRI, ...) cooperate in order to help physicians in clinical routine. One of the successes of this collaboration is the implementation of the concept of *augmented fluoroscopy*, which helps the surgeon to guide endoscopic tools towards the pathology. Lately, in cooperation with the team MIMESIS, we have proposed new methods for implicit modeling of the vasculature with the aim of obtaining near real-time simulation of the coil deployment in the aneurysm [2]. These works open the way towards near real-time patient-based simulations of interventional gestures both for training and for planning.

### **3.3. Parameter estimation**

Many problems in computer vision or image analysis can be formulated in terms of parameter estimation from image-based measurements. This is the case of many problems addressed in the team such as pose computation or image-guided estimation of 3D deformable models. Often traditional robust techniques which take into account the covariance on the measurements are sufficient to achieve reliable parameter estimation. However, depending on their number, their spatial distribution and the uncertainty on these measurements, some problems are very sensitive to noise and there is a considerable interest in considering how parameter estimation could be improved if additional information on the noise were available. Another common problem in our field of research is the need to estimate constitutive parameters of the models, such as (bio)-mechanical parameters for instance. Direct measurement methods are destructive and elaborating image based methods is thus highly desirable. Besides designing appropriate estimation algorithms, a fundamental question is to understand what group of parameters under study can be reliably estimated from a given experimental setup.

This line of research is relatively new in the team. One of the challenges is to improve image-based parameter estimation techniques considering sensor noise and specific image formation models. In a collaboration with the Pascal Institute (Clermont Ferrand), metrological performance enhancement for experimental solid mechanics has been addressed through the development of dedicated signal processing methods [7]. In the medical field, specific methods based on an adaptive evolutionary optimization strategy have been designed for estimating respiratory parameters [9]. In the context of designing realistic simulators for neuroradiology, we are now considering how parameters involved in the simulation could be adapted to fit real images.

## MORPHEO Project-Team

### 3. Research Program

#### 3.1. Shape Acquisition

Multiple camera setups allow to acquire shapes, i.e. geometry, as well as their appearances, i.e. photometry, with a reasonable level of precision. However fundamental limitations still exist, in particular today's state-of-the-art approaches do not fully exploit the redundancy of information over temporal sequences of visual observations. Despite an increasing interest of the computer vision communities in the past years, the problem is still far from solved other than in specific situations with restrictive assumptions and configurations. Our goal in this research axis is to fully leverage temporal aspects of the acquisition process and to open the acquisition process to different modalities, in particular Xrays.

#### 3.2. Generative / discriminative inference

Acquisition of 4D Models can often be conveniently formulated as an estimation or learning problem. Various generative models can be proposed for the problems of shape and appearance modeling over time sequences, and motion segmentation. The idea of these generative models is to predict the noisy measurements (e.g. pixel values, measured 3D points or speed quantities) from a set of parameters describing the unobserved scene state (e.g. shape and appearance), which in turn can be inverted with various inference algorithms. The advantages of this type of modeling are numerous to deal with noisy measurements, explicitly model dependencies between model parameters, hidden variables and observed quantities, and relevant priors over parameters; sensor models for different modalities can also easily be seamlessly integrated and jointly used, which remains central to our goals. A limitation of such algorithms is that classical algorithms to solve them rely on local iterative convergence schemes subject to local minima, or global restart schemes which avert this problem but with a significant computational penalty. This is why we also consider discriminative and deep learning approaches, which allow to formulate the parameter estimation as a direct regression from input quantities or pixel values, whose parameters are learned given a training set. This has the advantage of directly computing a solution from inputs, with robustness and speed benefits, as a standalone estimation algorithm or to initialize local convergence schemes based on generative modeling. A number of the approaches we propose thus leverage the advantages of both generative and such discriminative approaches.

#### 3.3. Shape Analysis

Shape analysis has received much attention from the scientific community and recovering the intrinsic nature of shapes is currently an active research domain. Of particular interest is the study of human and animal shapes and their associated articulated underlying structures, i.e. skeletons, since applications are numerous, either in the entertainment industry or for medical applications, among others. Our main goals in this research axis are : the understanding of a shape's global structure, and a pose-independent classification of shapes.

#### 3.4. Shape Tracking

Recovering the temporal evolution of a deformable surface is a fundamental task in computer vision, with a large variety of applications ranging from the motion capture of articulated shapes, such as human bodies, to the deformation of complex surfaces such as clothes. Methods that solve for this problem usually infer surface evolutions from motion or geometric cues. This information can be provided by motion capture systems or one of the numerous available static 3D acquisition modalities. In this inference, methods are faced with the challenging estimation of the time-consistent deformation of a surface from cues that can be sparse and noisy. Such an estimation is an ill posed problem that requires prior knowledge on the deformation to be introduced in order to limit the range of possible solutions. Our goal is to devise robust and accurate solutions based on new deformation models that fully exploit the geometric and photometric information available.



### **3.5. Dynamic Motion Modeling**

Multiple views systems can significantly change the paradigm of motion capture. Traditional motion capture systems provide 3D trajectories of a sparse set of markers fixed on the subject. These trajectories can be transformed into motion parameters on articulated limbs with the help of prior models of the skeletal structure. However, such skeletal models are mainly robotical abstractions that do not describe the true morphology and anatomical motions of humans and animals. On the other hand, 4D models (temporally consistent mesh sequences) provide dense motion information on body's shape while requiring less prior assumption. They represent therefore a new rich source of information on human and animal shape movements. The analysis of such data has already received some attention but most existing works model motion through static poses and do not consider yet dynamic information. Such information (e.g. trajectories and speed) is anyway required to analyse walking or running sequences. We will investigate this research direction with the aim to propose and study new dynamic models.

### **3.6. Shape Animation**

3D animation is a crucial part of digital media production with numerous applications, in particular in the game and motion picture industry. Recent evolutions in computer animation consider real videos for both the creation and the animation of characters. The advantage of this strategy is twofold: it reduces the creation cost and increases realism by considering only real data. Furthermore, it allows to create new motions, for real characters, by recombining recorded elementary movements. In addition to enable new media contents to be produced, it also allows to automatically extend moving shape datasets with fully controllable new motions. This ability appears to be of great importance with the recent advent of deep learning techniques and the associated need for large learning datasets. In this research direction, we will investigate how to create new dynamic scenes using recorded events.

## PERCEPTION Project-Team

### 3. Research Program

#### 3.1. Audio-Visual Scene Analysis

From 2006 to 2009, R. Horaud was the scientific coordinator of the collaborative European project POP (Perception on Purpose), an interdisciplinary effort to understand visual and auditory perception at the crossroads of several disciplines (computational and biological vision, computational auditory analysis, robotics, and psychophysics). This allowed the PERCEPTION team to launch an interdisciplinary research agenda that has been very active for the last five years. There are very few teams in the world that gather scientific competences spanning computer vision, audio signal processing, machine learning and human-robot interaction. The fusion of several sensorial modalities resides at the heart of the most recent biological theories of perception. Nevertheless, multi-sensor processing is still poorly understood from a computational point of view. In particular and so far, audio-visual fusion has been investigated in the framework of speech processing using close-distance cameras and microphones. The vast majority of these approaches attempt to model the temporal correlation between the auditory signals and the dynamics of lip and facial movements. Our original contribution has been to consider that audio-visual localization and recognition are equally important. We have proposed to take into account the fact that the audio-visual objects of interest live in a three-dimensional physical space and hence we contributed to the emergence of *audio-visual scene analysis* as a scientific topic in its own right. We proposed several novel statistical approaches based on supervised and unsupervised mixture models. The *conjugate mixture model* (CMM) is an unsupervised probabilistic model that allows to cluster observations from different modalities (e.g., vision and audio) living in different mathematical spaces [18], [2]. We thoroughly investigated CMM, provided practical resolution algorithms and studied their convergence properties. We developed several methods for sound localization using two or more microphones [1]. The *Gaussian locally-linear model* (GLLiM) is a partially supervised mixture model that allows to map high-dimensional observations (audio, visual, or concatenations of audio-visual vectors) onto low-dimensional manifolds with a partially known structure [6]. This model is particularly well suited for perception because it encodes both observable and unobservable phenomena. A variant of this model, namely *probabilistic piecewise affine mapping* has also been proposed and successfully applied to the problem of sound-source localization and separation [5]. The European projects HUMAVIPS (2010-2013) coordinated by R. Horaud and EARS (2014-2017), applied audio-visual scene analysis to human-robot interaction.

#### 3.2. Stereoscopic Vision

Stereoscopy is one of the most studied topics in biological and computer vision. Nevertheless, classical approaches of addressing this problem fail to integrate eye/camera vergence. From a geometric point of view, the integration of vergence is difficult because one has to re-estimate the epipolar geometry at every new eye/camera rotation. From an algorithmic point of view, it is not clear how to combine depth maps obtained with different eyes/cameras relative orientations. Therefore, we addressed the more general problem of binocular vision that combines the low-level eye/camera geometry, sensor rotations, and practical algorithms based on global optimization [12], [20]. We studied the link between mathematical and computational approaches to stereo (global optimization and Markov random fields) and the brain plausibility of some of these approaches: indeed, we proposed an original mathematical model for the complex cells in visual-cortex areas V1 and V2 that is based on steering Gaussian filters and that admits simple solutions [13]. This addresses the fundamental issue of how local image structure is represented in the brain/computer and how this structure is used for estimating a dense disparity field. Therefore, the main originality of our work is to address both computational and biological issues within a unifying model of binocular vision. Another equally important problem that still remains to be solved is how to integrate binocular depth maps over time. Recently, we have addressed this problem and proposed a semi-global optimization framework that starts with sparse yet reliable matches and proceeds with propagating them over both space and time. The concept of seed-match propagation has then been extended to TOF-stereo fusion [8].

### 3.3. Audio Signal Processing

Audio-visual fusion algorithms necessitate that the two modalities are represented in the same mathematical space. Binaural audition allows to extract sound-source localization (SSL) information from the acoustic signals recorded with two microphones. We have developed several methods, that perform sound localization in the temporal and the spectral domains. If a direct path is assumed, one can exploit the *time difference of arrival* (TDOA) between two microphones to recover the position of the sound source with respect to the position of the two microphones. The solution is not unique in this case, the sound source lies onto a 2D manifold. However, if one further assumes that the sound source lies in a horizontal plane, it is then possible to extract the azimuth. We used this approach to predict possible sound locations in order to estimate the direction of a speaker [2]. We also developed a geometric formulation and we showed that with four non-coplanar microphones the azimuth and elevation of a single source can be estimated without ambiguity [1]. We also investigated SSL in the spectral domain. This exploits the filtering effects of the head related transfer function (HRTF): there is a different HRTF for the left and right microphones. The interaural spectral features, namely the ILD (interaural level difference) and IPD (interaural phase difference) can be extracted from the short-time Fourier transforms of the two signals. The sound direction is encoded in these interaural features but it is not clear how to make SSL explicit in this case. We proposed a supervised learning formulation that estimates a mapping from interaural spectral features (ILD and IPD) to source directions using two different setups: audio-motor learning [5] and audio-visual learning [7].

### 3.4. Visual Reconstruction With Multiple Color and Depth Cameras

For the last decade, one of the most active topics in computer vision has been the visual reconstruction of objects, people, and complex scenes using a multiple-camera setup. The PERCEPTION team has pioneered this field and by 2006 several team members published seminal papers in the field. Recent work has concentrated onto the robustness of the 3D reconstructed data using probabilistic outlier rejection techniques combined with algebraic geometry principles and linear algebra solvers [23]. Subsequently, we proposed to combine 3D representations of shape (meshes) with photometric data [21]. The originality of this work was to represent photometric information as a scalar function over a discrete Riemannian manifold, thus *generalizing image analysis to mesh and graph analysis*. Manifold equivalents of local-structure detectors and descriptors were developed [22]. The outcome of this pioneering work has been twofold: the formulation of a new research topic now addressed by several teams in the world, and allowed us to start a three year collaboration with Samsung Electronics. We developed the novel concept of *mixed camera systems* combining high-resolution color cameras with low-resolution depth cameras [14], [10],[9]. Together with our start-up company 4D Views Solutions and with Samsung, we developed the first practical depth-color multiple-camera multiple-PC system and the first algorithms to reconstruct high-quality 3D content [8].

### 3.5. Registration, Tracking and Recognition of People and Actions

The analysis of articulated shapes has challenged standard computer vision algorithms for a long time. There are two difficulties associated with this problem, namely how to represent articulated shapes and how to devise robust registration and tracking methods. We addressed both these difficulties and we proposed a novel kinematic representation that integrates concepts from robotics and from the geometry of vision. In 2008 we proposed a method that parameterizes the occluding contours of a shape with its intrinsic kinematic parameters, such that there is a direct mapping between observed image features and joint parameters [19]. This deterministic model has been motivated by the use of 3D data gathered with multiple cameras. However, this method was not robust to various data flaws and could not achieve state-of-the-art results on standard dataset. Subsequently, we addressed the problem using probabilistic generative models. We formulated the problem of articulated-pose estimation as a maximum-likelihood with missing data and we devised several tractable algorithms [17], [16]. We proposed several expectation-maximization procedures applied to various articulated shapes: human bodies, hands, etc. In parallel, we proposed to segment and register articulated shapes represented with graphs by embedding these graphs using the spectral properties of graph Laplacians [4]. This turned out to be a very original approach that has been followed by many other researchers in computer vision and computer graphics.

## SIROCCO Project-Team

### 3. Research Program

#### 3.1. Introduction

The research activities on analysis, compression and communication of visual data mostly rely on tools and formalisms from the areas of statistical image modelling, of signal processing, of coding and information theory. However, the objective of better exploiting the Human Visual System (HVS) properties in the above goals also pertains to the areas of perceptual modelling and cognitive science. Some of the proposed research axes are also based on scientific foundations of computer vision (e.g. multi-view modelling and coding). We have limited this section to some tools which are central to the proposed research axes, but the design of complete compression and communication solutions obviously rely on a large number of other results in the areas of motion analysis, transform design, entropy code design, etc which cannot be all described here.

#### 3.2. Parameter Estimation and Inference

Bayesian estimation, Expectation-Maximization, stochastic modelling

Parameter estimation is at the core of the processing tools studied and developed in the team. Applications range from the prediction of missing data or future data, to extracting some information about the data in order to perform efficient compression. More precisely, the data are assumed to be generated by a given stochastic data model, which is partially known. The set of possible models translates the a priori knowledge we have on the data and the best model has to be selected in this set. When the set of models or equivalently the set of probability laws is indexed by a parameter (scalar or vectorial), the model is said parametric and the model selection resorts to estimating the parameter. Estimation algorithms are therefore widely used at the encoder to analyze the data. In order to achieve high compression rates, the parameters are usually not sent and the decoder has to jointly select the model (i.e. estimate the model parameters) and extract the information of interest.

#### 3.3. Data Dimensionality Reduction

Manifolds, locally linear embedding, non-negative matrix factorization, principal component analysis

A fundamental problem in many data processing tasks (compression, classification, indexing) is to find a suitable representation of the data. It often aims at reducing the dimensionality of the input data so that tractable processing methods can then be applied. Well-known methods for data dimensionality reduction include principal component analysis (PCA) and independent component analysis (ICA). The methodologies which will be central to several proposed research problems will instead be based on sparse representations, on locally linear embedding (LLE) and on the “non negative matrix factorization” (NMF) framework.

The objective of *sparse representations* is to find a sparse approximation of a given input data. In theory, given  $A \in \mathbb{R}^{m \times n}$ ,  $m < n$ , and  $\mathbf{b} \in \mathbb{R}^m$  with  $m \ll n$  and  $A$  is of full rank, one seeks the solution of  $\min\{\|\mathbf{x}\|_0 : A\mathbf{x} = \mathbf{b}\}$ , where  $\|\mathbf{x}\|_0$  denotes the  $L_0$  norm of  $x$ , i.e. the number of non-zero components in  $x$ . There exist many solutions  $x$  to  $Ax = b$ . The problem is to find the sparsest, the one for which  $x$  has the fewest non zero components. In practice, one actually seeks an approximate and thus even sparser solution which satisfies  $\min\{\|\mathbf{x}\|_0 : \|A\mathbf{x} - \mathbf{b}\|_p \leq \rho\}$ , for some  $\rho \geq 0$ , characterizing an admissible reconstruction error. The norm  $p$  is usually 2, but could be 1 or  $\infty$  as well. Except for the exhaustive combinatorial approach, there is no known method to find the exact solution under general conditions on the dictionary  $A$ . Searching for this sparsest representation is hence unfeasible and both problems are computationally intractable. Pursuit algorithms have been introduced as heuristic methods which aim at finding approximate solutions to the above problem with tractable complexity.

*Non negative matrix factorization* (NMF) is a non-negative approximate data representation<sup>0</sup>. NMF aims at finding an approximate factorization of a non-negative input data matrix  $V$  into non-negative matrices  $W$  and  $H$ , where the columns of  $W$  can be seen as *basis vectors* and those of  $H$  as coefficients of the linear approximation of the input data. Unlike other linear representations like PCA and ICA, the non-negativity constraint makes the representation purely additive. Classical data representation methods like PCA or Vector Quantization (VQ) can be placed in an NMF framework, the differences arising from different constraints being placed on the  $W$  and  $H$  matrices. In VQ, each column of  $H$  is constrained to be unitary with only one non-zero coefficient which is equal to 1. In PCA, the columns of  $W$  are constrained to be orthonormal and the rows of  $H$  to be orthogonal to each other. These methods of data-dependent dimensionality reduction will be at the core of our visual data analysis and compression activities.

### 3.4. Perceptual Modelling

Saliency, visual attention, cognition

The human visual system (HVS) is not able to process all visual information of our visual field at once. To cope with this problem, our visual system must filter out irrelevant information and reduce redundant information. This feature of our visual system is driven by a selective sensing and analysis process. For instance, it is well known that the greatest visual acuity is provided by the fovea (center of the retina). Beyond this area, the acuity drops down with the eccentricity. Another example concerns the light that impinges on our retina. Only the visible light spectrum lying between 380 nm (violet) and 760 nm (red) is processed. To conclude on the selective sensing, it is important to mention that our sensitivity depends on a number of factors such as the spatial frequency, the orientation or the depth. These properties are modeled by a sensitivity function such as the Contrast Sensitivity Function (CSF).

Our capacity of analysis is also related to our visual attention. Visual attention which is closely linked to eye movement (note that this attention is called *overt* while the *covert* attention does not involve eye movement) allows us to focus our biological resources on a particular area. It can be controlled by both top-down (i.e. goal-directed, intention) and bottom-up (stimulus-driven, data-dependent) sources of information<sup>0</sup>. This detection is also influenced by prior knowledge about the environment of the scene<sup>0</sup>. Implicit assumptions related to prior knowledge or beliefs play an important role in our perception (see the example concerning the assumption that light comes from above-left). Our perception results from the combination of prior beliefs with data we gather from the environment. A Bayesian framework is an elegant solution to model these interactions<sup>0</sup>. We define a vector  $\vec{v}_l$  of local measurements (contrast of color, orientation, etc.) and vector  $\vec{v}_c$  of global and contextual features (global features, prior locations, type of the scene, etc.). The salient locations  $S$  for a spatial position  $\vec{x}$  are then given by:

$$S(\vec{x}) = \frac{1}{p(\vec{v}_l | \vec{v}_c)} \times p(s, \vec{x} | \vec{v}_c) \quad (8)$$

The first term represents the bottom-up salience. It is based on a kind of contrast detection, following the assumption that rare image features are more salient than frequent ones. Most of existing computational models of visual attention rely on this term. However, different approaches exist to extract the local visual features as well as the global ones. The second term is the contextual priors. For instance, given a scene, it indicates which parts of the scene are likely the most salient.

<sup>0</sup>D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization", *Nature* 401, 6755, (Oct. 1999), pp. 788-791.

<sup>0</sup>L. Itti and C. Koch, "Computational Modelling of Visual Attention", *Nature Reviews Neuroscience*, Vol. 2, No. 3, pp. 194-203, 2001.

<sup>0</sup>J. Henderson, "Regarding scenes", *Directions in Psychological Science*, vol. 16, pp. 219-222, 2007.

<sup>0</sup>L. Zhang, M. Tong, T. Marks, H. Shan, H. and G.W. Cottrell, "SUN: a Bayesian framework for saliency using natural statistics", *Journal of Vision*, vol. 8, pp. 1-20, 2008.

### 3.5. Coding theory

OPTA limit (Optimum Performance Theoretically Attainable), Rate allocation, Rate-Distortion optimization, lossy coding, joint source-channel coding multiple description coding, channel modelization, oversampled frame expansions, error correcting codes.

Source coding and channel coding theory<sup>0</sup> is central to our compression and communication activities, in particular to the design of entropy codes and of error correcting codes. Another field in coding theory which has emerged in the context of sensor networks is Distributed Source Coding (DSC). It refers to the compression of correlated signals captured by different sensors which do not communicate between themselves. All the signals captured are compressed independently and transmitted to a central base station which has the capability to decode them jointly. DSC finds its foundation in the seminal Slepian-Wolf<sup>0</sup> (SW) and Wyner-Ziv<sup>0</sup> (WZ) theorems. Let us consider two binary correlated sources  $X$  and  $Y$ . If the two coders communicate, it is well known from Shannon's theory that the minimum lossless rate for  $X$  and  $Y$  is given by the joint entropy  $H(X, Y)$ . Slepian and Wolf have established in 1973 that this lossless compression rate bound can be approached with a vanishing error probability for long sequences, even if the two sources are coded separately, provided that they are decoded jointly and that their correlation is known to both the encoder and the decoder.

In 1976, Wyner and Ziv considered the problem of coding of two correlated sources  $X$  and  $Y$ , with respect to a fidelity criterion. They have established the rate-distortion function  $R_{*X|Y}(D)$  for the case where the side information  $Y$  is perfectly known to the decoder only. For a given target distortion  $D$ ,  $R_{*X|Y}(D)$  in general verifies  $R_{X|Y}(D) \leq R_{*X|Y}(D) \leq R_X(D)$ , where  $R_{X|Y}(D)$  is the rate required to encode  $X$  if  $Y$  is available to both the encoder and the decoder, and  $R_X$  is the minimal rate for encoding  $X$  without SI. These results give achievable rate bounds, however the design of codes and practical solutions for compression and communication applications remain a widely open issue.

---

<sup>0</sup>T. M. Cover and J. A. Thomas, Elements of Information Theory, Second Edition, July 2006.

<sup>0</sup>D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources." IEEE Transactions on Information Theory, 19(4), pp. 471-480, July 1973.

<sup>0</sup>A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder." IEEE Transactions on Information Theory, pp. 1-10, January 1976.

## STARS Project-Team

### 3. Research Program

#### 3.1. Introduction

Stars follows three main research directions: perception for activity recognition, semantic activity recognition, and software engineering for activity recognition. **These three research directions are interleaved:** *the software engineering* research direction provides new methodologies for building safe activity recognition systems and *the perception* and *the semantic activity recognition* directions provide new activity recognition techniques which are designed and validated for concrete video analytic and healthcare applications. Conversely, these concrete systems raise new software issues that enrich the software engineering research direction.

Transversely, we consider a *new research axis in machine learning*, combining a priori knowledge and learning techniques, to set up the various models of an activity recognition system. A major objective is to automate model building or model enrichment at the perception level and at the understanding level.

#### 3.2. Perception for Activity Recognition

**Participants:** François Brémond, Sabine Moisan, Monique Thonnat.

Computer Vision; Cognitive Systems; Learning; Activity Recognition.

##### 3.2.1. Introduction

Our main goal in perception is to develop vision algorithms able to address the large variety of conditions characterizing real world scenes in terms of sensor conditions, hardware requirements, lighting conditions, physical objects, and application objectives. We have also several issues related to perception which combine machine learning and perception techniques: learning people appearance, parameters for system control and shape statistics.

##### 3.2.2. Appearance Models and People Tracking

An important issue is to detect in real-time physical objects from perceptual features and predefined 3D models. It requires finding a good balance between efficient methods and precise spatio-temporal models. Many improvements and analysis need to be performed in order to tackle the large range of people detection scenarios.

**Appearance models.** In particular, we study the temporal variation of the features characterizing the appearance of a human. This task could be achieved by clustering potential candidates depending on their position and their reliability. This task can provide any people tracking algorithms with reliable features allowing for instance to (1) better track people or their body parts during occlusion, or to (2) model people appearance for re-identification purposes in mono and multi-camera networks, which is still an open issue. The underlying challenge of the person re-identification problem arises from significant differences in illumination, pose and camera parameters. The re-identification approaches have two aspects: (1) establishing correspondences between body parts and (2) generating signatures that are invariant to different color responses. As we have already several descriptors which are color invariant, we now focus more on aligning two people detection and on finding their corresponding body parts. Having detected body parts, the approach can handle pose variations. Further, different body parts might have different influence on finding the correct match among a whole gallery dataset. Thus, the re-identification approaches have to search for matching strategies. As the results of the re-identification are always given as the ranking list, re-identification focuses on learning to rank. "Learning to rank" is a type of machine learning problem, in which the goal is to automatically construct a ranking model from a training data.

Therefore, we work on information fusion to handle perceptual features coming from various sensors (several cameras covering a large scale area or heterogeneous sensors capturing more or less precise and rich information). New 3D RGB-D sensors are also investigated, to help in getting an accurate segmentation for specific scene conditions.

**Long term tracking.** For activity recognition we need robust and coherent object tracking over long periods of time (often several hours in videosurveillance and several days in healthcare). To guarantee the long term coherence of tracked objects, spatio-temporal reasoning is required. Modeling and managing the uncertainty of these processes is also an open issue. In Stars we propose to add a reasoning layer to a classical Bayesian framework modeling the uncertainty of the tracked objects. This reasoning layer can take into account the a priori knowledge of the scene for outlier elimination and long-term coherency checking.

**Controlling system parameters.** Another research direction is to manage a library of video processing programs. We are building a perception library by selecting robust algorithms for feature extraction, by insuring they work efficiently with real time constraints and by formalizing their conditions of use within a program supervision model. In the case of video cameras, at least two problems are still open: robust image segmentation and meaningful feature extraction. For these issues, we are developing new learning techniques.

### 3.3. Semantic Activity Recognition

**Participants:** François Brémond, Sabine Moisan, Monique Thonnat.

Activity Recognition, Scene Understanding, Computer Vision

#### 3.3.1. Introduction

Semantic activity recognition is a complex process where information is abstracted through four levels: signal (e.g. pixel, sound), perceptual features, physical objects and activities. The signal and the feature levels are characterized by strong noise, ambiguous, corrupted and missing data. The whole process of scene understanding consists in analyzing this information to bring forth pertinent insight of the scene and its dynamics while handling the low level noise. Moreover, to obtain a semantic abstraction, building activity models is a crucial point. A still open issue consists in determining whether these models should be given a priori or learned. Another challenge consists in organizing this knowledge in order to capitalize experience, share it with others and update it along with experimentation. To face this challenge, tools in knowledge engineering such as machine learning or ontology are needed.

Thus we work along the following research axes: high level understanding (to recognize the activities of physical objects based on high level activity models), learning (how to learn the models needed for activity recognition) and activity recognition and discrete event systems.

#### 3.3.2. High Level Understanding

A challenging research axis is to recognize subjective activities of physical objects (i.e. human beings, animals, vehicles) based on a priori models and objective perceptual measures (e.g. robust and coherent object tracks).

To reach this goal, we have defined original activity recognition algorithms and activity models. Activity recognition algorithms include the computation of spatio-temporal relationships between physical objects. All the possible relationships may correspond to activities of interest and all have to be explored in an efficient way. The variety of these activities, generally called video events, is huge and depends on their spatial and temporal granularity, on the number of physical objects involved in the events, and on the event complexity (number of components constituting the event).

Concerning the modeling of activities, we are working towards two directions: the uncertainty management for representing probability distributions and knowledge acquisition facilities based on ontological engineering techniques. For the first direction, we are investigating classical statistical techniques and logical approaches. For the second direction, we built a language for video event modeling and a visual concept ontology (including color, texture and spatial concepts) to be extended with temporal concepts (motion, trajectories, events ...) and other perceptual concepts (physiological sensor concepts ...).



### 3.3.3. Learning for Activity Recognition

Given the difficulty of building an activity recognition system with a priori knowledge for a new application, we study how machine learning techniques can automate building or completing models at the perception level and at the understanding level.

At the understanding level, we are learning primitive event detectors. This can be done for example by learning visual concept detectors using SVMs (Support Vector Machines) with perceptual feature samples. An open question is how far can we go in weakly supervised learning for each type of perceptual concept (i.e. leveraging the human annotation task). A second direction is to learn typical composite event models for frequent activities using trajectory clustering or data mining techniques. We name composite event a particular combination of several primitive events.

### 3.3.4. Activity Recognition and Discrete Event Systems

The previous research axes are unavoidable to cope with the semantic interpretations. However they tend to let aside the pure event driven aspects of scenario recognition. These aspects have been studied for a long time at a theoretical level and led to methods and tools that may bring extra value to activity recognition, the most important being the possibility of formal analysis, verification and validation.

We have thus started to specify a formal model to define, analyze, simulate, and prove scenarios. This model deals with both absolute time (to be realistic and efficient in the analysis phase) and logical time (to benefit from well-known mathematical models providing re-usability, easy extension, and verification). Our purpose is to offer a generic tool to express and recognize activities associated with a concrete language to specify activities in the form of a set of scenarios with temporal constraints. The theoretical foundations and the tools being shared with Software Engineering aspects, they will be detailed in section 3.4 .

The results of the research performed in perception and semantic activity recognition (first and second research directions) produce new techniques for scene understanding and contribute to specify the needs for new software architectures (third research direction).

## 3.4. Software Engineering for Activity Recognition

**Participants:** Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, François Brémont.

Software Engineering, Generic Components, Knowledge-based Systems, Software Component Platform, Object-oriented Frameworks, Software Reuse, Model-driven Engineering

The aim of this research axis is to build general solutions and tools to develop systems dedicated to activity recognition. For this, we rely on state-of-the art Software Engineering practices to ensure both sound design and easy use, providing genericity, modularity, adaptability, reusability, extensibility, dependability, and maintainability.

This research requires theoretical studies combined with validation based on concrete experiments conducted in Stars. We work on the following three research axes: *models* (adapted to the activity recognition domain), *platform architecture* (to cope with deployment constraints and run time adaptation), and *system verification* (to generate dependable systems). For all these tasks we follow state of the art Software Engineering practices and, if needed, we attempt to set up new ones.

### 3.4.1. Platform Architecture for Activity Recognition

In the former project teams Orion and Pulsar, we have developed two platforms, one (VSIP), a library of real-time video understanding modules and another one, LAMA [14], a software platform enabling to design not only knowledge bases, but also inference engines, and additional tools. LAMA offers toolkits to build and to adapt all the software elements that compose a knowledge-based system.

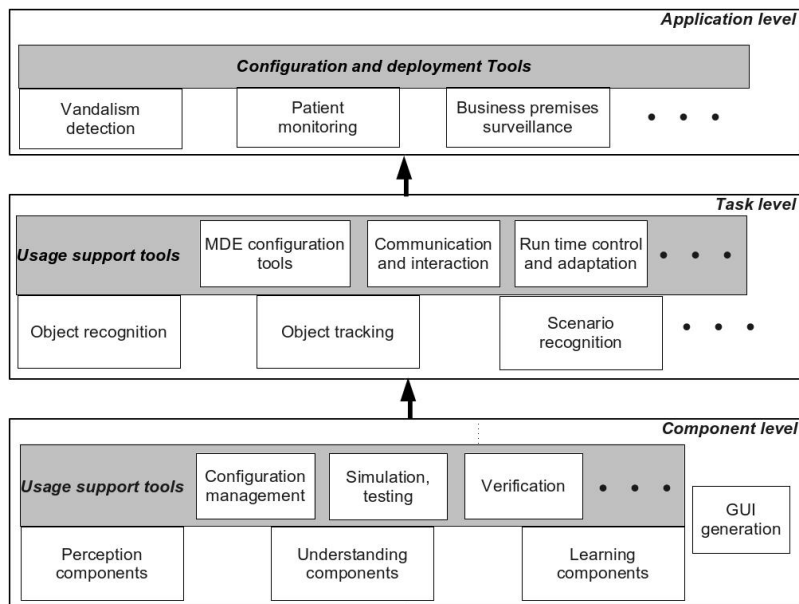


Figure 4. Global Architecture of an Activity Recognition The gray areas contain software engineering support modules whereas the other modules correspond to software components (at Task and Component levels) or to generated systems (at Application level).

Figure 4 presents our conceptual vision for the architecture of an activity recognition platform. It consists of three levels:

- The **Component Level**, the lowest one, offers software components providing elementary operations and data for perception, understanding, and learning.
  - *Perception components* contain algorithms for sensor management, image and signal analysis, image and video processing (segmentation, tracking...), etc.
  - *Understanding components* provide the building blocks for Knowledge-based Systems: knowledge representation and management, elements for controlling inference engine strategies, etc.
  - *Learning components* implement different learning strategies, such as Support Vector Machines (SVM), Case-based Learning (CBL), clustering, etc.

An Activity Recognition system is likely to pick components from these three packages. Hence, tools must be provided to configure (select, assemble), simulate, verify the resulting component combination. Other support tools may help to generate task or application dedicated languages or graphic interfaces.

- The **Task Level**, the middle one, contains executable realizations of individual tasks that will collaborate in a particular final application. Of course, the code of these tasks is built on top of the components from the previous level. We have already identified several of these important tasks: Object Recognition, Tracking, Scenario Recognition... In the future, other tasks will probably enrich this level.

For these tasks to nicely collaborate, communication and interaction facilities are needed. We shall also add MDE-enhanced tools for configuration and run-time adaptation.

- The **Application Level** integrates several of these tasks to build a system for a particular type of application, e.g., vandalism detection, patient monitoring, aircraft loading/unloading surveillance, etc.. Each system is parameterized to adapt to its local environment (number, type, location of sensors, scene geometry, visual parameters, number of objects of interest...). Thus configuration and deployment facilities are required.

The philosophy of this architecture is to offer at each level a balance between the widest possible genericity and the maximum effective reusability, in particular at the code level.

To cope with real application requirements, we shall also investigate distributed architecture, real time implementation, and user interfaces.

Concerning implementation issues, we shall use when possible existing open standard tools such as NuSMV for model-checking, Eclipse for graphic interfaces or model engineering support, Alloy for constraint representation and SAT solving for verification, etc. Note that, in Figure 4, some of the boxes can be naturally adapted from SUP existing elements (many perception and understanding components, program supervision, scenario recognition...) whereas others are to be developed, completely or partially (learning components, most support and configuration tools).

### 3.4.2. *Discrete Event Models of Activities*

As mentioned in the previous section (3.3) we have started to specify a formal model of scenario dealing with both absolute time and logical time. Our scenario and time models as well as the platform verification tools rely on a formal basis, namely the synchronous paradigm. To recognize scenarios, we consider activity descriptions as synchronous reactive systems and we apply general modeling methods to express scenario behavior.

Activity recognition systems usually exhibit many safeness issues. From the software engineering point of view we only consider software security. Our previous work on verification and validation has to be pursued; in particular, we need to test its scalability and to develop associated tools. Model-checking is an appealing technique since it can be automatized and helps to produce a code that has been formally proved. Our verification method follows a compositional approach, a well-known way to cope with scalability problems in model-checking.

Moreover, recognizing real scenarios is not a purely deterministic process. Sensor performance, precision of image analysis, scenario descriptions may induce various kinds of uncertainty. While taking into account this uncertainty, we should still keep our model of time deterministic, modular, and formally verifiable. To formally describe probabilistic timed systems, the most popular approach involves probabilistic extension of timed automata. New model checking techniques can be used as verification means, but relying on model checking techniques is not sufficient. Model checking is a powerful tool to prove decidable properties but introducing uncertainty may lead to infinite state or even undecidable properties. Thus model checking validation has to be completed with non exhaustive methods such as abstract interpretation.

### 3.4.3. *Model-Driven Engineering for Configuration and Control and Control of Video Surveillance systems*

Model-driven engineering techniques can support the configuration and dynamic adaptation of video surveillance systems designed with our SUP activity recognition platform. The challenge is to cope with the many—functional as well as nonfunctional—causes of variability both in the video application specification and in the concrete SUP implementation. We have used *feature models* to define two models: a generic model of video surveillance applications and a model of configuration for SUP components and chains. Both of them express variability factors. Ultimately, we wish to automatically generate a SUP component assembly from an application specification, using models to represent transformations [45]. Our models are enriched with intra- and inter-models constraints. Inter-models constraints specify models to represent transformations. Feature models are appropriate to describe variants; they are simple enough for video surveillance experts to express their requirements. Yet, they are powerful enough to be liable to static analysis [77]. In particular, the constraints can be analyzed as a SAT problem.

An additional challenge is to manage the possible run-time changes of implementation due to context variations (e.g., lighting conditions, changes in the reference scene, etc.). Video surveillance systems have to dynamically adapt to a changing environment. The use of models at run-time is a solution. We are defining adaptation rules corresponding to the dependency constraints between specification elements in one model and software variants in the other [44], [89], [82].

## THOTH Project-Team

### 3. Research Program

#### 3.1. Designing and learning structured models

The task of understanding image and video content has been interpreted in several ways over the past few decades, namely image classification, detecting objects in a scene, recognizing objects and their spatial extents in an image, estimating human poses, recovering scene geometry, recognizing activities performed by humans. However, addressing all these problems individually provides us with a partial understanding of the scene at best, leaving much of the visual data unexplained.

One of the main goals of this research axis is to go beyond the initial attempts that consider only a subset of tasks jointly, by developing novel models for a more complete understanding of scenes to address all the component tasks. We propose to incorporate the structure in image and video data explicitly into the models. In other words, our models aim to satisfy the complex sets of constraints that exist in natural images and videos. Examples of such constraints include: (i) relations between objects, like signs for shops indicate the presence of buildings, people on a road are usually walking or standing, (ii) higher-level semantic relations involving the type of scene, geographic location, and the plausible actions as a global constraint, e.g., an image taken at a swimming pool is unlikely to contain cars, (iii) relating objects occluded in some of the video frames to content in other frames, where they are more clearly visible as the camera or the object itself move, with the use of long-term trajectories and video object proposals.

This research axis will focus on three topics. The first is developing deep features for video. This involves designing rich features available in the form of long-range temporal interactions among pixels in a video sequence to learn a representation that is truly spatio-temporal in nature. The focus of the second topic is the challenging problem of modeling human activities in video, starting from human activity descriptors to building intermediate spatio-temporal representations of videos, and then learning the interactions among humans, objects and scenes temporally. The last topic is aimed at learning models that capture the relationships among several objects and regions in a single image scene, and additionally, among scenes in the case of an image collection or a video. The main scientific challenges in this topic stem from learning the structure of the probabilistic graphical model as well as the parameters of the cost functions quantifying the relationships among its entities. In the following we will present work related to all these three topics and then elaborate on our research directions.

- **Deep features for vision.** Deep learning models provide a rich representation of complex objects but in return have a large number of parameters. Thus, to work well on difficult tasks, a large amount of data is required. In this context, video presents several advantages: objects are observed from a large range of viewpoints, motion information allows the extraction of moving objects and parts, and objects can be differentiated by their motion patterns. We initially plan to develop deep features for videos that incorporate temporal information at multiple scales. We then plan to further exploit the rich content in video by incorporating additional cues, such as the detection of people and their body-joint locations in video, minimal prior knowledge of the object of interest, with the goal of learning a representation that is more appropriate for video understanding. In other words, a representation that is learned from video data and targeted at specific applications. For the application of recognizing human activities, this involves learning deep features for humans and their body-parts with all their spatiotemporal variations, either directly from raw video data or “pre-processed” videos containing human detections. For the application of object tracking, this task amounts to learning object-specific deep representations, further exploiting the limited annotation provided to identify the object.
- **Modeling human activities in videos.** Humans and their activities are not only one of the most frequent and interesting subjects in videos but also one of the hardest to analyze owing to the

complexity of the human form, clothing and movements. As part of this task, the Thoth project-team plans to build on state-of-the-art approaches for spatio-temporal representation of videos. This will involve using the dominant motion in the scene as well as the local motion of individual parts undergoing a rigid motion. Such motion information also helps in reasoning occlusion relationships among people and objects, and the state of the object. This novel spatio-temporal representation ultimately provides the equivalent of object proposals for videos, and is an important component for learning algorithms using minimal supervision. To take this representation even further, we aim to integrate the proposals and the occlusion relationships with methods for estimating human pose in videos, thus leveraging the interplay among body-joint locations, objects in the scene, and the activity being performed. For example, the locations of shoulder, elbow and wrist of a person drinking coffee are constrained to move in a certain way, which is completely different from the movement observed when a person is typing. In essence, this step will model human activities by dynamics in terms of both low-level movements of body-joint locations and global high-level motion in the scene.

- **Structured models.** The interactions among various elements in a scene, such as, the objects and regions in it, the motion of object parts or entire objects themselves, form a key element for understanding image or video content. These rich cues define the structure of visual data and how it evolves spatio-temporally. We plan to develop a novel graphical model to exploit this structure. The main components in this graphical model are spatio-temporal regions (in the case of video or simply image regions), which can represent object parts or entire objects themselves, and the interactions among several entities. The dependencies among the scene entities are defined with a higher order or a global cost function. A higher order constraint is a generalization of the pairwise interaction term, and is a cost function involving more than two components in the scene, e.g., several regions, whereas a global constraint imposes a cost term over the entire image or video, e.g., a prior on the number of people expected in the scene. The constraints we plan to include generalize several existing methods, which are limited to pairwise interactions or a small restrictive set of higher-order costs. In addition to learning the parameters of these novel functions, we will focus on learning the structure of the graph itself—a challenging problem that is seldom addressed in current approaches. This provides an elegant way to go beyond state-of-the-art deep learning methods, which are limited to learning the high-level interaction among parts of an object, by learning the relationships among objects.

### 3.2. Learning of visual models from minimal supervision

Today's approaches to visual recognition learn models for a limited and fixed set of visual categories with fully supervised classification techniques. This paradigm has been adopted in the early 2000's, and within it enormous progress has been made over the last decade.

The scale and diversity in today's large and growing image and video collections (such as, e.g., broadcast archives, and personal image/video collections) call for a departure from the current paradigm. This is the case because to answer queries about such data, it is unfeasible to learn the models of visual content by manually and precisely annotating every relevant concept, object, scene, or action category in a representative sample of everyday conditions. For one, it will be difficult, or even impossible to decide a-priori what are the relevant categories and the proper granularity level. Moreover, the cost of such annotations would be prohibitive in most application scenarios. One of the main goals of the Thoth project-team is to develop a new framework for learning visual recognition models by actively exploring large digital image and video sources (off-line archives as well as growing on-line content), and exploiting the weak supervisory signal provided by the accompanying metadata (such as captions, keywords, tags, subtitles, or scripts) and audio signal (from which we can for example extract speech transcripts, or exploit speaker recognition models).

Textual metadata has traditionally been used to index and search for visual content. The information in metadata is, however, typically sparse (e.g., the location and overall topic of newscasts in a video archive <sup>0</sup>)

<sup>0</sup>For example at the Dutch national broadcast archive Netherlands Institute of Sound and Vision, with whom we collaborated in the EU FP7 project AXES, typically one or two sentences are used in the metadata to describe a one hour long TV program.

and noisy (e.g., a movie script may tell us that two persons kiss in some scene, but not when, and the kiss may occur off screen or not have survived the final cut). For this reason, metadata search should be complemented by visual content based search, where visual recognition models are used to localize content of interest that is not mentioned in the metadata, to increase the usability and value of image/video archives. *The key insight that we build on in this research axis is that while the metadata for a single image or video is too sparse and noisy to rely on for search, the metadata associated with large video and image databases collectively provide an extremely versatile source of information to learn visual recognition models.* This form of “embedded annotation” is rich, diverse and abundantly available. Mining these correspondences from the web, TV and film archives, and online consumer generated content sites such as Flickr, Facebook, or YouTube, guarantees that the learned models are representative for many different situations, unlike models learned from manually collected fully supervised training data sets which are often biased.

The approach we propose to address the limitations of the fully supervised learning paradigm aligns with “Big Data” approaches developed in other areas: we rely on the orders-of-magnitude-larger training sets that have recently become available with metadata to compensate for less explicit forms of supervision. This will form a sustainable approach to learn visual recognition models for a much larger set of categories with little or no manual intervention. Reducing and ultimately removing the dependency on manual annotations will dramatically reduce the cost of learning visual recognition models. This in turn will allow such models to be used in many more applications, and enable new applications based on visual recognition beyond a fixed set of categories, such as natural language based querying for visual content. This is an ambitious goal, given the sheer volume and intrinsic variability of the every day visual content available on-line, and the lack of a universally accepted formalism for modeling it. Yet, the potential payoff is a breakthrough in visual object recognition and scene understanding capabilities.

This research axis is organized into the following three sub-tasks:

- **Weakly supervised learning.** For object localization we will go beyond current methods that learn one category model at a time and develop methods that learn models for different categories concurrently. This allows “explaining away” effects to be leveraged, i.e., if a certain region in an image has been identified as an instance of one category, it cannot be an instance of another category at the same time. For weakly supervised detection in video we will consider detection proposal methods. While these are effective for still images, recent approaches for the spatio-temporal domain need further improvements to be similarly effective. Furthermore, we will exploit appearance and motion information jointly over a set of videos. In the video domain we will also continue to work on learning recognition models from subtitle and script information. The basis of leveraging the script data which does not have a temporal alignment with the video, is to use matches in the narrative in the script and the subtitles (which do have a temporal alignment with the video). We will go beyond simple correspondences between names and verbs relating to self-motion, and match more complex sentences related to interaction with objects and other people. To deal with the limited amount of occurrences of such actions in a single movie, we will consider approaches that learn action models across a collection of movies.
- **Online learning of visual models.** As a larger number of visual category models is being learned, online learning methods become important, since new training data and categories will arrive over time. We will develop online learning methods that can incorporate new examples for existing category models, and learn new category models from few examples by leveraging similarity to related categories using multi-task learning methods. Here we will develop new distance-based classifiers and attribute and label embedding techniques, and explore the use of NLP techniques such as skipgram models to automatically determine between which classes transfer should occur. Moreover, NLP will be useful in the context of learning models for many categories to identify synonyms, and to determine cases of polysemy (e.g. jaguar car brand v.s. jaguar animal), and merge or refine categories accordingly. Ultimately this will result in methods that are able to learn an “encyclopedia” of visual models.
- **Visual search from unstructured textual queries.** We will build on recent approaches that learn

recognition models on-the-fly (as the query is issued) from generic image search engines such as Google Images. While it is feasible to learn models in this manner in a matter of seconds, it is challenging to use the model to retrieve relevant content in real-time from large video archives of more than a few thousand hours. To achieve this requires feature compression techniques to store visual representations in memory, and cascaded search techniques to avoid exhaustive search. This approach, however, leaves untouched the core problem of how to associate visual material with the textual query in the first place. The second approach we will explore is based on image annotation models. In particular we will go beyond image-text retrieval methods by using recurrent neural networks such as Elman networks or long short-term memory (LSTM) networks to generate natural language sentences to describe images.

### 3.3. Large-scale learning and optimization

We have entered an era of massive data acquisition, leading to the revival of an old scientific utopia: it should be possible to better understand the world by automatically converting data into knowledge. It is also leading to a new economic paradigm, where data is a valuable asset and a source of activity. Therefore, developing scalable technology to make sense of massive data has become a strategic issue. Computer vision has already started to adapt to these changes.

In particular, very high dimensional models such as deep networks are becoming highly popular and successful for visual recognition. This change is closely related to the advent of big data. On the one hand, these models involve a huge number of parameters and are rich enough to represent well complex objects such as natural images or text corpora. On the other hand, they are prone to overfitting (fitting too closely to training data without being able to generalize to new unseen data) despite regularization; to work well on difficult tasks, they require a large amount of labelled data that has been available only recently. Other cues may explain their success: the deep learning community has made significant engineering efforts, making it possible to learn in a day on a GPU large models that would have required weeks of computations on a traditional CPU, and it has accumulated enough empirical experience to find good hyper-parameters for its networks.

To learn the huge number of parameters of deep hierarchical models requires scalable optimization techniques and large amounts of data to prevent overfitting. This immediately raises two major challenges: how to learn without large amounts of labeled data, or with weakly supervised annotations? How to efficiently learn such huge-dimensional models? To answer the above challenges, we will concentrate on the design and theoretical justifications of deep architectures including our recently proposed deep kernel machines, with a focus on weakly supervised and unsupervised learning, and develop continuous and discrete optimization techniques that push the state of the art in terms of speed and scalability.

This research axis will be developed into three sub-tasks:

- **Deep kernel machines for structured data.** Deep kernel machines combine advantages of kernel methods and deep learning. Both approaches rely on high-dimensional models. Kernels implicitly operate in a space of possibly infinite dimension, whereas deep networks explicitly construct high-dimensional nonlinear data representations. Yet, these approaches are complementary: Kernels can be built with deep learning principles such as hierarchies and convolutions, and approximated by multilayer neural networks. Furthermore, kernels work with structured data and have well understood theoretical principles. Thus, a goal of the Thoth project-team is to design and optimize the training of such deep kernel machines.
- **Large-scale parallel optimization.** Deep kernel machines produce nonlinear representations of input data points. After encoding these data points, a learning task is often formulated as a *large-scale convex optimization problem*; for example, this is the case for linear support vector machines, logistic regression classifiers, or more generally many empirical risk minimization formulations. We intend to pursue recent efforts for making convex optimization techniques that are dedicated to machine learning more scalable. Most existing approaches address scalability issues either in model size (meaning that the function to minimize is defined on a domain of very high dimension), or in the amount of training data (typically, the objective is a large sum of elementary functions). There is



thus a large room for improvements for techniques that jointly take these two criteria into account.

- **Large-scale graphical models.** To represent structured data, we will also investigate graphical models and their optimization. The challenge here is two-fold: designing an adequate cost function and minimizing it. While several cost functions are possible, their utility will be largely determined by the efficiency and the effectiveness of the optimization algorithms for solving them. It is a combinatorial optimization problem involving billions of variables and is NP-hard in general, requiring us to go beyond the classical approximate inference techniques. The main challenges in minimizing cost functions stem from the large number of variables to be inferred, the inherent structure of the graph induced by the interaction terms (e.g., pairwise terms), and the high-arity terms which constrain multiple entities in a graph.

### 3.4. Datasets and evaluation

Standard benchmarks with associated evaluation measures are becoming increasingly important in computer vision, as they enable an objective comparison of state-of-the-art approaches. Such datasets need to be relevant for real-world application scenarios; challenging for state-of-the-art algorithms; and large enough to produce statistically significant results.

A decade ago, small datasets were used to evaluate relatively simple tasks, such as for example interest point matching and detection. Since then, the size of the datasets and the complexity of the tasks gradually evolved. An example is the Pascal Visual Object Challenge with 20 classes and approximately 10,000 images, which evaluates object classification and detection. Another example is the ImageNet challenge, including thousands of classes and millions of images. In the context of video classification, the TrecVid Multimedia Event Detection challenges, organized by NIST, evaluate activity classification on a dataset of over 200,000 video clips, representing more than 8,000 hours of video, which amounts to 11 months of continuous video.

Almost all of the existing image and video datasets are annotated by hand; it is the case for all of the above cited examples. In some cases, they present limited and unrealistic viewing conditions. For example, many images of the ImageNet dataset depict upright objects with virtually no background clutter, and they may not capture particularly relevant visual concepts: most people would not know the majority of subcategories of snakes cataloged in ImageNet. This holds true for video datasets as well, where in addition a taxonomy of action and event categories is missing.

Our effort on data collection and evaluation will focus on two directions. First, we will design and assemble video datasets, in particular for action and activity recognition. This includes defining relevant taxonomies of actions and activities. Second, we will provide data and define evaluation protocols for weakly supervised learning methods. This does not mean of course that we will forsake human supervision altogether: some amount of ground-truth labeling is necessary for experimental validation and comparison to the state of the art. Particular attention will be paid to the design of efficient annotation tools.

Not only do we plan to collect datasets, but also to provide them to the community, together with accompanying evaluation protocols and software, to enable a comparison of competing approaches for action recognition and large-scale weakly supervised learning. Furthermore, we plan to set up evaluation servers together with leaderboards, to establish an unbiased state of the art on held out test data for which the ground-truth annotations are not distributed. This is crucial to avoid tuning the parameters for a specific dataset and to guarantee a fair evaluation.

- **Action recognition.** We will develop datasets for recognizing human actions and human-object interactions (including multiple persons) with a significant number of actions. Almost all of today's action recognition datasets evaluate classification of short video clips into a number of predefined categories, in many cases a number of different sports, which are relatively easy to identify by their characteristic motion and context. However, in many real-world applications the goal is to identify and localize actions in entire videos, such as movies or surveillance videos of several hours. The actions targeted here are "real-world" and will be defined by compositions of atomic actions into higher-level activities. One essential component is the definition of relevant taxonomies of actions

and activities. We think that such a definition needs to rely on a decomposition of actions into poses, objects and scenes, as determining all possible actions without such a decomposition is not feasible. We plan to provide annotations for spatio-temporal localization of humans as well as relevant objects and scene parts for a large number of actions and videos.

- **Weakly supervised learning.** We will collect weakly labeled images and videos for training. The collection process will be semi-automatic. We will use image or video search engines such as Google Image Search, Flickr or YouTube to find visual data corresponding to the labels. Initial datasets will be obtained by manually correcting whole-image/video labels, i.e., the approach will evaluate how well the object model can be learned if the entire image or video is labeled, but the object model has to be extracted automatically. Subsequent datasets will feature noisy and incorrect labels. Testing will be performed on PASCAL VOC'07 and ImageNet, but also on more realistic datasets similar to those used for training, which we develop and manually annotate for evaluation. Our dataset will include both images and videos, the categories represented will include objects, scenes as well as human activities, and the data will be presented in realistic conditions.
- **Joint learning from visual information and text.** Initially, we will use a selection from the large number of movies and TV series for which scripts are available on-line, see for example <http://www.dailyscript.com> and <http://www.weeklyscript.com>. These scripts can easily be aligned with the videos by establishing correspondences between script words and (timestamped) spoken ones obtained from the subtitles or audio track. The goal is to jointly learn from visual content and text. To measure the quality of such a joint learning, we will manually annotate some of the videos. Annotations will include the space-time locations of the actions as well as correct parsing of the sentence. While DVDs will, initially, receive most attention, we will also investigate the use of data obtained from web pages, for example images with captions, or images and videos surrounded by text. This data is by nature more noisy than scripts.

## WILLOW Project-Team

### 3. Research Program

#### 3.1. 3D object and scene modeling, analysis, and retrieval

This part of our research focuses on geometric models of specific 3D objects at the local (differential) and global levels, physical and statistical models of materials and illumination patterns, and modeling and retrieval of objects and scenes in large image collections. Our past work in these areas includes research aimed at recognizing rigid 3D objects in cluttered photographs taken from arbitrary viewpoints (Rothganger *et al.*, 2006), segmenting video sequences into parts corresponding to rigid scene components before recognizing these in new video clips (Rothganger *et al.*, 2007), retrieval of particular objects and buildings from images and videos (Sivic and Zisserman, 2003) and (Philbin *et al.*, 2007), and a theoretical study of a general formalism for modeling central and non-central cameras using the formalism and terminology of classical projective geometry (Ponce, 2009 and Batog *et al.*, 2010).

We have also developed multi-view stereopsis algorithms that have proven remarkably effective at recovering intricate details and thin features of compact objects and capturing the overall structure of large-scale, cluttered scenes. We have obtained a US patent 8,331,615<sup>0</sup> for the corresponding software (PMVS, <https://github.com/pmoulon/CMVS-PMVS>) which is available under a GPL license and used for film production by ILM and Weta as well as by Google in Google Maps. It is also the basic technology used by Iconem, a start-up founded by Y. Ubelmann, a Willow collaborator. We have also applied our multi-view-stereo approach to model archaeological sites together with developing representations and efficient retrieval techniques to enable matching historical paintings to 3D models of archaeological sites (Russel *et al.*, 2011).

Our current efforts in this area are outlined in detail in Section 7.1 .

#### 3.2. Category-level object and scene recognition

The objective in this core part of our research is to learn and recognize quickly and accurately thousands of visual categories, including materials, objects, scenes, and broad classes of temporal events, such as patterns of human activities in picnics, conversations, etc. The current paradigm in the vision community is to model/learn one object category (read 2D aspect) at a time. If we are to achieve our goal, we have to break away from this paradigm, and develop models that account for the tremendous variability in object and scene appearance due to texture, material, viewpoint, and illumination changes within each object category, as well as the complex and evolving relationships between scene elements during the course of normal human activities.

Our current work in this area is outlined in detail in Section 7.2 .

#### 3.3. Image restoration, manipulation and enhancement

The goal of this part of our research is to develop models, and methods for image/video restoration, manipulation and enhancement. The ability to "intelligently" manipulate the content of images and video is just as essential as high-level content interpretation in many applications: This ranges from restoring old films or removing unwanted wires and rigs from new ones in post production, to cleaning up a shot of your daughter at her birthday party, which is lovely but noisy and blurry because the lights were out when she blew the candles, or editing out a tourist from your Roman holiday video. Going beyond the modest abilities of current "digital zoom" (bicubic interpolation in general) so you can close in on that birthday cake, "deblock" a football game on TV, or turn your favorite DVD into a blue-ray, is just as important.

---

<sup>0</sup>The patent: "Match, Expand, and Filter Technique for Multi-View Stereopsis" was issued December 11, 2012 and assigned patent number 8,331,615.

In this context, we believe there is a new convergence between computer vision, machine learning, and signal processing. For example: The idea of exploiting self-similarities in image analysis, originally introduced in computer vision for texture synthesis applications (Efros and Leung, 1999), is the basis for non-local means (Buades *et al.*, 2005), one of today's most successful approaches to image restoration. In turn, by combining a powerful sparse coding approach to non-local means (Dabov *et al.*, 2007) with modern machine learning techniques for dictionary learning (Mairal *et al.*, 2010), we have obtained denoising and demosaicking results that are the state of the art on standard benchmarks (Mairal *et al.*, 2009).

Our current work is outlined in detail in Section 7.3 .

### 3.4. Human activity capture and classification

From a scientific point of view, visual action understanding is a computer vision problem that until recently has received little attention outside of extremely specific contexts such as surveillance or sports. Many of the current approaches to the visual interpretation of human activities are designed for a limited range of operating conditions, such as static cameras, fixed scenes, or restricted actions. The objective of this part of our project is to attack the much more challenging problem of understanding actions and interactions in unconstrained video depicting everyday human activities such as in sitcoms, feature films, or news segments. The recent emergence of automated annotation tools for this type of video data (Everingham, Sivic, Zisserman, 2006; Laptev, Marszałek, Schmid, Rozenfeld, 2008; Duchenne, Laptev, Sivic, Bach, Ponce, 2009) means that massive amounts of labelled data for training and recognizing action models will at long last be available.

Our research agenda in this scientific domain is described below and our recent results are outlined in detail in Section 7.4 .

- **Weakly-supervised learning and annotation of human actions in video.** We aim to leverage the huge amount of video data using readily-available annotations in the form of video scripts. Scripts, however, often provide only imprecise and incomplete information about the video. We address this problem with weakly-supervised learning techniques both at the text and image levels.
- **Descriptors for video representation** Video representation has a crucial role for recognizing human actions and other components of a visual scene. Our work in this domain aims to develop generic methods for representing video data based on realistic assumptions. In particular, we develop deep learning methods and design new trainable representations for various tasks such as human action recognition, person detection, segmentation and tracking.