RESEARCH CENTER

FIELD
Perception, Cognition, Interaction

Activity Report 2012

Section New Results

Edition: 2013-04-24

ALPAGE Project-Team

6. New Results

6.1. Advances in symbolic and hybrid parsing with DyALog and FRMG

Participants: Éric Villemonte de La Clergerie, François Barthélemy, Julien Martin.

Within the team is developed a wide-coverage French meta-grammar (FRMG) and a efficient hybrid TAG/TIG parser based on the DyALog logic programming environment [120] and on the Le*fff* morphological and syntactic lexicon [105]. It relies on the notion of factorized grammar, themselves generated from a representation that lies at a higher level of abstraction, named Meta-Grammars [122]. At that level, linguistic generalizations can be expressed, which in turn makes it possible to transfer meta-grammars from one language to a closely related one. The hybrid TAG/TIG parser generator itself implements all kinds of parsing optimizations: lexicalization (in particular via hypertags), left-corner guiding, top/bottom feature analysis, TIG analysis (with multiple adjoining), and others. The recent evolutions go towards an hybridization with statistical approaches.

6.1.1. Tuning FRMG's disambiguation mechanism

Continuing works initiated in 2011 on the exploitation of the dependency version of the French TreeBank (FTB), Éric de La Clergerie has explored the tuning of FRMG's rule base disambiguation mechanism using a larger set of features and weight learned from the FTB. In 2011, this approach led to on improvement from 82.31% to 84.54% in terms of accuracy (LAS - Labelled Attachment Score) on the test part of the FTB. By increasing the set of features, in particularly using higher-order dependency features (on parent edge and sibling edges), and a better understanding of the iterative tuning mechanism, it was possible to reach 85.95% LAS. This tuning mechanism is based on the idea of adding or subtracting some weight to a disambiguation rule given some specific contexts (provided by the features), where the delta is progressively learned from the accuracy of the disambiguation rule in terms of edge selection or rejection. The learning algorithm presents some relationships with the perceptron approach, but the use of a more standard implementation of the perceptron led to less interesting gains.

During the same time, the coverage of FRMG was improved (to reach for instance 94% of full parses on the FTB).

6.1.2. Synchronous Tree-Adjoining Grammars

A preliminary work has been done to implement *Synchronous Tree-Adjoining Grammars* (STAGs) in DyALog, relying on the notion of *Thread Automata* [119]. Synchronous Tree Adjoining Grammars is an instance of formalism where the order of the components of a tree structure is not fully determined. This leads to combinatorial alternatives when parsing, while a tree-structure corresponding to the input string has to be build. A specific front-end has been written to implement STAGs. The work on the back-end is still in progress, with the goal to have a common intermediate representation for several mildly context-sensitive formalisms where some node operations non-deterministically pick a node out of a finite set of nodes. STAGs are an instance of such formalisms, Multi-Component Tree Adjoining Grammars (MCTAGs) are another instance. The intermediate representation consists in Thread Automata (TA), an extension of Push-Down Automata where several threads of computations are considered and only one is active at any time.

6.1.3. Adding weights and probabilities to DyALog

Weights can already be used during the disambiguation phase of the FRMG parser, implemented in DyALog. However, a deeper implementation of weights and probabilities in DyALog was initiated in 2012 by Julien Martin during his Master internship. By enriching the structure of the backpointers (relating the items to their parent items), it is now possible to maintain an ordered weighted list of derivations, to update the scheduling of items wrt their weight, to update the weights of all the descendants of an item I when updating I's weight. The motivation is of course to be able to favor the best analysis first during parsing. A second objective (which has

been implemented) is the possibility to extract the n-best parses after parsing (but keeping a shared derivation forest). A third objective, remaining to be done, is related to the use of beam search techniques to prune the search space during parsing. A longer-term objective is the abstraction of this work to be able to work on semi-rings.

6.2. Tree transformation

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

In 2011, the conversion of native FRMG dependencies into the CONLL dependency scheme was the occasion to explore new ideas about tree transformation (for dependencies), based on the notion of two-level transformation with a first level relying on local transformation rules and a second level being controlled by constraints carried by the first level edges. During his Master internship, Corentin Ribeyre has formalized and re-implemented this approach in a more systematic and generic way. This work was also completed by the use of example-based learning techniques to quickly learn the local transformation rules of the first level. The line of research is motivated by possibility to quickly develop a reduced set of transformation rules (thanks to the examples and the constraint level) for a large variety of applications, such as information extration but also conversion toward a deep syntax level or a shallow semantic level. A poster paper was presented at TAG+11 [29].

6.3. lexical knowledge acquisition and visualization

Participants: Éric Villemonte de La Clergerie, Mickael Morardo, Benoît Sagot.

In relation with our collaboration with Lingua & Machina (cf section 4.4), Mikael Morardo has enriched the interfaces of the WEB platform Libellex for the visualization and validation of more complex lexical resource. In particular, the focus has been on the development of a graph-based view with the javascript Library `d3.js` to represent large lexical networks. The current implementation is powerful enough to deal with large networks of several teens of thousands of connections, allowing the visualization of fragments of the network and an easy navigation. Because the graph-view proved to be both intuitive and efficient, the previous list-based view for terminology was partially re-implemented in the new graph-view. It was also extended for visualizing and validating more complex lexical networks, like the French Wordnet WOLF coupled with the original English WordNet (cf 5.9).

The graph-based view was used to explore several networks built using Harris' distributional hypothesis (through a clustering algorithm) on the output of FRMG for several corpora. Because terminology was now be visualized at the same time, the clustering algorithm was modified to be able to take into account a list of terms (also automatically extracted from the parsed corpora) .

6.4. Advances in statistical parsing

Participants: Marie Candito, Benoît Crabbé, Djamé Seddah, Enrique Henestroza Anguiano.

6.4.1. Statistical Parsing

We have achieved **state-of-the art results for French statistical parsing**, adapting existing techniques for French, a language with a morphology richer than English, either for constituency parsing [110], [113] or dependency parsing [68]. We made available The Bonsai parsing chain [1] (cf. 5.4), that gathers preprocessing tools and models for French dependency parsing into an easy-to-use parsing tool for French. We designed our parsing pipeline with modularity in mind: our parsing models are interchangeable. For instance, dependencies output can either be generated from a PCFG-LA based parser associated with a functional role labeler or from any dependency parsers trained on our dependency treebank [68]. Tokens can either be raw words, POS tagged lemmas or word clusters [69].

[1] http://alpage.inria.fr/statgram/frdep/fr_stat_dep_parsing.html

We have innovated in the tuning of tagsets to optimize both grammar induction and unknown word handling [75], thus providing the best parsing models for French [111]. Then we have contributed on three main points:

1. conversion of the French Treebank [55] used as constituency training data into a dependency treebank [4], which is now used by several teams for dependency parsing;
2. an original method to reduce lexical data sparseness by replacing tokens by unsupervised word clusters, or morphological clusters [64], [112];
3. a postprocessing step that uses specialized statistical models for parse correction [81].

For the last 18 to 12 months, we have been increasingly focused in increasing the robustness of our parsing models by (a) validating our approach on other morphologically-rich languages; (b) other domains and (c) on user generated content. All of those challenging the current state-of-the-art in statistical parsing.

6.4.2. Multilingual parsing

Applying the techniques we developed for reducing lexical data, which is commonly found in morphologically-rich languages (MRLs) and optimizing the POS tagset, we integrated lexical information through data driven lemmatisation [112] and POS tagging [79]. This provided state-of-the-art results in parsing Romance languages such as Italian [35] and Spanish [26]. In the latter case, we mixed the outputs of two morphological analyzers and generated a version of the treebank where each morphological gold information was replaced by a predicted one. Relying on a rich lexicon developed within the Alexina framework (cf. 5.8) and accurate morphological treatment (cf. 6.5), this method brings more robustness to treebank-based parsing models.

6.4.3. Out-of-domain parsing : resources and parsing techniques

Statistical parsing is known to lead to parsers that exhibit quite degraded performance on input text that varies from the sentences used for training. Alpage has devoted a major effort on providing both evaluation resources and parser adaptation techniques, to increase robustness of statistical parsing for French. We have investigated several degrees of distance between the training corpus, the French Treebank, which is made of sentences from the *Le Monde* newspaper: we first focused on parsing well-edited texts, but from domains with varying difference with respect to the national newspaper *Le Monde* type of text. We then turned our attention to parsing user-generated content, hence potentially not only from a different domain than news, but also with great "noise" with respect to well-edited texts, and extremely divergent linguistic phenomena (see next subsection). As far as out-of-domain well-edited text, we have supervised the annotation and release of the **Sequoia Treebank** [47] (https://www.rocq.inria.fr/alpage-wiki/tiki-index.php?page=CorpusSequoia), a corpus of 3200 sentences annotated for part-of-speech and syntactic structure, from four subdomains : sentences from the regional newspaper *L'Est Républicain*, from the French Wikipedia, from the Europarl Corpus (European parliamentary debates), and from reports of the European Medicine Agency. We have proposed a word clustering technique, with clusters computed over a "*bridge*" corpus that couples indomain and target domain raw texts, to improve parsing performance on target domain, without degrading performance on indomain texts (contrary to usual adaptation techniques such as self-training). Preliminary experiments were performed on the biomedical domain only [67] and confirmed on the whole Sequoia Treebank [47].

6.4.4. Robust parsing of user-generated content

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 [116]. 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 analyze and interpret automatically.

In order to fully prepare a shift toward more robustness, we started to develop a richly annotated corpus of user-generated French text, the French Social Media Bank, which includes not only POS, constituency and functional information, but also a layer of "normalized" text[37]. This corpus is fully available and constitutes the first data set on Facebook data and the first instance of user generated content for an MRL.

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 MElt, a state-of-the art POS tagging system [15]. 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 [93], as Alpage was ranked second and third [38]. For achieving this result, we brought together a preliminary implementation of a normalization wrapper around the MElt POS tagger followed by a state-of-the art statistical parser improved by several domain adaptation techniques originally developed for parsing edited out-of-domain texts (cf. previous section).

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.

6.4.5. Precise recovery of unbounded dependencies

We focused on a linguistic phenomena known as long-distance dependencies. These are dependencies involved a fronted element that depends on a head that is potentially embedded in the clause the element is in front of. This embedding make such dependencies very hard to recover for a parser. Though this phenomena is rare, the corresponding dependencies are generally part of predicate-argument structures, and are thus very important to recover for downstream semantic applications. We have assessed the low parsing performance of long-distance dependencies (LDDs) for French, proposed an explicit annotation of such dependencies in the French Treebank and the Sequoia Treebank, and evaluated several parsing architectures with the aim of maintaining high general performance and good performance on LDDs [22]. We found that using a non-projective parser helps for LDDs but degrades overall performance, while using pseudo-projective parsing [88] which transforms in a reversible way a non-projective treebank into a projective one) is the best strategy, in order to take advantage of the better performance of projective parsers.

6.5. Computational morphology and automatic morphological analysis

Participants: Benoît Sagot [correspondant], Marion Baranes, Virginie Mouilleron, Damien Nouvel.

Since 2011 and, Alpage members have started interacting with formal morphologists for taking part in the development and implementation of new morphological models and resources. Concerning inflectional morphology, this work has led to new versions of the morphological layer of the ALEXINA formalism, to new ALEXINA lexicons for several languages of choice (Kurdish languages and German, as mentioned above, but also Maltese and Latin, see the section on ALEXINA), and to studies about the quantitative assessment of morphological complexity, currently an active area of research in morphology, have been pursued following previous work published in 2011 [109], [126]. Concerning constructional morphology (derivation, composition) and borrowings, studies and experiments have been carried out in the context of the ANR EDyLex project and that of the collaboration with *viavoo* [45], following here as well experiments carried out in 2011 [124], [115], [127].

6.6. Advances in lexical morphology and syntax

Participants: Benoît Sagot [correspondant], Laurence Danlos, Éric Villemonte de La Clergerie.

The Alexina framework (cf. 5.8) [105] has been developed and used for developing various lexicons, in particular the Le*fff*, that are used in many tools such as POS-taggers [15] and parsers.

In 2012, the new developments within Alexina have been fourfold:

- A large amount of work has been made for developing a new morphological layer to Alexina, in collaboration with a specialist of formal morphology.
- In the context of this collaboration, new Alexina lexicons have been developed with a special focus on linguistic relevance and exhaustivity within a well-defined subset of lexical entries (e.g., Latin verbs, 1st-binyan Maltese verbs).
- The development of a new large-scale NLP-oriented Alexina lexicon has been initiated, namely that of DeLex, an Alexina lexicon for German. It is currently restricted to the morphological layer (no valency information yet) but already generates 2 million inflected lexical entries. The underlying morphological grammar makes use of the new morphological layer mentioned above.
- Following previous work, merging experiments between syntactic resources and the Le*fff* [30] and comparison experiments between such resources and the Le*fff* as reference lexicon for the FRMG parser have been carried out [43]. In the latter series of experiments, the Le*fff* has proven better, or rather more suitable, that other (converted) resources.

6.7. Named Entity Recognition and Entity Linking

Participants: Rosa Stern, Benoît Sagot.

Identifying named entities is a widely studied issue in Natural Language Processing, because named entities are crucial targets in information extraction or retrieval tasks, but also for preparing further NLP tasks (e.g., parsing). Therefore a vast amount of work has been published that is dedicated to named entity *recognition*, i.e., the task of identification of named entity *mentions* (spans of text denoting a named entity), and sometimes *types*. However, real-life applications need not only identify named entity mentions, but also know which real entity they refer to; this issue is addressed in tasks such as knowledge base population with entity resolution and linking, which require an inventory of entities is required prior to those tasks in order to constitute a reference.

6.7.1. Cooperation of symbolic and statistical methods for named entity recognition and typing

Named entity recognition and typing is achieved both by symbolic and probabilistic systems. We have performed an experiment [62] for making the rule-based system NP, SXPipe's high-precision named entity recognition system developed at Alpage on AFP news corpora and which relies on the *Aleda* named entity database, interact with LIANE, a high-recall probabilistic system developed by Frédéric Béchet (LIF) and trained on oral transcriptions from the ESTER corpus. We have shown that a probabilistic system such as LIANE can be adapted to a new type of corpus in a non-supervised way thanks to large-scale corpora automatically annotated by NP. This adaptation does not require any additional manual annotation and illustrates the complementarity between numeric and symbolic techniques for tackling linguistic tasks.

6.7.2. Nomos, a statistical entity linking system

For information extraction from news wires, entities such as persons, locations or organizations are especially relevant in a knowledge acquisition context. Through a process of named entity recognition and entity linking applied jointly, we aim at the extraction and complete identification of these relevant entities, which are meant to enrich textual content in the form of *metadata*. In order to store and access extracted knowledge in a structured and coherent way, we aim at populating an ontological reference base with these metadata. We have pursued our efforts in this direction, using an approach where NLP tools have early access to Linked Data resources and thus have the ability to produce metadata integrated in the Linked Data framework. In particular, we have studied how the entity linking process in this task must deal with noisy data, as opposed to the general case where only correct entity identification is provided.

We use the symbolic named entity recognition system NP, a component of SXPipe, and use it as a mention detection module. Its output is then processed through our entity linking system, which is based on a supervised model learned from examples of linked entities. Since our named entity recognition is not deterministic, as opposed to other entity linking tasks where the gold named entity recognition results are provided, it is configured to remain ambiguous and non-deterministic, i.e., its output preserves a number of ambiguities which are usually resolved at this level. In particular, no disambiguation is made in the cases of multiple possible mentions boundaries (e.g., *{Paris}+{Hilton}* vs. *{Paris Hilton}*). In order to cope with possible false mention matches, which should be discarded as linking queries, the named entity recognition output is made more ambiguous by adding a *not-an-entity* alternative to each mention's candidate set for linking. The entity linking module's input therefore consists in multiple possible readings of sentences. For each reading, this module must perform entity linking on every possible entity mention by selecting their most probable matching entity. Competing readings are then ranked according to the score of entities (or sequence of entities) ranked first in each of them. The reading with no entity should also receive a score in order to be included in the ranking. The motivation for this joint task lies in the frequent necessity of accessing contextual and referential information in order to complete an accurate named entity recognition; thus the part where named entity recognition usually resolves a number of ambiguities is left for the entity linking module, which uses contextual and referential information about entities.

We have realized a first implementation of our system, as well as experiments and evaluation results. In particular, when using knowledge about entities to perform entity linking, we discuss the usefulness of domain specific knowledge and the problem of domain adaptation.

In 2012, improvements have been made to Nomos by combining the NP named entity detection module with LIANE, a probabilistic system developed by Frédéric Béchet (LIF) in order to better predict possible false matches. The linking step has also been enriched with the use of a more complete and autonomous knowledge base derived from Wikipedia, as well as new parameters and ranking functions for the prediction of the mention/entity alignment.

In the context of this linking task for the processing of AFP corpora and content enrichment with metadata, we conducted a deep study of Semantic Web recent developments and especially of the Linked Data initiatives in order to consider the integration of AFP metadata in these knowledge representation frameworks. On this topic as well as the enlarged view of entity linking for semantic annotation of textual content, discussions have taken place with Eric Charton (CRIM, Montréal, Canada) during 2012 Fall.

The Nomos system as well as the general process of content enrichment with metadata and reference base population has been presented at a dedicated workshop at NAACL in June 2012 (AKBC-WEKEX 2012).

6.8. Advances in lexical semantics

Participants: Benoît Sagot [correspondant], Marion Richard, Sarah Beniamine.

In 2012, several contributions to the WOLF have been finalized and/or published. In particular, various successful attempts to enhance the coverage of the WOLF have been integrated within the master resource [23], [19], [31], [24]. A more original work has also been achieved, targeted at improving the precision of the resource by automatically detecting probable outliers [32]. This latter work has been integrated within the dedicated sloWTool platform, and these outliers partly validated by Slovene students of Romance studies. In parallel, a medium-scale manual validation effort has been achieved at Alpage thanks to the work of two Master students funded by the ANR EDyLex project, which has led to the validation of a vast majority of so-called "basic" synsets, i.e., what can be expected to be the most useful part of the resource.

The result of all this work has been integrated in a preliminary first non-alpha version of the WOLF, version WOLF 1.0b.

6.9. Techniques for transferring lexical resources from one language to a closely-related one

Participants: Yves Scherrer, Benoît Sagot.

Developing lexical resources is a costly activity, which means that large resources only exist for a small number of languages. In our work, we address this issue by transferring linguistic annotations from a language with large resources to a closely related language which lacks such resources. This research activity, funded by the Labex EFL, has started in October 2012.

First results include the development of a method to create bilingual dictionaries without any parallel data, depending solely on surface form similarities and their regularities. The resulting bilingual dictionaries are used to transfer part-of-speech annotations from one language to the other. At the moment, our methods are being tested with Wikipedia texts from various languages and dialects closely related to German, such as Dutch and Pfälzisch. We plan to extend this work to data from other language groups and to other types of linguistic annotations, for instance syntactic or semantic resources.

6.10. Modelling the acquisition of linguistic categories by children

Participants: Benoît Crabbé, Luc Boruta, Isabelle Dautriche.

This task breaks in two sub-tasks: acquisition of phonemic categories, and acquisition of syntactic categories.

Although we are only able to distinguish between a finite, small number of sound categories – i.e., a given language's phonemes – no two sounds are actually identical in the messages we receive. Given the pervasiveness of sound-altering processes across languages – and the fact that every language relies on its own set of phonemes – the question of the acquisition of allophonic rules by infants has received a considerable amount of attention in recent decades. How, for example, do English-learning infants discover that the word forms [kæt] and [kat] refer to the same animal species (i.e. *cat*), whereas [kæt] and [bæt] (i.e. *cat*∼*bat*) do not? What kind of cues may they rely on to learn that [sɪŋkɪŋ] and [θɪŋkɪŋ] (*sinking*∼*thinking*) can not refer to the same action? The work presented in this dissertation builds upon the line of computational studies initiated by [90], wherein research efforts have been concentrated on the definition of sound-to-sound dissimilarity measures indicating which sounds are realizations of the same phoneme. We show that solving Peperkamp et al.'s task does not yield a full answer to the problem of the discovery of phonemes, as formal and empirical limitations arise from its pairwise formulation. We proceed to circumvent these limitations, reducing the task of the acquisition of phonemes to a partitioning-clustering problem and using multidimensional scaling to allow for the use of individual phones as the elementary objects. The results of various classification and clustering experiments consistently indicate that effective indicators of allophony are not necessarily effective indicators of phonemehood. Altogether, the computational results we discuss suggest that allophony and phonemehood can only be discovered from acoustic, temporal, distributional, or lexical indicators when—on average—phonemes do not have many allophones in a quantified representation of the input. This subtask has seen the Phd defense of Luc Boruta whose Phd thesis : *"Indicators of allophony and phonemehood"* was successfully defended in September 2012.

As for syntactic categorization, the task is concerned with modelling and implementing psychologically motivated models of language treatment and acquisition. Contrary to classical Natural Language Processing applications, the main aim was not to create engineering solutions to language related tasks, but rather to test and develop psycholinguistic theories. In this context, the study was concerned with the question of learning word categories, such as the categories of Noun and Verb. It is established experimentally that 2-year-old children can identify novel nouns and verbs. It has been suggested that this can be done using distributional cues as well as prosodic cues. While the plain distributional hypothesis had been tested quite extensively, the importance of prosodic cues has not been addressed in a computational simulation. We provided a formulation for modelling this hypothesis using unsupervised and semi-supervised forms of Bayesian learning (EM) both offline and online. This activity started with the master thesis of A. Gutman and has seen this year the start of a new Phd student : I. Dautriche.

6.11. Modelling and extracting discourse structures

Participants: Laurence Danlos, Charlotte Roze.

6.11.1. Lexical semantics of discourse connectives

Discourse connectives are words or phrases that indicate senses holding between two spans of text. The theoretical approaches accounting for these senses, such as text coherence, cohesion, or rhetorical structure theory, share at least one common feature: they acknowledge that many connectives can indicate different senses depending on their context. LEXCONN is a lexical database for French connectives [16].

The French connectives "*en réalité*" and "*en effet*" have been the topic of numerous studies but none of them was formalized. [53] gives a formalization of the conditions the two arguments of these connectives should meet. This formalization is based on factivity information as modeled in the FactBank corpus developed by Roser Sauri.

Sometimes, the sense of connectives is unique but its arguments are hard to determine. In particular, the second argument of an adverbial connective is not always equivalent to its syntactic arguments. This raises problems at the syntax-semantics interface which are described in [52]. The method to handle theses problems in a discursive parser will be studied in the ANR project POLYMNIE, which is headed by Sylvain Podogolla (Inria Lorraine) and which started in October 2012.

6.11.2. Discursive annotation

We plan to annotate the French corpus FTB (French Tree Bank) at the discursive level, in order to obtain the FDTB (French Discourse Tree Bank). The methodology that will be used is close to the one used in the PDTB (Penn Discourse Tree Bank). The first steps of this long term project are presented in [48], [49], [51].

This work is based on a new hierarchy of discourse relations and this new hierarchy was presented at an European workshop organized by the project MULDICO.

6.12. Modelling word order preferences in French

Participants: Juliette Thuilier, Benoît Crabbé, Margaret Grant.

We study the problem of choice in the ordering of French words using statistical models along the lines of [60] and [61]. This work aims at describing and model preferences in syntax, bringing additional elements to Bresnan's thesis, according to which the syntactic competence of human beings can be largely simulated by probabilistic models. We previously investigated the relative position of attributive adjectives with respect to the noun.

This year has seen the Phd thesis defense of Juliette Thuilier in September 2012.

In collaboration with Anne Abeillé (Laboratoire de Linguistique Formelle, Université Paris 7), we extended our corpora study with psycholinguistic questionnaires, in order to show that statistical models are reflecting some linguistic knowledge of French speakers. The preliminary results confirm that animacy is not a relevant factor in ordering French complements.

As regards to corpus work, we are extending the database with spontaneous speech corpora (CORAL-ROM and CORPAIX) and a wider variety of verbal lemmas, in order to enhance sample representativeness and statistical modelling. This activity has lead to the development of an extension of the French Treebank for oral corpora (approx 2000 sentences).

In a cross-linguistic perspective, we plan to strengthen the comparison with the constraints observed in other languages such as English or German with the recruitment of a new postdoc arriving at the beginning of 2013.

As can be seen from the outline above, this line of research brings us closer to cognitive sciences. We hope, in the very long run, that these investigations will bring new insights on the design of probabilistic parsers or generators. In NLP, the closest framework implementing construction grammars is Data Oriented Parsing (DOP).

METISS Project-Team

6. New Results

6.1. Audio and speech content processing

Audio segmentation, speech recognition, motif discovery, audio mining

6.1.1. Audio motif discovery

Participants: Frédéric Bimbot, Laurence Catanese.

This work was performed in close collaboration with Guillaume Gravier from the Texmex project-team.

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio contents.

Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. The developed algorithms allows discovering and collecting occurrences of repeating patterns in the absence of prior acoustic and linguistic knowledge, or training material.

Former work extended the principles of seeded discovery to near duplicate detection and spoken document retrieval from examples [41].

In 2012, the work achieved consisted in consolidating previously obtained results with the motif discovery algorithm and making implementation choices regardless of the structure and the code, in order to minimize the computation time. This has lead to the creation of a software prototype called MODIS.

After the code has been thoroughly optimised, further optimizations to improve the system performances was to change the method used for the search of similarities between patterns. A new functionality has been added to get rid of unrelevant patterns like silence in speech. New versions of dynamic time warping have been implemented, as well as the possibility to downsample the input sequence during the process, which allows a huge gain of computation time.

The Inria/Metiss team has participated to the IRIT P5 evaluation for repetitive musical motifs discovery. The motif discovery software has been adapted to respect the input and output format defined for the task. The run has been made on a evaluation corpus comprised of French radio broadcast from YACAST.

This work has been carried out in the context of the Quaero Project.

6.1.2. Landmark-driven speech recognition

Participant: Stefan Ziegler.

This work is supervised by Guillaume Gravier and Bogdan Ludusan from the Texmex project-team.

Our previous studies indicate that acoustic-phonetic approaches to ASR, while they cannot achieve state-of-the-art ASR performance by themselves, can prevent HMM-based ASR from degrading, by integrating additional knowledge into the decoding.

In our previous framework we inserted knowledge into the decoding by detecting time frames (referred to as landmarks) which estimate the presence of the active broad phonetic class. This enables the use of a modified version of the viterbi decoding that favours states that are coherent with the detected phonetic knowledge[65].

In 2012 we focused on two major issues. First, we aimed at finding new ways to model and detect phonetic landmarks. Our second focus was on the extension of our landmark detector towards a full acoustic-phonetic framework, to model speech by a variety of articulatory features.

Our new approach for the classification and detection of speech units focuses on developping landmark-models that are different from existing frame-based approaches to landmark detection[64]. In our approach, we use segmentation to model any time-variable speech unit by a fixed-dimensional observation vector. After training any desired classifier, we can estimate the presence of a desired speech unit by searching for each time frame the corresponding segment, that provides the maximum classification score.

We used this segment-based landmark-detection inside a standalone acoustic-phonetic framework that models speech as a stream of articulatory features. In this framework we first search for relevant broad phonetic landmarks, before attaching each landmark with the full set of articulatory features.

Integrating these articulatory feature streams into a standard HMM-based speech recognizer by weighted linear combination improves speech recognition up to 1.5

Additionally, we explored the possibilities of using stressed syllables as an information to guide the viterbi decoding. This work was carried under the leaderhip of Bogdan Ludusan from the team TEXMEX at IRISA [56].

6.1.3. Speech-driven functionalities for interactive television

Participants: Grégoire Bachman, Guylaine Le Jan, Nathan Souviraà-Labastie, Frédéric Bimbot.

In the context of the collaborative ReV-TV project, the Metiss research group has contributed to technological solutions for the demonstration of new concepts of interactive television, integrating a variety of modalities (audio/voice, gesture, image, haptic feed-back).

The focus has been to provide algorithmic solutions to some advanced audio processing and speech recognition tasks, in particular : keywords recognition, lip synchronisation for an avatar, voice emotion recognition and interactive vocal control.

The main challenges adressed in the project have been to robustify state-of-the-art based technologies to the diversity of adverse conditions, to provide real-time response and to ensure the smooth integration of the various interactive technologies involved in the project.

The work of the project has resulted in a demonstration which was presented at the Forum Imagina 2012

6.2. Recent results on sparse representations

Sparse approximation, high dimension, scalable algorithms, dictionary design, graph wavelets

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the field of sparse representations, which is at the core of the FET-Open European project (FP7) SMALL (Sparse Models, Algorithms and Learning for Large-Scale Data, see section 8.2.1.1), the ANR project ECHANGE (ECHantillonnage Acoustique Nouvelle GEnération, see section 8.1.1.2), and the ERC project PLEASE (projections, Learning and Sparsity for Efficient Data Processing, see section 8.2.1.2).

6.2.1. A new framework for sparse representations: analysis sparse models

Participants: Rémi Gribonval, Sangnam Nam, Nancy Bertin, Srdjan Kitic.

Main collaboration: Mike Davies, Mehrdad Yaghoobi (Univ. Edinburgh), Michael Elad (The Technion).

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of *few* columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome. Within the SMALL project, we initiated a research programme dedicated to this analysis model, in the context of a generic missing data problem (e.g., compressed sensing, inpainting, source separation, etc.). We obtained a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. We also considered a number of pursuit algorithms for solving the missing data problem, including an L1-based and a new greedy method called GAP (Greedy Analysis Pursuit). Our simulations demonstrated the appeal of the analysis model, and the success of the pursuit techniques presented.

These results have been published in conferences and in a journal paper [42]. Other algorithms based on iterative cosparse projections [83] as well as extensions of GAP to deal with noise and structure in the cosparse representation have been developed, with applications to toy MRI reconstruction problems and acoustic source localization and reconstruction from few measurements [58].

6.2.2. Theoretical results on sparse representations and dictionary learning

Participants: Rémi Gribonval, Sangnam Nam, Nancy Bertin.

Main collaboration: Karin Schnass (EPFL), Mike Davies (University of Edinburgh), Volkan Cevher (EPFL), Simon Foucart (Université Paris 5, Laboratoire Jacques-Louis Lions), Charles Soussen (Centre de recherche en automatique de Nancy (CRAN)), Jérôme Idier (Institut de Recherche en Communications et en Cybernétique de Nantes (IRCCyN)), Cédric Herzet (Equipe-projet FLUMINANCE (Inria - CEMAGREF, Rennes)), Morten Nielsen (Department of Mathematical Sciences [Aalborg]), Gilles Puy, Pierre Vandergheynst, Yves Wiaux (EPFL), Mehrdad Yaghoobi, Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (Inria, Paris)), Boaz Ophir, Michael Elad (Technion), Mark D. Plumbley (Queen Mary, University of London).

Sparse recovery conditions for Orthogonal Least Squares : We pursued our investigation of conditions on an overcomplete dictionary which guarantee that certain ideal sparse decompositions can be recovered by some specific optimization principles / algorithms. We extended Tropp's analysis of Orthogonal Matching Pursuit (OMP) using the Exact Recovery Condition (ERC) to a first exact recovery analysis of Orthogonal Least Squares (OLS). We showed that when ERC is met, OLS is guaranteed to exactly recover the unknown support. Moreover, we provided a closer look at the analysis of both OMP and OLS when ERC is not fulfilled. We showed that there exist dictionaries for which some subsets are never recovered with OMP. This phenomenon, which also appears with ℓ_1 minimization, does not occur for OLS. Finally, numerical experiments based on our theoretical analysis showed that none of the considered algorithms is uniformly better than the other. This work has been submitted for publication in a journal [86]. More recently, we obtained simpler coherence-based conditions [85].

Performance guarantees for compressed sensing with spread spectrum techniques : We advocate a compressed sensing strategy that consists of multiplying the signal of interest by a wide bandwidth modulation before projection onto randomly selected vectors of an orthonormal basis. Firstly, in a digital setting with random modulation, considering a whole class of sensing bases including the Fourier basis, we prove that the technique is universal in the sense that the required number of measurements for accurate recovery is optimal and independent of the sparsity basis. This universality stems from a drastic decrease of coherence between the sparsity and the sensing bases, which for a Fourier sensing basis relates to a spread of the original signal spectrum by the modulation (hence the name "spread spectrum"). The approach is also efficient as sensing matrices with fast matrix multiplication algorithms can be used, in particular in the case of Fourier measurements. Secondly, these results are confirmed by a numerical analysis of the phase transition of the l1-minimization problem. Finally, we show that the spread spectrum technique remains effective in an analog setting with chirp modulation for application to realistic Fourier imaging. We illustrate these findings in the context of radio interferometry and magnetic resonance imaging. This work has been accepted for publication in a journal [45].

Dictionary learning : An important practical problem in sparse modeling is to choose the adequate dictionary to model a class of signals or images of interest. While diverse heuristic techniques have been proposed in the litterature to learn a dictionary from a collection of training samples, there are little existing results which provide an adequate mathematical understanding of the behaviour of these techniques and their ability to recover an ideal dictionary from which the training samples may have been generated.

In 2008, we initiated a pioneering work on this topic, concentrating in particular on the fundamental theoretical question of the identifiability of the learned dictionary. Within the framework of the Ph.D. of Karin Schnass, we developed an analytic approach which was published at the conference ISCCSP 2008 [13] and allowed us to describe "geometric" conditions which guarantee that a (non overcomplete) dictionary is "locally identifiable" by ℓ^1 minimization.

In a second step, we focused on estimating the number of sparse training samples which is typically sufficient to guarantee the identifiability (by ℓ^1 minimization), and obtained the following result, which is somewhat surprising considering that previous studies seemed to require a combinatorial number of training samples to guarantee the identifiability: the local identifiability condition is typically satisfied as soon as the number of training samples is roughly proportional to the ambient signal dimension. The outline of the second result was published in conferences [12], [25]. These results have been published in the journal paper [15].

Analysis Operator Learning for Overcomplete Cosparse Representations : Besides standard dictionary learning, we also considered learning in the context of the cosparse model. We consider the problem of learning a low-dimensional signal model from a collection of training samples. The mainstream approach would be to learn an overcomplete dictionary to provide good approximations of the training samples using sparse synthesis coefficients. This famous sparse model has a less well known counterpart, in analysis form, called the cosparse analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete analysis operator. We consider two approaches to learn an analysis operator from a training corpus.

The first one uses a constrained optimization program based on L1 optimization. We derive a practical learning algorithm, based on projected subgradients, and demonstrate its ability to robustly recover a ground truth analysis operator, provided the training set is of sufficient size. A local optimality condition is derived, providing preliminary theoretical support for the well-posedness of the learning problem under appropriate conditions. Extensions to deal with noisy training samples are currently investigated, and a journal paper is under revision [87].

In the second approach, analysis "atoms" are learned sequentially by identifying directions that are orthogonal to a subset of the training data. We demonstrate the effectiveness of the algorithm in three experiments, treating synthetic data and real images, showing a successful and meaningful recovery of the analysis operator.

Connections between sparse approximation and Bayesian estimation: Penalized least squares regression is often used for signal denoising and inverse problems, and is commonly interpreted in a Bayesian framework as a Maximum A Posteriori (MAP) estimator, the penalty function being the negative logarithm of the prior. For example, the widely used quadratic program (with an ℓ^1 penalty) associated to the LASSO / Basis Pursuit Denoising is very often considered as MAP estimation under a Laplacian prior in the context of additive white Gaussian noise (AWGN) reduction.

A first result, which we published last year, highlights the fact that, while this is *one* possible Bayesian interpretation, there can be other equally acceptable Bayesian interpretations. Therefore, solving a penalized least squares regression problem with penalty $\phi(x)$ need not be interpreted as assuming a prior $C \cdot \exp(-\phi(x))$ and using the MAP estimator. In particular, we showed that for *any* prior P_X, the minimum mean square error (MMSE) estimator is the solution of a penalized least square problem with some penalty $\phi(x)$, which can be interpreted as the MAP estimator with the prior $C \cdot \exp(-\phi(x))$. Vice-versa, for *certain* penalties $\phi(x)$, the solution of the penalized least squares problem is indeed the MMSE estimator, with a certain prior P_X. In general $dP_X(x) \neq C \cdot \exp(-\phi(x))dx$.

A second result, obtained in collaboration with Prof. Mike Davies and Prof. Volkan Cevher (a paper is under revision) characterizes the "compressibility" of various probability distributions with applications to underdetermined linear regression (ULR) problems and sparse modeling. We identified simple characteristics of probability distributions whose independent and identically distributed (iid) realizations are (resp. are not) compressible, i.e., that can be approximated as sparse. We prove that many priors which MAP Bayesian interpretation is sparsity inducing (such as the Laplacian distribution or Generalized Gaussian distributions with exponent p<=1), are in a way inconsistent and do not generate compressible realizations. To show this, we identify non-trivial undersampling regions in ULR settings where the simple least squares solution outperform oracle sparse estimation in data error with high probability when the data is generated from a sparsity inducing prior, such as the Laplacian distribution [39].

6.3. Emerging activities on compressive sensing, learning and inverse problems

Compressive sensing, acoustic wavefields, audio inpainting,

6.3.1. Nearfield acoustic holography (ECHANGE ANR project)

Participants: Rémi Gribonval, Nancy Bertin.

Main collaborations: Albert Cohen (Laboratoire Jacques-Louis Lions, Université Paris 6), Laurent Daudet, Gilles Chardon, François Ollivier, Antoine Peillot (Institut Jean Le Rond d'Alembert, Université Paris 6)

Compressed sensing is a rapidly emerging field which proposes a new approach to sample data far below the Nyquist rate when the sampled data admits a sparse approximation in some appropriate dictionary. The approach is supported by many theoretical results on the identification of sparse representations in overcomplete dictionaries, but many challenges remain open to determine its range of effective applicability. METISS has chosen to focus more specifically on the exploration of Compressed Sensing of Acoustic Wavefields, and we have set up the ANR collaborative project ECHANGE (ECHantillonnage Acoustique Nouvelle GEnération) which began in January 2009. Rémi Gribonval is the coordinator of the project.

In 2010, the activity on ECHANGE has concentrated on Nearfield acoustic holography (NAH), a technique aiming at reconstructing the operational deflection shapes of a vibrating structure, from the near sound field it generates. In this application scenario, the objective is either to improve the quality of the reconstruction (for a given number of sensors), or reduce the number of sensors, or both, by exploiting a sparsity hypothesis which helps regularizing the inverse problem involved.

Contributions of the team in this task spans: notations and model definitions, experimental setting design and implementation, choice of an adapted dictionary in which the sparsity hypothesis holds, improved acquisition strategies through pseudo-random sensor arrays and/or spatial multiplexing of the inputs, experimental study of robustness issues, and theoretical study of potential success guarantees based on the restricted isometry property (which revealed being not verified in our case, despite improved experimental performance).

A paper about robustness issues and spatial multiplexing (an alternative to building antennas with random sensor position) was published in GRETSI last year and as a journal paper this year [38].

6.3.2. Sparse reconstruction for underwater acoustics (ECHANGE ANR project)

Participants: Rémi Gribonval, Nancy Bertin.

Main collaborations: Jacques Marchal, Pierre Cervenka (UPMC Univ Paris 06)

Underwater acoustic imaging is traditionally performed with beamforming: beams are formed at emission to insonify limited angular regions; beams are (synthetically) formed at reception to form the image. We proposed to exploit a natural sparsity prior to perform 3D underwater imaging using a newly built flexible-configuration sonar device. The computational challenges raised by the high-dimensionality of the problem were highlighted, and we described a strategy to overcome them. As a proof of concept, the proposed approach was used on real data acquired with the new sonar to obtain an image of an underwater target. We discussed the merits of the obtained image in comparison with standard beamforming, as well as the main challenges lying ahead, and the bottlenecks that will need to be solved before sparse methods can be fully exploited in the context of underwater compressed 3D sonar imaging. This work has been published in [61] and a journal paper is in preparation.

6.3.3. Audio inpainting (SMALL FET-Open project)

Participants: Rémi Gribonval, Nancy Bertin, Corentin Guichaoua.

Main collaborations: Amir Adler, Michael Elad (Computer Science Department, The Technion, Israel); Maria G. Jafari, Mark D. Plumbley (Centre for Digital Music, Department of Electronic Engineering, Queen Mary University of London, U.K.).

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. It consists in reconstructing a set of missing pixels in an image based on the observation of the remaining pixels. Sparse representations have proved to be particularly appropriate to address this problem. However, inpainting audio data has never been defined as such so far.

METISS has initiated a series of works about audio inpainting, from its definition to methods to address it. This research has begun in the framework of the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL (Sparse Models, Algorithms and Learning for Large-Scale data) which began in January 2009. Rémi Gribonval is the coordinator of the project. The research on audio inpainting has been conducted by Valentin Emiya in 2010 and 2011.

The contributions consist of:

- defining audio inpainting as a general scheme where missing audio data must be estimated: it covers a number of existing audio processing tasks that have been addressed separately so far – click removal, declipping, packet loss concealment, unmasking in time-frequency;
- proposing algorithms based on sparse representations for audio inpainting (based on Matching Pursuit and on l_1 minimization);
- addressing the case of audio declipping (*i.e.* desaturation): thanks to the flexibility of our inpainting algorithms, they can be constrained so as to include the structure of signals due to clipping in the objective to optimize. The resulting performance are significantly improved. This work will appear as a journal paper [33].

Current and future works deal with developping advanced sparse decomposition for audio inpainting, including several forms of structured sparsity (*e.g.* temporal and multichannel joint-sparsity), dictionary learning for inpainting, and several applicative scenarios (declipping, time-frequency inpainting).

6.3.4. Blind Calibration of Compressive Sensing systems

Participants: Rémi Gribonval, Cagdas Bilen.

Main collaborations: Gilles Chardon, Laurent Daudet (Institut Langevin), Gilles Puy (EPFL)

We consider the problem of calibrating a compressed sensing measurement system under the assumption that the decalibration consists in unknown gains on each measure. We focus on blind calibration, using measures performed on a few unknown (but sparse) signals. A naive formulation of this blind calibration problem, using l1 minimization, is reminiscent of blind source separation and dictionary learning, which are known to be highly non-convex and riddled with local minima. In the considered context, we show that in fact this formulation can be exactly expressed as a convex optimization problem, and can be solved using off-the-shelf algorithms. Numerical simulations demonstrate the effectiveness of the approach even for highly uncalibrated measures, when a sufficient number of (unknown, but sparse) calibrating signals is provided. We observe that the success/failure of the approach seems to obey sharp phase transitions. This work has been published at ICASSP 2012 [54], and an extension dealing with the problem of phase-only decalibration, using techniques revolving around low-rank matrix recovery, has been submitted to ICASSP 2013. A journal version is in preparation.

6.3.5. Compressive Gaussian Mixture estimation

Participants: Rémi Gribonval, Anthony Bourrier.

Main collaborations: Gilles Blanchard (University of Potsdam), Patrick Perez (Technicolor R&D, FR)

When fitting a probability model to voluminous data, memory and computational time can become prohibitive. In this paper, we pro- pose a framework aimed at fitting a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical moments of the underlying probability distribution. Deriving a reconstruction algorithm by analogy with compressive sensing, we experimentally show that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. Our algorithm provides good reconstruction and scales to higher dimensions than previous probability mixture estimation algorithms, while consuming less memory in the case of numerous data. It also provides a privacy-preserving data analysis tool, since the sketch does not disclose information about individual datum it is based on. This work has been submitted for publication at ICASSP 2013.

6.3.6. Nearest neighbor search for arbitrary kernels with explicit embeddings

Participants: Rémi Gribonval, Anthony Bourrier.

Main collaborations: Hervé Jégou (TEX-MEX team), Patrick Perez (Technicolor R&D, FR)

Many algorithms have been proposed to handle efficient search in large databases for simple metrics such as the Euclidean distance. However, few approaches apply to more sophisticated Positive Semi-Definite (PSD) kernels. In this document, we propose for such kernels to use the concept of explicit embedding and to cast the search problem into a Euclidean space. We first describe an exact nearest neighbor search technique which relies on bounds on the approximation of the kernel. We show that, in the case of SIFT descriptors, one can retrieve the nearest neighbor with probability 1 by computing only a fraction of the costly kernels between the query and the database vectors. We then propose to combine explicit embedding with a recent Euclidean approximate nearest neighbor search method and show that it leads to significant improvements with respect to the state-of-the-art methods which rely on an implicit embedding. The database vectors being indexed by short codes, the approach is shown to scale to a dataset comprising 200 million vectors on a commodity server. This work has been submitted for journal publication [74]

6.4. Music Content Processing and Music Information Retrieval

Acoustic modeling, non-negative matrix factorisation, music language modeling, music structure

6.4.1. Music language modeling

Participants: Frédéric Bimbot, Dimitris Moreau, Stanisław Raczyński, Emmanuel Vincent.

Main collaboration: S. Fukayama (University of Tokyo, JP)

Music involves several levels of information, from the acoustic signal up to cognitive quantities such as composer style or key, through mid-level quantities such as a musical score or a sequence of chords. The dependencies between mid-level and lower- or higher-level information can be represented through acoustic models and language models, respectively.

We pursued our pioneering work on music language modeling, with a particular focus on the joint modeling of "horizontal" (sequential) and "vertical" (simultaneous) dependencies between notes by log-linear interpolation of the corresponding conditional distributions. We identified the normalization of the resulting distribution as a crucial problem for the performance of the model and proposed an exact solution to this problem [81]. We also applied the log-linear interpolation paradigm to the joint modeling of melody, key and chords, which evolve according to different timelines [80]. In order to synchronize these feature sequences, we explored the use of beat-long templates consisting of several notes as opposed to short time frames containing a fragment of a single note.

The limited availability of multi-feature symbolic music data is currently an issue which prevents the training of the developed models on sufficient amounts of data for the unsupervised probabilistic approach to significantly outperform more conventional approaches based on musicological expertise. We outlined a procedure for the semi-automated collection of large-scale multifeature music corpora by exploiting the wealth of music data available on the web (audio, MIDI, leadsheets, lyrics, etc) together with algorithms for the automatic detection and alignment of matching data. Following this work, we started collecting pointers to data and developing such algorithms.

6.4.2. Music structuring

Participants: Frédéric Bimbot, Gabriel Sargent, Emmanuel Vincent.

External collaboration: Emmanuel Deruty (as an independant consultant)

The structure of a music piece is a concept which is often referred to in various areas of music sciences and technologies, but for which there is no commonly agreed definition. This raises a methodological issue in MIR, when designing and evaluating automatic structure inference algorithms. It also strongly limits the possibility to produce consistent large-scale annotation datasets in a cooperative manner.

This year, our methodology for the *semiotic* annotation of music pieces has developed [72] and concretized into a set of principles, concepts and conventions for locating the boundaries and determining metaphoric labels of music segments [53] [71]. The method relies on a new concept for characterizing the inner organization of music segments called the System & Contrast (S&C) model [73]. At the time of writing this text, the annotation of over 400 music pieces is being finalized and will be released to the MIR scientific community.

In parallel to this work aiming at specifying the task of music structure description, we have designed, implemented and tested new algorithms for segmenting and labeling music into structural units. The segmentation process is formulated as a cost optimization procedure, accounting for two terms : the first one corresponds to the characterization of structural segments by means of the fusion of audio criteria, whereas the second term relies on a regularity constraint on the resulting segmentation. Structural labels are estimated as a probabilistic automaton selection process. A recent development of this work has included the S&C model in the algorithm.

Different systems based on these principles have been tested in the context of the Quaero Project and the MIREX international evaluation campaigns in 2010, 2011 and 2012 (see for instance [66], in 2012).

6.5. Source separation

Source separation, sparse representations, probabilistic model, source localization

6.5.1. A general framework for audio source separation

Participants: Frédéric Bimbot, Rémi Gribonval, Nobutaka Ito, Emmanuel Vincent.

Main collaborations: H. Tachibana (University of Tokyo, JP), N. Ono (National Institute of Informatics, JP)

Source separation is the task of retrieving the source signals underlying a multichannel mixture signal. The state-of-the-art approach consists of representing the signals in the time-frequency domain and estimating the source coefficients by sparse decomposition in that basis. This approach relies on spatial cues, which are often not sufficient to discriminate the sources unambiguously. Recently, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [44], which generalizes a number of existing techniques including our former study on spectral GMMs [34]. This framework makes it possible to quickly design a new model adapted to the data at hand and estimate its parameters via the EM algorithm. As such, it is expected to become the basis for a number of works in the field, including our own.

Since the EM algorithm is sensitive to initialization, we devoted a major part of our work to reducing this sensitivity. One approach is to use some prior knowledge about the source spatial covariance matrices, either via probabilistic priors [75] or via deterministic subspace constraints [76]. The latter approach was the topic of the PhD thesis of Nobutaka Ito who defended this year [30]. A complementary approach is to initialize the parameters in a suitable way using source localization techniques specifically designed for environments involving multiple sources and possibly background noise [37].

6.5.2. Exploiting filter sparsity for source localization and/or separation

Participants: Alexis Benichoux, Emmanuel Vincent, Rémi Gribonval, Frédéric Bimbot.

Main collaboration: Simon Arberet (EPFL)

Estimating the filters associated to room impulse responses between a source and a microphone is a recurrent problem with applications such as source separation, localization and remixing.

We considered the estimation of multiple room impulse responses from the simultaneous recording of several known sources. Existing techniques were restricted to the case where the number of sources is at most equal to the number of sensors. We relaxed this assumption in the case where the sources are known. To this aim, we proposed statistical models of the filters associated with convex log-likelihoods, and we proposed a convex optimization algorithm to solve the inverse problem with the resulting penalties. We provided a comparison between penalties via a set of experiments which shows that our method allows to speed up the recording process with a controlled quality tradeoff. A journal paper including extensive experiments with real data is in preparation.

We also investigated the filter estimation problem in a blind setting, where the source signals are unknown. We proposed an approach for the estimation of sparse filters from a convolutive mixture of sources, exploiting the time-domain sparsity of the mixing filters and the sparsity of the sources in the time-frequency (TF) domain. The proposed approach is based on a wideband formulation of the cross-relation (CR) in the TF domain and on a framework including two steps: (a) a clustering step, to determine the TF points where the CR is valid; (b) a filter estimation step, to recover the set of filters associated with each source. We proposed for the first time a method to blindly perform the clustering step (a) and we showed that the proposed approach based on the wideband CR outperforms the narrowband approach and the GCC-PHAT approach by between 5 dB and 20 dB. This work has been submitted for publication as a journal paper.

On a more theoretical side, we studied the frequency permutation ambiguity traditionnally incurred by blind convolutive source separation methods. We focussed on the filter permutation problem in the absence of scaling, investigating the possible use of the temporal sparsity of the filters as a property enabling permutation correction. The obtained theoretical and experimental results highlight the potential as well as the limits of sparsity as an hypothesis to obtain a well-posed permutation problem. This work has been published in a conference [52] and is accepted for publication as a journal paper, to appear in 2013.

6.5.3. Towards real-world separation and remixing applications

Participants: Nancy Bertin, Frédéric Bimbot, Jules Espiau de Lamaestre, Jérémy Paret, Laurent Simon, Nathan Souviraà-Labastie, Joachim Thiemann, Emmanuel Vincent.

Shoko Araki, Jonathan Le Roux (NTT Communication Science Laboratories, JP)

We participated in the organization of the 2011 Signal Separation Evaluation Campaign (SiSEC) [51], [59]. Following our founding role in the organization of this campaign, we wrote an invited paper summarizing the outcomes of the three first editions of this campaign from 2007 to 2010 [47]. While some challenges remain, this paper highlighted that progress has been made and that audio source separation is closer than ever to successful industrial applications. This is also exemplified by the ongoing i3DMusic project and the recently signed contracts with Canon Research Centre France and MAIA Studio.

In order to exploit our know-how for these real-world applications, we investigated issues such as how to implement our algorithms in real time [60], how to reduce artifacts [40] and how best to exploit extra information or human input. In addition, while the state-of-the-art quality metrics previously developed by METISS remain widely used in the community, we proposed some improvements to the perceptually motivated metrics introduced last year [62].

6.5.4. Source separation for multisource content indexing

Participants: Kamil Adiloğlu, Emmanuel Vincent.

Main collaborations: Jon Barker (University of Sheffield, UK), Mathieu Lagrange (IRCAM, FR), Alexey Ozerov (Technicolor R&D, FR)

Another promising real-world application of source separation concerns information retrieval from multi-source data. Source separation may then be used as a pre-processing stage, such that the characteristics of each source can be separately estimated. The main difficulty is not to amplify errors from the source separation stage through subsequent feature extraction and classification stages. To this aim, we proposed a principled Bayesian approach to the estimation of the uncertainty about the separated source signals [50], [69], [68] and propagated this uncertainty to the features. We then exploited it in the training of the classifier itself, thereby greatly increasing classification accuracy [43].

This work was applied both to singer identification in polyphonic music [55] and to speech and speaker recognition in real-world nonstationary noise environments. In order to motivate further work by the community, we created a new international evaluation campaign on that topic (CHiME) in 2011 and analyzed the outcomes of the first edition [36].

Some work was also devoted to the modeling of similarity between sound events [32].

PAROLE Project-Team

6. New Results

6.1. Speech Analysis and Synthesis

Participants: Anne Bonneau, Vincent Colotte, Dominique Fohr, Yves Laprie, Joseph di Martino, Slim Ouni, Sébastien Demange, Fadoua Bahja, Agnès Piquard-Kipffer, Utpala Musti.

Signal processing, phonetics, health, perception,articulatory models, speech production, learning language, hearing help, speech analysis, acoustic cues, speech synthesis

6.1.1. Acoustic-to-articulatory inversion

6.1.1.1. Annotation of X-ray films and construction of articulatory models

Two databases have been annotated this year: one composed of 15 short sentences representing more than 1000 X-ray images and a second about CVCVs which has already been annotated by hand on sheets of papers. In the latter case we adapted tools of Xarticul software in order to enable a fast processing of these annotations.

Since images of the first database have been digitized from old films there are several spurious jumps and we thus developed tools to remove them during the construction of articulatory models. The big difference with previous databases processed is the presence of more consonants.

The articulatory model is supplemented by a clipping algorithm in order to take into account contacts between tongue and palate.

6.1.1.2. Articulatory copy synthesis

Acoustic features and articulatory gestures have always been studied separately. Articulatory synthesis could offer a nice solution to study both domains simultaneously. We thus explored how X-ray images could be used to synthesize speech. The first step consisted of connecting the 2D geometry given by mediosagittal images of the vocal tract with the acoustic simulation. Last year we thus developed an algorithm to compute the centerline of the vocal tract, i.e. a line which is approximately perpendicular to the wave front. The centerline is then used to segment the vocal tract into elementary tubes whose acoustic equivalents are fed into the acoustic simulation.

The frequency simulation enables the impact of local modifications of the vocal tract geometry to be evaluated easily. This is useful to investigate the contribution of the sagittal to area transformation in the synthetic speech spectrum. However, the sequence of area functions alone does not suffice to synthesize speech since consonants involve very fine temporal details (closure of the vocal tract and then release of the constriction for stops and fricatives for instance) which additionally have to be synchronized with the temporal evolution of the glottis area. Scenarii have thus been designed for VCV sequences and more generally for any consonant clusters. The idea consists of choosing relevant X-ray images near the VCV to be synthesized. These images can be duplicated just before the closure of the vocal tract, modified to simulate the constriction release for a stop...

This procedure has been applied successfully to copy sentences and VCV for four X-ray films of the DOCVACIM database http://www2i.misha.fr/flora/jsp/index.jsp. The next objective will be to develop a complete articulatory synthesis system.

6.1.1.3. Inversion from cepstral coefficients

The two main difficulties of inversion from cepstral coefficients are: (i) the comparison of cepstral vectors from natural speech and cepstral vectors generated by the articulatory synthesizer and (ii) the access to the articulatory codebook.

Last year we developed a bilinear frequency warping optimized to compensate for the articulatory model mismatch. However, the spectral tilt was not taken into account. We thus combined it with affine adaptation of the very first cepstral coefficients in order to take into account the spectral tilt. It turns out that the new adaptation enables a more relevant comparison of cepstral vectors since the geometric precision of the best solution is less than 1mm.

The second difficulty consists of exploring the articulatory codebook efficiently. Indeed, only a small number of hypercuboids could correspond to the input cepstral vector. The issue is to eliminate all cuboids, which cannot give rise to the input cepstral vector. This is easy when using formants as input data since all cuboids can be indexed easily with extreme values of formants. But this becomes impossible with cepstral vectors because the effect of the excitation source cannot be removed completely from cepstral coefficients. We thus use spectral peaks to access the codebook. However, there exist some spurious spectral peaks, and at the same time some peaks can be absent. We thus designed a lax matching between spectral peaks, which enables the comparison of a series of spectral peaks of the original speech with peaks calculated on synthetic speech. This matching algorithm allows the exploration to focus on 5% of the codebook instead of 40% when using only the peak corresponding to F2 is used.

6.1.1.4. Acoustic-to-articulatory inversion using a generative episodic memory

We have developed an episodic based inversion method. Episodic modeling is interesting for two reasons. First, it does not rely on any assumption about the mapping relationship between acoustic and articulatory, but rather it relies on real synchronized acoustic and articulatory data streams. Second, the memory structurally embeds the naturalness of the articulatory dynamics as speech segments (called episodes) instead of single observations as for the codebook based methods. Estimating the unknown articulatory trajectories from a particular acoustic signal, with an episodic memory, consists in finding the sequence of episodes, which acoustically best explains the input acoustic signal. We refer to such a memory as a concatenative memory (C-Mem) as the result is always expressed as a concatenation of episodes. Actually a C-Mem lacks from generalization capabilities as it contains only several examples of a given phoneme and fails to invert an acoustic signal, which is not similar to the ones it contains. However, if we look within each episode we can find local similarities between them. We proposed to take advantage of these local similarities to build a generative episodic memory (G-Mem) by creating inter-episodes transitions. The proposed G-Mem allows switching between episodes during the inversion according to their local similarities. Care is taken when building the G-Mem and specifically when defining the inter-episodes transitions in order to preserve the naturalness of the generated trajectories. Thus, contrary to a C-Mem the G-Mem is able to produce totally unseen trajectories according to the input acoustic signal and thus offers generalization capabilities. The method was implemented and evaluated on the MOCHA corpus, and on a corpus that we recorded using an AG500 articulograph. The results showed the effectiveness of the proposed G-Mem which significantly outperformed standard codebook and C-Mem based approaches. Moreover similar performances to those reported in the literature with recently proposed methods (mainly parametric) were reached.

The paradigm of episodic memories was also used for speech recognition. We do not extend the acoustic feature with any explicit articulatory measurements but instead we used the articulatory-acoustic generative episodic memories (G-mem). The proposed recognizer is made of different memories each specialized for a particular articulator. As all the articulators do not contribute equally to the realization of a particular phoneme, the specialized memories do not perform equally regarding each phoneme. We showed, through phone string recognition experiments that combining the recognition hypotheses resulting from the different articulatory specialized memories leads to significant recognition improvements.

6.1.2. Using Articulography for Speech production

Since we have an articulograph (AG500, Carstens Medizinelektronik) available, we can easily acquire articulatory data required to study speech production. The articulograph is used to record the movement of the tongue (this technique is called electromagnetography - EMA). The AG500 has a very good time resolution (200Hz), which allows capturing all articulatory dynamics. It has also a good precision. In fact, we performed recently an comparative study to assess the precision of the articulograph AG500 in comparison to a concurrent

articulograph NDI Wave. In this study, we found that both systems presented similar results. We showed also that the accuracy is relatively independent of the sensor velocity, but decreases with the distance from magnetic center of the system [31].

To make the best use of the articulograph, we developed an original visualization software, VisArtico, which allows displaying the data acquired by an articulograph. It is possible to display the tongue contour and the lips contour animated simultaneously with acoustics. The software helps to find the midsagittal plane of the speaker and find the palate contour. In addition, VisArtico allows labeling phonetically the articulatory data[30].

We continuousely work on the usage this platform to acquire articulatory data that were used for articulatory-to-acoustic inversion but also to study the co-variation of speech clarity and coarticulatory patterns in Arabic [18]. The results revealed evident relationship between speech clarity and coarticulation: more coarticulation in formal speech and in strong prosodic position.

6.1.3. Speech synthesis

Visual data acquisition was performed simultaneously with acoustic data recording, using an improved version of a low-cost 3D facial data acquisition infrastructure. The system uses two fast monochrome cameras, a PC, and painted markers, and provides a sufficiently fast acquisition rate to enable an efficient temporal tracking of 3D points. The recorded corpus consisted of the 3D positions of 252 markers covering the whole face. The lower part of the face was covered by 70% of all the markers (178 markers), where 52 markers were covering only the lips so as to enable a fine lip modeling. The corpus was made of 319 medium-sized French sentences uttered by a native male speaker and corresponding to about 25 minutes of speech,.

We designed a first version of the text to acoustic-visual speech synthesis based on this corpus. The system uses bimodal diphones (an acoustic component and a visual one) and unit selection techniques (see 3.2.4). We have introduced visual features in the selection step of the TTS process. The result of the selection is the path in the lattice of candidates found in the Viterbi algorithm, which minimizes a weighted linear combination of three costs: the target cost, the acoustic joined cost, and the visual joined cost. Finding the best set of weights is a difficult problem by itself mainly because of their highly different nature (linguistic, acoustic, and visual considerations). To this end, we developed a method to determine automatically the weights applied to each cost, using a series of metrics that assess quantitatively the performance of synthesis.

The visual target cost includes visual and articulatory information. We implemented and evaluated two techniques: (1) Phonetic category modification, where the purpose was to change the current characteristics of some phonemes which were based on phonetic knowledge. The changes modified the target and candidate description for the target cost to better take into account their main characteristics as observed in the audio-visual corpus. The expectation was that their synthesized visual speech component would be more similar to the real visual speech after the changes. (2) Continuous visual target cost, where the visual target cost component is now considered as real value, and thus continuous, based on the articulatory feature statistics. This year, we continued working on improving the quality of the synthesis. This was done by continuously testing new strategies of weight tuning and improving our selection technique [26].

6.1.4. Phonemic discrimination evaluation in language acquisition and in dyslexia and dysphasia

6.1.4.1. Phonemic segmentation in reading and reading-related skills acquisition in dyslexic children and adolescents

Our computerized tool EVALEC was published [56] after the study of reading level and reading related skills of 400 hundred children from grade 1 to grade 4 (from age 6 to age 10) [58]. This research was supported by a grant from the French Ministery of Health (Contrat 17-02-001, 2002-2005). This first compurerized battery of tests in French language assessing reading and related skills (phonemic segmentation, phonological short term memory) comparing results both to chronological age controls and reading level age control in order to diagnostic Dyslexia. Both processing speed and accuracy scores are taken into account. This battery of tests is used by speech and langage therapists. We keep on examining the reliability (group study) and the prevalence (multiple case study) of 15 dyslexics' phonological deficits in reading and reading related skills in comparaison with a hundred reading level children [57], and by the mean of longitudinal studies of children

from age 5 to age 17 [55]. This year, we started the development of a project which examined multimodal speech both with SLI, dyslexics and control children (30 children). Our goal is to examine visual contribution to speech perception accross differents experiments with a natural face (syllables with several conditions). Our goal is to search what can improve intelligibility in children who have sévère langague acquisition difficulties.

6.1.4.2. Langage acquisition and langage disabilities (deaf chidren, dysphasic children)

Providing help for improving French language acquisition for hard of hearing (HOH) children or for children with language disabilities was one of our goal : ADT (Action of Technological Development) Handicom [piquardkipffer:2010:inria-00545856:2]. The originality of this project was to combine psycholinguistical and speech analyses researchs. New ways to learn to speak/read were developed. A collection of three digital books has been written by Agnès Piquard-Kipffer for both 2-6, 5-9, 8-12 year old children (kindergarten, 1-4th grade) to train speaking and reading acquisition regarding their relationship with speech perception and audio-visual speech perception. A web interface has been created (using Symfony and AJAX technologies) in order to create others books for language impaired children. A workflow which transforms a text and an audio source in a video of digital head has been developed. This worklow includes an automatic speech alignment, a phonetic transcription, a speech synthetizer, a French cued speech coding and speaking digital head. A series of studies (simple cases studies, 5 deaf children and 5 SLI children and group studies with 2 kindergarten classes) were proposed to investigate the linguistical, audio-visual processing.... presumed to contribute to language acquisition in deaf children. Publication are submitted.

6.1.5. Enhancement of esophageal voice

6.1.5.1. Detection of F0 in real-time for audio: application to pathological voices

The work first rested on the CATE algorithm developed by Joseph Di Martino and Yves Laprie, in Nancy, 1999.The CATE (Circular Autocorrelation of the Temporal Excitation) algorithm is based on the computation of the autocorrelation of the temporal excitation signal which is extracted from the speech log-spectrum. We tested the performance of the parameters using the Bagshaw database, which is constituted of fifty sentences, pronounced by a male and a female speaker. The reference signal is recorded simultaneously with a microphone and a laryngograph in an acoustically isolated room. These data are used for the calculation of the contour of the pitch reference. When the new optimal parameters from the CATE algorithm were calculated, we carried out statistical tests with the C functions provided by Paul BAGSHAW. The results obtained were very satisfactory and a first publication relative to this work was accepted and presented at the ISIVC 2010 conference. At the same time, we improved the voiced / unvoiced decision by using a clever majority vote algorithm electing the actual F0 index candidate. A second publication describing this new result was published at the ISCIT 2010 conference. Recently we developed a new algorithm based on a wavalet transform applied to the cepstrum excitation. The resuts obtained were satisfactory. This work has been published in the ICMCS 2012 conference [14].

6.1.5.2. Voice conversion techniques applied to pathological voice repair

Voice conversion is a technique that modifies a source speaker's speech to be perceived as if a target speaker had spoken it. One of the most commonly used techniques is the conversion by GMM (Gaussian Mixture Model). This model, proposed by Stylianou, allows for efficient statistical modeling of the acoustic space of a speaker. Let "x" be a sequence of vectors characterizing a spectral sentence pronounced by the source speaker and "y" be a sequence of vectors describing the same sentence pronounced by the target speaker. The goal is to estimate a function F that can transform each source vector as nearest as possible of the corresponding target vector. In the literature, two methods using GMM models have been developed: In the first method (Stylianou), the GMM parameters are determined by minimizing a mean squared distance between the transformed vectors and target vectors. In the second method (Kain), source and target vectors are combined in a single vector "z". Then, the joint distribution parameters of source and target speakers is estimated using the EM optimization technique. Contrary to these two well known techniques, the transform function F, in our laboratory, is statistically computed directly from the data: no needs of EM or LSM techniques are necessary. On the other hand, F is refined by an iterative process. The consequence of this strategy is that the estimation of F is robust and is obtained in a reasonable lapse of time. This interesting result was published and presented at the ISIVC

2010 conference. Recently,we realized that one of the most important problems in speech conversion is the prediction of the excitation. In order to solve this problem we developed a new strategy based on the prediction of the ceptrum excitation pulses. This interesting result has been published in the SIIE 2012 conference [13].

6.1.5.3. Signal reconstruction from short-time Fourier transform magnitude spectra

Joseph Di Martino and Laurent Pierron developed in 2010 an algorithm for real-time signal reconstruction from short-time Fourier magnitude spectra. Such an algorithm has been designed in order to enable voice conversion techniques we are developing in Nancy for pathological voice repair. Recently Mouhcine Chami, an assistant-professor of the INPT institute at Rabat (Morocco) proposed a hardware implementation of this algorithm using FPGAs. This implementation has been publised in the SIIE 2012 conference [17].

6.1.6. Perception and production of prosodic contours in L1 and L2

6.1.6.1. Language learning (feedback on prosody)

A corpus, made up of 8 English sentences and 40 English isolated words has been recorded. Thirty three speakers pronounced the corpus under different conditions : without any audio feedback (first condition), with audio feedback (second condition, experiment realized one week after the first one). In order to test the permanence of the improvement due to feedback, a set of words and all the sentences were then pronounced without feedback (third condition, experiment realized after the second one). An English teacher helped us in the composition of the corpus and recorded it. Parts of this corpus have already been used to test the automatic speech alignment methods developed under the framework of ALLEGRO and implemented in jsnoori (ADT). The feedback will be progressively transferred from Winsnoori to Jsnoori.

6.1.6.2. Production of prosodic contour

The study of French contours (various types of continuations, end of sentences ...) confirmed the existence of patterns which are typical of French prosody. In order to determine the impact of French (the native language) on a second language pronunciation (English), a series of prosodic contours extracted from English sentences uttered by French speakers have been compared to French prosodic countours. To that purpose, French speakers recorded similar sentences in French and in English. Analysis of results is in progress. First results tend to show the impact of the native language ([15] and [10]).

6.1.7. Pitch detection

Over the last two years, we have proposed two new real time pitch detection algorithms (PDAs) based on the circular autocorrelation of the glottal excitation, weighted by temporal functions, derived from the CATE [53] original algorithm (Circular Autocorrelation of the Temporal Excitation), proposed initially by J. Di Martino and Y. Laprie. In fact, this latter algorithm is not constructively real time because it uses a post-processing technique for the Voiced/Unvoiced (V/UV) decision. The first algorithm we developed is the eCATE algorithm (enhanced CATE) that uses a simple V/UV decision less robust than the one proposed later in the eCATE+ algorithm.

We propose a recent modified version called the eCATE++ algorithm which focuses especially on the detection of the F0, the tracking of the pitch and the voicing decision in real time. The objective of the eCATE++ algorithm consists in providing low classification errors in order to obtain a perfect alignment with the pitch contours extracted from the Bagshaw database by using robust voicing decision methods. The main improvement obtained in this study concerns the voicing decision, and we show that we reach good results for the two corpora of the Bagshaw database. This algorithm is under a submission process in an international journal.

6.2. Automatic Speech Recognition

Participants: Sébastien Demange, Dominique Fohr, Christian Gillot, Jean-Paul Haton, Irina Illina, Denis Jouvet, Odile Mella, Luiza Orosanu, Othman Lachhab.

telecommunications, stochastic models, acoustic models, language models, automatic speech recognition, training, robustness

6.2.1. Core recognition

6.2.1.1. Broadcast News Transcription

A complete speech transcription system, named ANTS (see section 5.6), was initially developed in the framework of the Technolangue evaluation campaign ESTER for French broadcast news transcription. This year, in the context of the ETAPE evaluation campaign about transcription of radio and TV debates, the speech transcription system was improved. Large amounts of text data have been collected over the web. This new collected web data, plus new text and speech resources have made possible the creation and training of new acoustic models and new language models. Moreover new processing steps have been included in the transcription system, leading to much better performance than with the initial system. Several system variants have been developed, and for the ETAPE evaluation campaign, their results have been combined.

Extensions of the ANTS system have been studied, including the possibility to use the sphinx recognizers, and unsupervised adaptation processes. Training scripts for building acoustic models for the Sphinx recognizers are now available and take benefit of parallel computations on the computer cluster for a rapid optimization of the model parameters The Sphinx models are also used for speech/text alignment on both French and English speech data. A new speech transcription program has been developed for efficient decoding on the computer cluster, and easy modification of the decoding steps (speaker segmentation and clustering, data classification, speech decoding in one or several passes, ...). It handles both the Julius and Sphinx (versions 3 and 4) decoders.

This year, in the context of the ETAPE evaluation campaign, which deals with the transcription of radio and TV shows, mainly debates, the Julius-based and Sphinx-based transcription systems have been improved. Several system variants have been developed (relying on different features, and/or different normalization schemes, different processing steps, and different unsupervised adaptation processes); and, combining the output of the various systems led to significantly improved performance.

The recently proposed approach to grapheme-to-phoneme conversion based on a probabilistic method: Conditional Random Fields (CRF) was investigated further. CRF gives a long term prediction, and assumes a relaxed state independence condition. The proposed system was validated in a speech recognition context. Our approach compared favorably with the performance of the state-of-the-art Joint-Multigram Models (JMM) for the quality of the pronunciations, and it was also shown that combining the pronunciation variants generated by both the CRF-based and the JMM-based apporaches improves performance [21].

Concerning grapheme-to-phoneme conversion, a special attention was paid to infering the pronunciation variants of proper names [34], and the usage of additional information corresponding to the language origin of the proper name was investigated.

6.2.1.2. Non-native speakers

The performance of automatic speech recognition (ASR) systems drastically drops with non native speech. The main aim of non-native enhancement of ASRs is to make available systems tolerant to pronunciation variants by integrating some extra knowledge (dialects, accents or non-native variants).

Our approach is based on acoustic model transformation and pronunciation modeling for multiple non-native accents. For acoustic model transformation, two approaches are evaluated: MAP and model re-estimation. For pronunciation modeling, confusion rules (alternate pronunciations) are automatically extracted from a small non-native speech corpus. We presents [9] a novel approach to introduce confusion rules in the recognition system which are automatically learned through pronunciation modelling. The modified HMM of a foreign spoken language phoneme includes its canonical pronunciation along with all the alternate non-native pronunciations, so that spoken language phonemes pronounced correctly by a non-native speaker could be recognized. We evaluate our approaches on the European project HIWIRE non-native corpus which contains English sentences pronunced by French, Italian, Greek and Spanish speakers. Two cases are studied: the native language of the test speaker is either known or unknown. Our approach gives better recognition results than the classical acoustic adaptation of HMM when the foreign origin of the speaker is known. We obtain 22% WER reduction compared to the reference system.

6.2.1.3. Language Model

Christian Gillot has defended his Ph.D. thesis on the 17th September 2012. In his thesis, he proposes a new approach to estimate the language model probabilities for an automatic speech recognition system. The most commonly used language models in the state of the art are based on n-grams smoothed with Kneser-Ney method. Such models make use of occurrence counts of words sequences up to a maximum length (typically 5 words). These counts are computed on a huge training corpus. Christian's Ph.D. thesis starts by an empirical study of the errors of a state-of-the-art speech recognition system in French, which shows that there are many regular language phenomena that are out of reach of the n-gram models. This thesis thus explores a dual approach of the prevailing statistical paradigm by using memory models that process efficiently specific phenomena, in synergy with the n-gram models which efficiently capture the main trends in the corpus. The notion of similarity between long n-grams is studied in order to identify the relevant contexts to take into account in a first similarity language model. The data extracted from the corpus is combined via a Gaussian kernel to compute a new score. The integration of this non-probabilistic model improves the performance of a recognition system. A second model is then introduced, which is probabilistic and thus allows for a better integration of the similarity approach with the existing models. This second model improves the performance on texts in terms of perplexity. Some future works are further described, where the memory-based paradigm is transposed from the estimation of the n-gram probability up to the language model itself. The principle is to combine individual models together, where each model represents a specific syntactic structure, and also to combine these specific models with a standard n-gram model. The objective is to let specific models compensate for some weaknesses of n-gram models, which cannot capture sparse and rare phenomena, nor patterns that do not occur at all in the the training corpus. This approach hence opens new interesting perspectives in particular for domain adaptation.

6.2.1.4. Speech recognition for interaction in virtual worlds

Automatic speech recognition was investigated for vocal interaction in virtual worlds, in the context of serious games in the EMOSPEECH project. For training the language models, the text dialogs recorded by the TALARIS team (Midiki corpus) on the same serious game (but in a text-based interaction), have been manually corrected and used on addition of available broadcast news corpus. Different language models have then been created using different vocabulary sizes. The acoustic models were adapted from the radio broadcast news models, using state-of-the-art Maximum A Posteriori adaptation algorithm. This reduces the mismatch in recording conditions between the game devices and the original models trained on radio streams. A client-server speech recognition demonstrator has been developed. The client runs on an iPad; it records the speech input, sends it to the server, waits for the speech recognition answer, and finally displays the results. The server runs on a PC, relies on the sphinx4 decoder for decoding the received speech signal, and then sends the results to the iPad client.

6.2.2. Speech recognition modeling

Robustness of speech recognition to multiple sources of speech variability is one of the most difficult challenge that limits the development of speech recognition technologies. We are actively contributing to this area via the development of the following advanced modeling approaches.

6.2.2.1. Detailed modeling

Detailed acoustic modeling was further investigated using automatic classification of speaker data. With such an approach it is possible to go beyond the traditional four class models (male vs female, studio quality vs telephone quality). However, as the amount of training data for each class gets smaller when the number of classes increases, this limits the amount of classes that can efficiently be trained. Hence, we have investigated introducing a classification marging in the classification process. With such a marging, which handle boundary classification uncertainty, speech data at the class-boundary may belong to several classes. This increases the amount of training data in each class, which makes the class acoustic model parameters more reliable, and finally improved the overall recognition performance [22]. Combining maximum likelihood linear regression (MLLR) and maximum a posteriori (MAP) adaptation techniques leads to better speech recognition performance, and makes it possible to use more classes [35].

The approach was later improved by introducing a classification process which relies on phonetic acoustic models and the Kullback Leibler divergence measure to build maximally dissimilar clusters. This approach lead to better recognition results than the likelihood based classification approach used in previous experiments [20].

These class-based speech recognition systems were combined with more traditional gender-based system in the ETAPE campaign for the evaluation of speech transcription systems on French radio and TV shows.

6.2.2.2. Training HMM acouctic models

At the beginning of his second internship at Inria Nancy research laboratory, Othman Lachhab focused on the finalization of a speech recognition system based on context-independent HMMs models, using bigram probabilities for the phonotactic constraints and a model of duration following a normal distribution $\mathcal{N}(\mu, \sigma^2)$ incorporated directly in the Viterbi search process. Currently, he built a reference system for speaker-independent continuous phone recognition using Context- Independent Continuous Density HMM (CI-CDHMM) modeled by Gaussian Mixture Models (GMMs). In this system he developed his own training technique, based on a statistical algorithm estimating the classical optimal parameters. This new training process compares favorably with already published HMM technology on the same test corpus (TIMIT) and has been published in the ICMCS 2012 conference [23].

6.2.3. Speech/text alignment

6.2.3.1. Evaluation of speech/text alignment tools

Speech-text alignment tools are frequently used in speech technology and research: for instance, for training or assessing of speech recognition systems, the extraction of speech units in speech synthesis or in foreign language learning. We designed the software CoALT (Comparing Automatic Labelling Tools) for comparing two automatic labellers or two speech-text alignment tools, ranking them, and displaying statistics about their differences.

The main feature of CoALT is that a user can define its own criteria for evaluating and comparing the speech-text alignment tools since the required quality for labelling depends on the targeted application. Beyond ranking, our tool provides useful statistics for each labeller and above all about their differences and can emphasize the drawbacks and advantages of each labeller. We have applied our software for the French and English languages [19] but it can be used for another language by simply defining the list of the phonetic symbols and optionally a set of phonetic rules.

6.2.3.2. Alignment with non-native speech

Non-native speech alignment with text is one critical step in computer assisted foreign language learning. The alignement is necessary to analyze the learner's utterance, in view of providing some prosody feedback (as for example bad duration of some syllables - too short or too long -). However, non-native speech alignement with text is much more complicated than native speech alignment. This is due to the pronunciation deviations observed on non-native speech, as for example the replacement of some target language phonemes by phonemes of the mother tongue, as well as errors in the pronunciations. Moreover, these pronunciation deviations are strongly speaker dependent (i.e. they depend on the mother tongue of the speaker, and on its fluency in the target foreign lanaguage) which makes their prediction difficult.

However, the first step in automatic computer assisted language learning is to check that the pronunced word or utterance corresponds to the expected sentence, otherwise, if the user has not pronunced the correct words it is useless to proceed further with a detailed analysis of the pronunciation to check for possible misspronunciations. In order to decide if the pronunced utterance corresponds to the expected word or sentence, a force phonetic alignment of the sentence is compared to free decoding of the same sentence. Several comparison features are then defined, such as the number of matching phonemes, the percentage of frames having the save category label, ..., as well as the likelihood ratio. A classifier is then used to decide whether text and speech utterance match or not [36], [28].

These non-native phonetic alignments processes developed in the framework of the ALLEGRO project are currently under implementation in the JSNOORI software, and the processing should be completed by the developpement of automatic feedback procedures.

6.3. Speech-to-Speech Translation and Langage Modeling

Participants: Kamel Smaïli, David Langlois, Sylvain Raybaud, Motaz Saad, Denis Jouvet, Cyrine Nasri.

machine translation, statistical models

Sylvain Raybaud has just defended his thesis untitled "De l'utilisation de mesures de confiance en traduction automatique : évaluation, post-édition et application à la traduction de la parole.". His contributions are the following: study and evaluation of confidence measures for Machine Translation, an original algorithm to automatically build an artificial corpus with errors for training the confidence measures, development of an entire speech-to-text translation system.

In the scope of Confidence Measures, we participated to the World Machine Translation evaluation campaign (WMT2012 http://www.statmt.org/wmt12/quality-estimation-task.html). More precisely, we proposed a Quality Estimation system to the Quality Estimation shared task. The goal was to predict the quality of translations generated by an automatic system. Each translated sentence is given a score between 1 and 5. The score is obtained using several numerical or boolean features calculated according to the source and target sentences. We perform a linear regression of the feature space against scores in the range [1:5]. To this end, we use a Support Vector Machine. We experiment with two kernels: linear and radial basis function. In our system we use the features from the shared task baseline system and our own features (based on the work from the Sylvain Raybaud's thesis). This leads to 66 features. To deal with this large number of features, we propose an in-house feature selection algorithm. Our system came 5th among 19 systems. This work was publish in [24]. In the continuation of this research, we contributed to the development of a Quality Estimation tool (quest: https://github.com/lspecia/quest). For that, David Langlois was invited by Lucia Specia at University of Sheffield, Computer Sciences department, Natural Language Processing group. We added our own features into quest. This tool is dedicated to be available for the research community.

Another objective of our research work, with the Cyrine Nasri's Phd thesis, is to retrieve bilingual phrases for machine translation. As in fact, current statistical machine translation systems usually build an initial word-to-word alignment before learning phrase translation pairs. This operation needs many matching between different single words of both considered languages. We propose a new approach for phrase-based machine translation which does not need any word alignments. It is based on inter-lingual triggers determined by Multivariate Mutual Information. This algorithm segments sentences into phrases and finds their alignments simultaneously. Inspite of the youth of this method, experiments showed that the results are competitive but needs some more efforts in order to overcome the one of state-of-the-art methods.

Another aspect of the research of the group is to work on under resourced language related to Arabic. In fact, in several countries through the Arabic world, only few people speak the modern standard Arabic language. People speak something which is inspired from Arabic but could be very different from the modern standard Arabic. This one is reserved for the official broadcast news, official discourses and so on. The study of dialect is more difficult than any other natural language because it should be noted that this language is not written. A preliminary work has been done knowing that our final objective is to propose a machine translation between the different Arabic dialects and modern standrad Arabic. This issue is very difficult and challenging because no corpus does exist, vernaculars are different even within the same country, etc.

Last, Motaz Saad has started his thesis in November 2011. His objective is to work on opinion analysis in multilingual documents from internet. During this year, he retrieved comparable corpus from the web, and proposed a method to align these corpora at document level. He proposed algorithms to measure the degree of comparability between documents. He submitted his work to the International Conference on Corpus Linguistics (CICL2013).

In the framework of the ETAPE evaluation campaign a new machine learning based process was developed to select the most relevant lexicon to be used for the transcription of the speech data (radio and TV shows). The approach relies on a neural network trained to distinguish between words that are relevant for the task and those that are not. After training, the neural network (NN) is applied to each possible word (extracted from a very large text corpus). Then the words that have the largest NN output score are selected for creating the speech recognition lexicon. Such an approach can handle counts of occurences of the words in various data subsets, as well as other complementary informations, and thus offer more perspectives than the traditional unigram-based selection procedures.

SEMAGRAMME Team

6. New Results

6.1. Syntax-Semantics Interface

6.1.1. Graph Rewriting

Bruno Guillaume and Guy Perrier have proposed a system for annotating the French Treebank with semantic dependencies [12], [14]. This system (Synsem_FTB) is based on Graph Rewriting. Graph Rewriting is a framework which is well-suited for syntax-semantic interface because it allows for a modular development of large systems. Each modelled linguistic phenomenon is described by a small set of local rewriting rules. The whole transformation is then described by a sequence of modules to apply successively to the input structure. Another benefit of the Graph Rewriting formalism is that it handles the ambiguity in a natural way with the use of non confluent rewriting systems.

The Synsem_FTB system produces a semantic annotation in the framework of DMRS starting from an annotation with surface syntactic dependencies. It contains 34 modules that can be split in two main parts; the first part produces a deep syntax annotation of the input and the second one rewrites deep syntax to semantics.

With respect to previous works, the system of rewriting rules itself has been improved: it has a larger coverage (causative constructions, rising verbs, ...) and the order between modules has been studied in a more systematic way.

The rewriting calculus has been enriched on two points: the use of rules to make a link with lexicons, especially with the lexicon of verbs Dicovalence, and the introduction of filters to discard inconsistent annotations at some computation steps.

This system has been experimented on the whole French Treebank with the Grew software, which implements the used rewriting calculus.

6.1.2. Passive Sentences

Chris Blom, Philippe de Groote, Yoad Winter, and Joost Zwarts have proposed a unified syntactic-semantic account of passive sentences and sentences with an unspecified object [18]. For both constructions, they use *option types* for introducing implicit arguments into the syntactic-semantic categorial mechanism. They show the advantages of this approach over previous proposals in the domains of scope and unaccusatives. Unlike pure syntactic treatments, option types immediately derive the obligatory narrow scope of existential quantification over an implicit argument's slot. Unlike purely semantic, event-based treatments, their solution naturally accounts for syntactic contrasts between passives and unaccusatives.

6.1.3. Intensionalization

Makoto Kanazawa and Philippe de Groote have defined a general *intensionalization* procedure that turns an extensional semantics for a language into an intensionalized one that is capable of accommodating *truly intensional* lexical items without changing the compositional semantic rules [48]. They have proved some formal properties of this procedure and have clarified its relation to the procedure implicit in Montague's PTQ.

6.1.4. Plural

Sai Qian and Maxime Amblard have modeled the semantics of plurality in continuation semantics [13]. Two types of discourse antecedents formations, inherited from the classical treatment, namely summation and abstraction, are studied in detail. Solutions for each phenomenon are provided respectively by introducing two new functions `Sum` and `Abs`, for obtaining the semantic interpretations.

6.2. Discourse Dynamics

In a joint work with a psycho-linguistist (Michel Musiol, ATILF) and a philosopher (Manuel Rebuschi, Archives Poincaré), are developing a formal analysis of pathological conversations involving schizophrenic speakers [16]. Such conversations give rise to manifest incongruities or ruptures that can be seen as mere contradictions by any "normal" speaker. Our analysis relies both on semantic and pragmatic features of conversation. We propose a SDRT-inspired [20] account of pathological conversations, and we apply it to two relevant excerpts. We conclude with a short discussion about the localization of inconsistencies by schizophrenics, either in semantics or in pragmatics, and its importance for our understanding of thought disorders.

ALICE Project-Team

5. New Results

5.1. A Runtime Cache for Interactive Procedural Modeling

Participant: Sylvain Lefebvre.

This work further explores hashing techniques that we developed over the past years. In particular, we considered modifying our hashing scheme to create a run-time cache. The cache avoids expensive computations when texturing implicit surfaces with complex procedural functions. This is a result from a collaboration with the Karlsruhe Institute of Technology which was funded by an Inria COLOR grant and has been published this year in the journal "Computers & Graphics" [14].

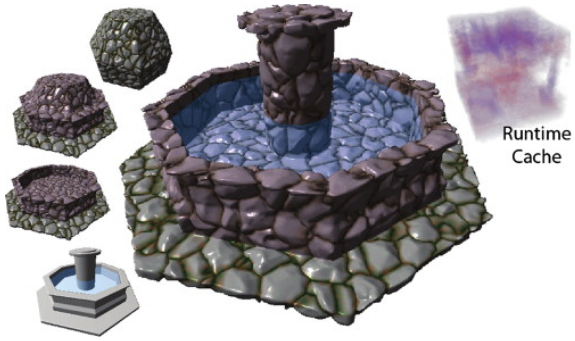


Figure 1. A Runtime Cache for Interactive Procedural Modeling.

5.2. Texture Synthesis

Participants: Sylvain Lefebvre, Bruno Jobard.

We continued investigating on Gabor Noise and considered fitting the parameters of our Gabor noise texturing technique from example images. This required a new formulation of our noise, allowing us to solve the problem as a basis pursuit denoising optimization. This is the result of a collaboration with the team REVES / Inria Sophia-Antipolis, the Katholieke Universiteit of Leuven and Université Paris Descartes. This work has been presented at the SIGGRAPH conference this year [8].

We also revisited techniques for texture synthesis explicitly copying and assembling large patches of an example image to form a new texture. We accelerate this process through a parallel implementation which both optimizes for the shape of the patches and a deformation along their boundary to better match edges. This work is part of the PhD thesis of Anass Lasram and has been presented this year at the Eurographics/ ACM SIGGRAPH Symposium on High Performance Graphics, [19].

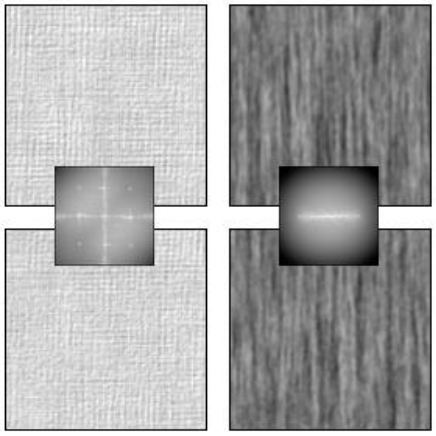


Figure 2. Gabor Noise by Example.

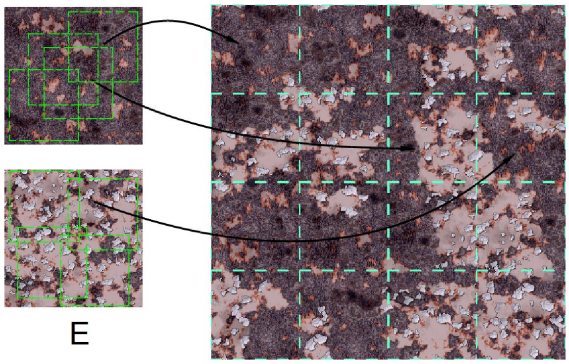


Figure 3. Parallel patch–based texture synthesis.

We also studied ways of helping the user to select the parameters of procedural texture generators, by proposing two contributions :

- We studied how to summarize the appearances generated by complex procedural textures in a small preview image. The challenge is to capture the large variety of appearances despite a limited pixel space. We formulate the problem as a layout of high-dimensional samples in a regular grid, and optimize for it through a modified Self Organizing Map algorithm. This work is part of the PhD thesis of Anass Lasram, and is a collaboration with our industrial partner Allegorithmic. This work has been published this year in the journal "Computer Graphics Forum", [10].
- The parameters of complex procedural textures are typically chosen through a slider-based interface. We augment this interface with preview images which predict how the texture will change when manipulating the slider. This greatly simplifies the process of choosing parameters for these textures. This work is part of the PhD thesis of Anass Lasram, and is a collaboration with our industrial partner Allegorithmic. This work has been published this year as EUROGRAPHICS short paper, [18].

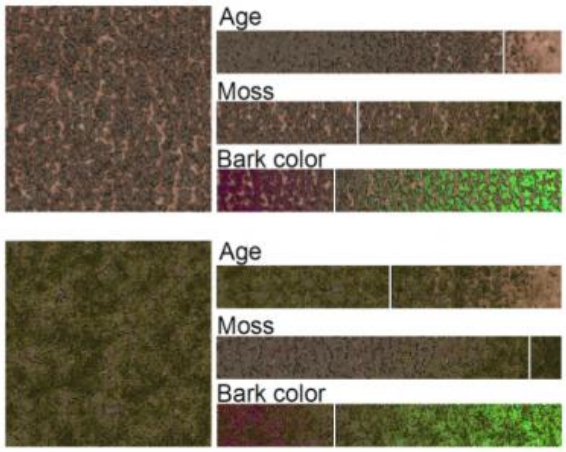


Figure 4. Scented Sliders for Procedural Textures.

5.3. Algorithms and analysis

Participants: Laurent Alonso, Samuel Hornus.

Data structure for fast witness complexes: Samuel Hornus is currently pursuing work started while a post-doc in Sophia Antipolis, on data structure for the fast construction of witness complexes; these are sub complexes of Delaunay triangulations that can be faster to compute for low dimensional data embedded in high dimensional ambiant space.

Analysis of Boyer and Moore's MJRTY Algorithm: Given a set of n elements each of which is either red or blue, Boyer and Moore's algorithm uses pairwise equal/not equal color comparisons to determine the majority color. We analyze the average behavior of their algorithm, proving that if all 2^n possible inputs are equally likely, the average number of color comparisons used is $n - \sqrt{2n/\pi} + O(1)$ and have variance in $\frac{\pi-2}{\pi}n - \frac{\sqrt{2\,n}}{\sqrt{\pi}} + O(1)$. This work has been submitted to SIAM Journal On Computing.

5.4. Visualizing $2D$ Flows with Animated Arrow Plots

Participants: Bruno Jobard, Nicolas Ray, Dmitry Sokolov.

Flow fields are often represented as a set of static arrows in illustration of scientific vulgarization, documentary, meteorology, etc. This simple and schematic representation lets an observer intuitively interpret the main properties of a flow: its orientation and velocity magnitude (Figure 5).

We have investigated how to automatically generate dynamic versions of such representations for $2D$ unsteady flow fields. As a result, we designed an algorithm able to smoothly animate arrows along the flow while controlling their density in the domain over time. Beside keeping an even distribution of arrows over time, we made significant efforts to remove disturbing rendering artefacts such as the apparition of a new arrow, the removing of existing arrows, and the representation of field where the velocity is null. This work has been published as a research report, [24].

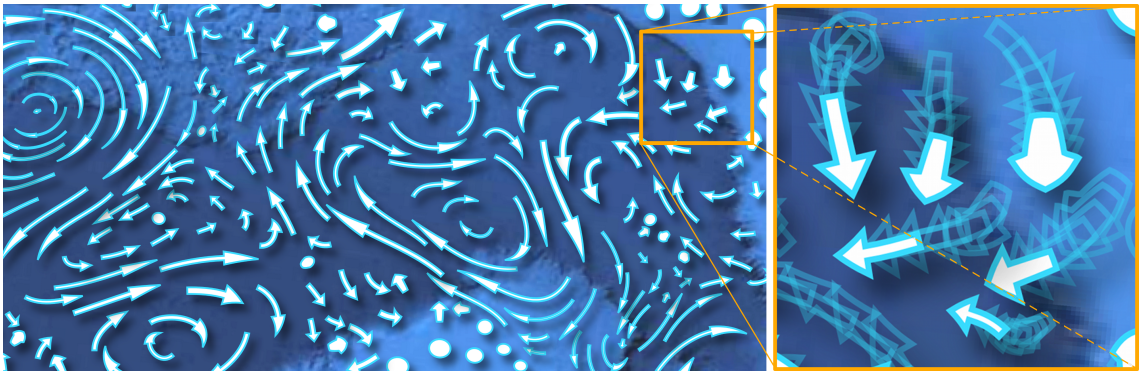


Figure 5. Ocean currents visualized with a set of dynamic arrows. The Close-up shows the arrow trajectories and the morphing of their glyphs.

5.5. Fixing normal constraints for generation of polycubes

Participants: Nicolas Ray, Dmitry Sokolov.

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

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

We have worked on a method able to determine if there exists a deformation of the surface that respects a given orientation constraint on each point. We have also design an automatic solution that can fix constraints that would prevent the existence of a deformation into a polycube (Figure 6).

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

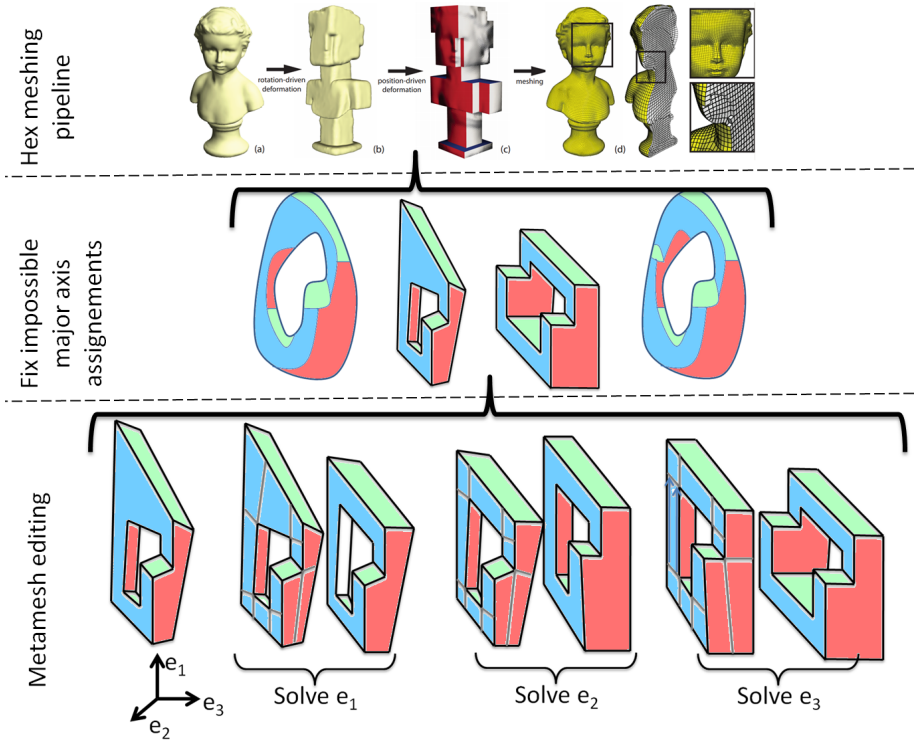


Figure 6. ***Upper row:*** *the surface is deformed to make its normals closer to major axis, but to reach an equality, we need to have a coherent "wished orientation" of the faces.* ***Middle row:*** *we define a valid deformation into a polycube by editing the "wished orientation".* ***Lower row*** *the resolution is performed a dimension at a time.*

5.6. Control of the differential behaviour of the joining curve between two fractal curves

Participants: Dmitry Sokolov, S. Podkorytov, C. Gentil, S. Lanquetin.

The general objective of our work is to create a geometric modeller based on iterative processes. Iterative processes can be used to describe a wide array of shapes inaccessible to standard methods such as fractal curves or sets. Our work is based on Boundary Controlled Iterative System (BCIFS). BCIFS upgrades the standard iterative process such as Iterated Function System (IFS) with B-Rep structure. We can describe objects with familiar B-rep structure, where each cell is a fractal object. For instance, if we consider a polyhedron, then each face is a fractal surface, and each edge is a fractal curve. Objects modelled with BCIFS not necessary have the fractal properties, objects such as B-splines curves and surfaces can be modelled as well. So with BCIFS formalism we can operate with both standard and fractal objects.

With this objective in mind, we have to provide tools that work with fractal objects in the same manner as with objects of classical topology. In this project we focus on the constructing of an intermediate curve between two other curves defined by different iterative construction processes. Similar problem often arises with subdivision surfaces, when the goal is to connect two surfaces with different subdivision masks. We start by dealing with curves, willing to later generalize our approach to surfaces. We formalise the problem with Boundary Controlled Iterated Function System model. Then we deduct the conditions that guaranties continuity of the intermediate curve. These conditions determine the structure of subdivision matrices. By studying the eigenvalues of the subdivision operators, we characterise the differential behaviour at the connection points between the curves and the intermediate one. This behaviour depends on the nature of the initial curves and coefficients of the subdivision matrices. We also suggest a method to control the differential behaviour by adding intermediate control points (Figure 7). This work was presented at the Symposium on Solid and Physical Modeling [23].

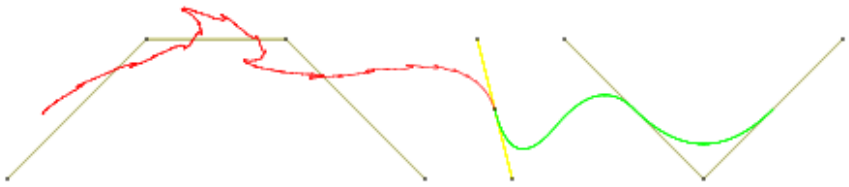


Figure 7. Two intermediate curves between the fractal curve and B-spline. Three control point are used to control the shape of the curve

5.7. Approximate convex hull of affine iterated function system attractors

Participants: Dmitry Sokolov, A. Mishkinis, C. Gentil, S. Lanquetin.

In this paper, we present an algorithm to construct an approximate convex hull of the attractors of an affine iterated function system (IFS). We construct a sequence of convex hull approximations for any required precision using the self-similarity property of the attractor in order to optimize calculations. Due to the affine properties of IFS transformations, the number of points considered in the construction is reduced. The time complexity of our algorithm is a *linear* function of the number of iterations and the number of points in the output convex hull. The number of iterations and the execution time increases logarithmically with increasing accuracy. In addition, we introduce a method to simplify the approximation of the convex hull without loss of accuracy. Figure 8 gives and illustration. This work was published at the Chaos, Solitons & Fractals journal [12].

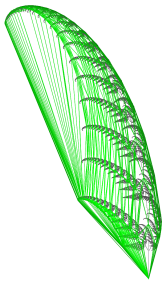


Figure 8. Approximate convex hull for a 3D IFS attractor.

5.8. Shift-Based Parallel Image Compositing on InfiniBand Fat-Trees

Participant: Xavier Cavin.

In this work, we propose a new parallel image compositing algorithm, called Shift-Based, relying on a well-known communication pattern called shift permutation. Indeed, shift permutation is one of the possible ways to get the maximum cross bisectional bandwidth provided by an InfiniBand fat-tree cluster. We show that our Shift-Based algorithm scales on any number of processing nodes (with peak performance on specific counts), allows overlapping communications with computations and exhibits contention free network communications. This is demonstrated with the image compositing of very high resolution images at interactive frame rates. This work is a collaboration with the SED service of Inria (Olivier Demengeon). It has been presented this year at the Eurographics Symposium on Parallel Graphics and Visualization, [17].

5.9. Multi view data processing

Participants: Rhaleb Zayer, Alejandro Galindo, Kun Liu.

Direct use of denoising and mesh reconstruction algorithms on point clouds originating from multi-view images is often oblivious to the reprojection error. This can be a severe limitation in applications which require accurate point tracking, e.g., metrology. we propose a method for improving the quality of such data without forfeiting the original matches. We formulate the problem as a robust smoothness cost function constrained by a bounded reprojection error. The arising optimization problem is addressed as a sequence of unconstrained optimization problems by virtue of the barrier method. Experimental results on synthetic and acquired data compare our approach to alternative techniques. This work has been presented this year at the 8th International Symposium on Visual Computing, [20].

5.10. Deformation modeling of slender objects

Participants: Rhaleb Zayer, Alejandro Galindo, Kun Liu.

A desirable property when modeling/editing slender curve-like objects is the ability to emulate the deformation behavior of natural objects (e.g. cables, ropes). Taking such physical considerations into account needs also to abide to editing requirements such as interactivity and full access and control of all degrees of freedom (positional and rotational constraints) during interaction. We regard editing as a static deformation problem but our treatment differs from standard finite element methods in the sense that the interpolation is based on deformation modes rather than the classic shape functions. A careful choice of these modes allows capturing the deformation behavior of the individual curve segments, and devising the underlying mathematical model from simple and tractable physical considerations. In order to correctly handle arbitrary user input (e.g.

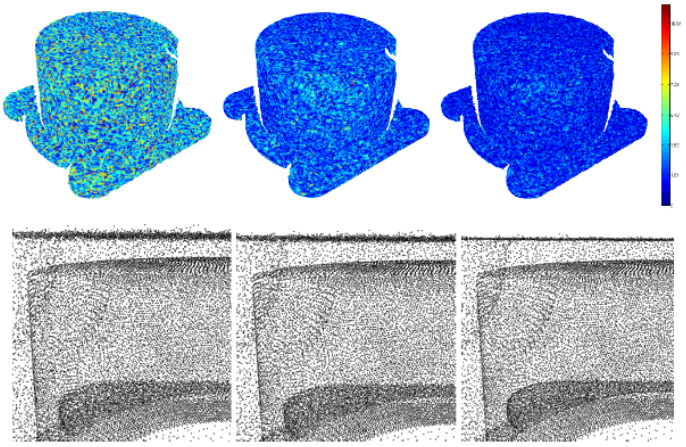


Figure 9. Example of denoising.

dragging vertices in a fast and excessive manner), our approach operates in the nonlinear regime. The arising geometric nonlinearities are addressed effectively through the modal representation without requiring complicated fitting strategies. In this way, we circumvent commonly encountered locking and stability issues while conveying a natural sense of flexibility of the shape at hand. Experiments on various editing scenarios including closed and non-smooth curves demonstrate the robustness of the proposed approach. This work has been published this year in the journal "Computers & Graphics", [15].

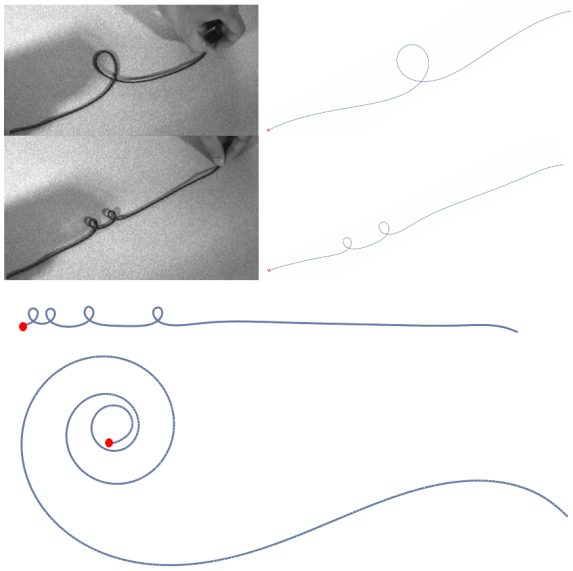


Figure 10. Example of curves.

5.11. Temporally consistent 3D meshing from video data

Participants: Dobrina Boltcheva, Phuong Ho, Bruno Lévy.

This work is a part of the ANR Morpho project (Morpho) which aims at combined analysis of human shapes and motions. In particular, the goal is to study how motions relate to human shapes or how shapes deform in typical motions. During this year, we addressed the first challenge which is building temporally consistent 3D meshes from silhouette images. We have already achieved a very fast meshing algorithm for each frame based on the Centroidal Voronoi Tessellation which has been previously developed in our team. Actually, we are investigating different ways for adding the temporal consistency within our optimisation framework. In particular, we are studding a strategy based on the optimal transport paradigm.

5.12. Re-meshing surfaces

Participants: Nicolas Bonneel, Bruno Lévy, David Lopez, Vincent Nivoliers, DongMing Yan.

In the frame of the ERC GOODSHAPE project, we continued to develop new methods to optimize the sampling of 3D objects. In particular, we studied how to sample a surface with generalized primitives, such as line segments and deformable graphs [11]. We also focused on the problem of remeshing a surface with quads, or fiting a polynomial surface to an input mesh. We proposed a method that minimizes an approximation of the integrated squared distance, based on a restricted Voronoi diagram [22]. Still on the same topic of mesh quadrangulation, we co-published a survey with other international experts of this field [16].

We also worked on anisotropic surface meshing, and developed a technique based on embedding into higher dimensional space and a fast computation of the restricted Voronoi diagram [21].

AVIZ Project-Team

6. New Results

6.1. Tangible Visualization

Participants: Pierre Dragicevic [correspondant], Petra Isenberg, Yvonne Jansen, Jean-Daniel Fekete.

The goal of tangible visualization is to move data and controls to the physical world in order to exploit peoples' natural abilities to perceive and to manipulate objects, and to collaborate through these objects. This is a new topic in information visualization. Physical objects can be used either to represent data (physical visualizations) or to interact with data (physical controls). We studied both.

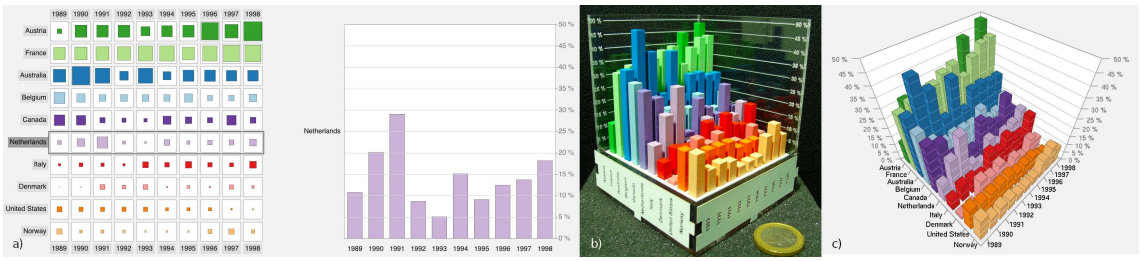


Figure 8. Education expenses of Country by Year shown under three conditions: a) on-screen 2D control; b) on-screen 3D bar chart; c) physical 3D bar chart.

Physical visualizations already exist in the form of data sculptures. Data sculptures are an increasingly popular data-driven media whose purposes are essentially artistic, communicative or educational. We are maintaining a public list of such visualizations at http://www.aviz.fr/Research/PassivePhysicalVisualizations (for passive visualizations) and at http://www.aviz.fr/Research/ActivePhysicalVisualizations (for visualizations with electronics included). But despite prolific work from the art and design communities, physical visualizations have been largely ignored in infovis research. In particular, there is no study on whether physical visualizations can help carry out actual information visualization tasks. We carried out the first infovis study comparing physical to on-screen visualizations. We focused on 3D visualizations, as these are common among physical visualizations but known to be problematic on computers. Taking 3D bar charts as an example (Figure 8), we showed that moving visualizations to the physical world can improve users' efficiency at information retrieval tasks. In contrast, augmenting on-screen visualizations with stereoscopic rendering alone or with prop-based manipulation was of limited help. Our work suggests that the efficiency of physical visualizations stem from features that are unique to physical objects, such as their ability to be touched and their perfect visual realism. These findings provide empirical motivation for current research on fast digital fabrication and self-reconfiguring materials.

We also studied how physical artifacts can help explore and interact with on-screen visualizations. One project consisted in building customizable tangible remote controllers for interacting with visualizations on wall-sized displays [34] (see http://www.aviz.fr/trc). Such controllers are especially suited to visual exploration tasks where users need to move to see details of complex visualizations. In addition, we conducted a controlled user study suggesting that tangibles make it easier for users to focus on the visual display while they interact. Another project explored the concept of stackable tangibles designed to support faceted information seeking in a variety of contexts (see http://www.aviz.fr/stackables). Each Stackable tangible represents search parameters

that can be shared amongst collaborators, modified during an information seeking process, and stored and transferred. Stackables were designed to support collaborative browsing and search in large data spaces. They are useful in meetings, for sharing results from individual search activities, and for realistic datasets including multiple facets with large value ranges.

For more information, see http://www.aviz.fr/phys.

Figure 9. Four stackables. The left shows Stackables with their filter selection interface. The right two show the selected filters.

6.2. EVE : Evolutionary Visual Exploration

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

Evolutionary Visual Exploration (EVE) is a new approach that combines visual analytics with stochastic optimisation to aid the exploration of multidimensional datasets characterised by a large number of possible views or projections. A prototype tool (EvoGraphDice) has been built as an extension of GraphDice, this work has been funded by the System@tics project CSDL, see Figure 10 .

Starting from dimensions whose values are automatically calculated by a PCA, an interactive evolutionary algorithm progressively builds (or evolves) non-trivial viewpoints in the form of linear and non-linear dimension combinations, to help users discover new interesting views and relationships in their data. The criteria for evolving new dimensions is not known a priori and is partially specified by the user via an interactive interface: (i) The user selects views with meaningful or interesting visual patterns and provides a satisfaction score. (ii) The system calibrates a fitness function (optimised by the evolutionary algorithm) to take into account the user input, and then calculates new views. Our method leverages automatic tools to detect interesting visual features and human interpretation to derive meaning, validate the findings and guide the exploration without having to grasp advanced statistical concepts. To validate our method, we conducted an observational study with five domain experts. Our results show that EvoGraphDice can help users quantify qualitative hypotheses and try out different scenarios to dynamically transform their data. Importantly, it allowed our experts to think laterally, better formulate their research questions and build new hypotheses for further investigation.

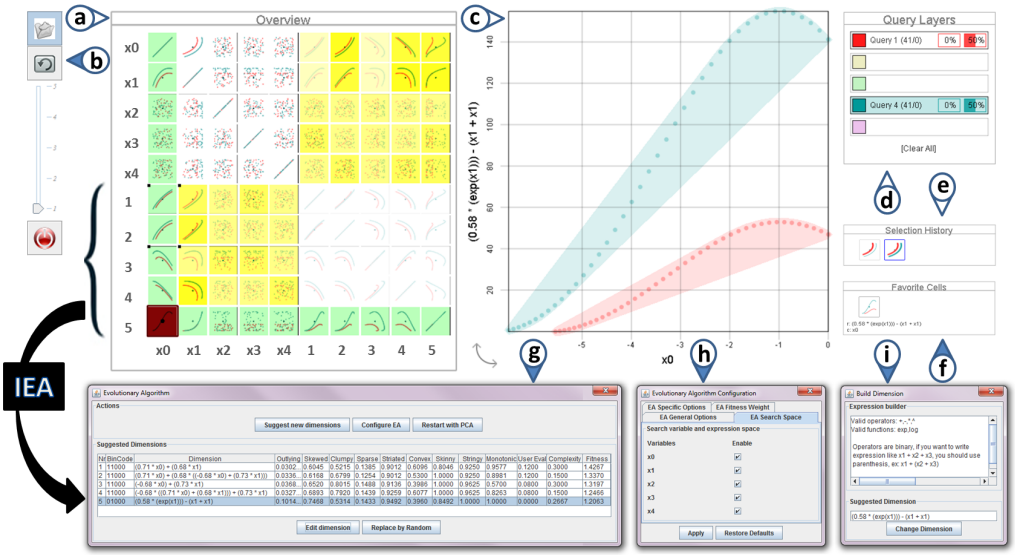


Figure 10. EvoGraphDice prototype showing an exploration session of a synthetic dataset. New extensions to the GraphDice system are indicated by coloured label arrows. Widgets: (a) an overview scatterplot matrix showing the original data set of 5 dimensions (x0..x4) and the new dimensions (1..5) as suggested by the evolutionary algorithm. (b) a tool bar with (top to bottom)"favourite" toggle button, "evolve" button , a slider to evaluate cells and a restart (PCA) button. (c) main plot view. (d) the selection query window. (e) the selection history tool. (f) the favourite cells window. (g) IEA main control window. (h) window to "limit the search space". (i) dimension editor.

6.3. Perception of Visual Variables on Wall-Sized Displays

Participants: Anastasia Bezerianos [correspondant], Petra Isenberg.

We ran two user studies on the perception of visual variables on tiled high-resolution wall-sized displays [9]. We contribute an understanding of, and indicators predicting how, large variations in viewing distances and viewing angles affect the accurate perception of angles, areas, and lengths. Our work, thus, helps visualization researchers with design considerations on how to create effective visualizations for these spaces. The first study showed that perception accuracy was impacted most when viewers were close to the wall but differently for each variable (angle, area , length). Our second study examined the effect of perception when participants could move freely compared to when they had a static viewpoint. We found that a far but static viewpoint was as accurate but less time consuming than one that included free motion. Based on our findings, we recommend encouraging viewers to stand further back from the display when conducting perception estimation tasks. If tasks need to be conducted close to the wall display, important information should be placed directly in front of the viewer or above, and viewers should be provided with an estimation of the distortion effects predicted by our work—or encouraged to physically navigate the wall in specific ways to reduce judgement error. For more information, see http://www.aviz.fr/Research/WallVariables.

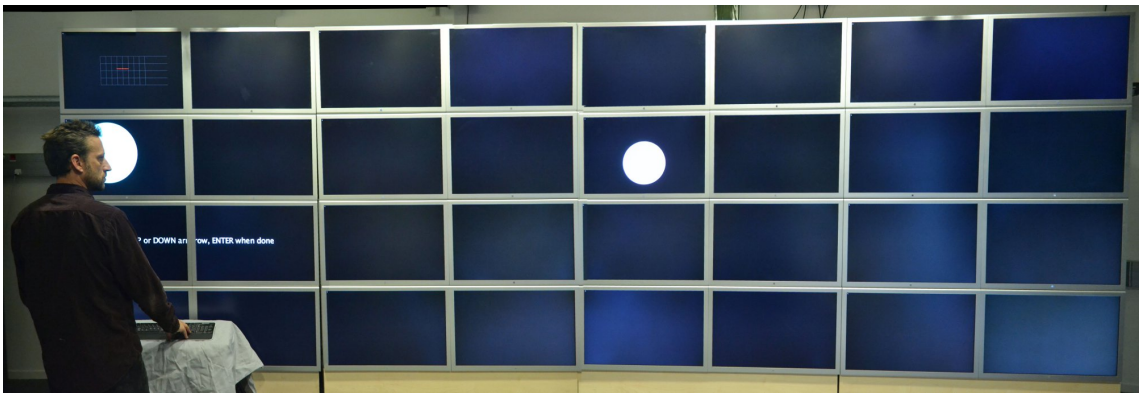


Figure 11. A participant during one trial of a user study on the WILD wall-sized display.

6.4. Sketchyness in Visualization

Participants: Tobias Isenberg [correspondant], Petra Isenberg, Jo Wood, Jason Dykes, Aidan Slingsby, Nadia Boukhelifa, Anastasia Bezerianos, Jean-Daniel Fekete.

AVIZ, in collaboration with City University London, studied how sketchiness can be used, both as a visual style and as a way to represent qualitative uncertainty.

We first studied Handy, an alternative renderer for the Processing graphics environment developed by our collaborators at the City University London [24]. It allows higher-level graphical features such as bar charts, line charts, treemaps and node-link diagrams to be drawn in a sketchy style with a specified degree of sketchiness. Our evaluation concentrated on two core aspects: the perception of sketchiness as a visual variable and higher-level impact of sketchiness on the perception of a whole graphic drawn in this style. Results suggest relative area judgment is compromised by sketchy rendering and that its influence is dependent on the shape being rendered. We showed that degree of sketchiness may be judged on an ordinal scale but that its judgement varies strongly between individuals. We evaluated higher-level impacts of sketchiness through user testing of scenarios that encourage user engagement with data visualization and willingness to critique visualization design. Results suggest that where a visualization is clearly sketchy, engagement may be increased and

that attitudes to participating in visualization annotation are more positive. The results of this work have implications for effective information visualization design that go beyond the traditional role of sketching as a tool for prototyping or its use for an indication of general uncertainty.

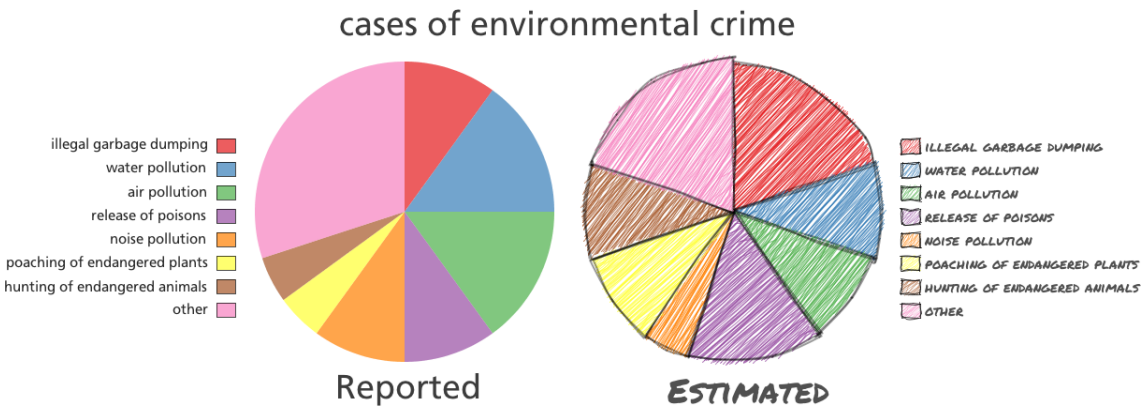


Figure 12. A pie chart drawn in regular and sketchy style.

On this last issue, we have studied whether sketchiness was an effective rendering style for conveying qualitative uncertainty [10]. We compared sketchiness to blur, intensity and dashes and obtained mixed results, showing that sketchiness is not worse than the other visual encodings but that none of them are intuitive and all of them are very limited in range, although still usable for common cases. More work is needed to asses how sketchiness can be best used and to find out more effective encodings for conveying uncertainty in a spontaneous/intuitive way.

For more information, see http://www.aviz.fr/Research/SketchyRendering and http://www.aviz.fr/Research/UncertaintySketchy.

6.5. Supporting Judgment and Decision Making with Visualizations

Participants: Pierre Dragicevic [correspondant], Luana Micallef, Jean-Daniel Fekete.

People have difficulty understanding statistical information and are unaware of their wrong judgments. Cognitive biases abound, particularly in Bayesian reasoning (see http://youtu.be/D8VZqxcu0I0 for a classic example). Psychology studies suggest that the way Bayesian problems are represented can impact comprehension, but few visual designs have been evaluated and only populations with a specific background have been involved. We conducted a study where a textual and six visual representations for three classic problems were compared using a diverse subject pool through crowdsourcing []. Visualizations included area-proportional Euler diagrams, glyph representations, and hybrid diagrams combining both. Our findings were inconsistent with previous studies in that subjects' accuracy was remarkably low and did not significantly improve when a visualization was provided with the text. A follow-up experiment confirmed that simply adding a visualization to a textual Bayesian problem is of little help for crowdsource workers. It however revealed that communicating statistical information with a diagram, giving no numbers and using text to merely set the scene significantly reduces probability estimation errors. Thus, novel representations that holistically combine text and visualizations and that promote the use of estimation rather than calculation need to be investigated. We also argued for the need to carry out more studies in settings that better capture real-life rapid decision making than laboratories. We proposed the use of crowdsourcing to partly address this concern, as crowdsourcing captures a more diverse and less intensely focused population than university students. Doing so, we hope that appropriate representations that facilitate reasoning for both laymen and professionals, independent of

their background, knowledge, abilities and age will be identified. By effectively communicating statistical and probabilistic information, physicians will interpret diagnostic results more adequately, patients will take more informed decisions when choosing medical treatments, and juries will convict criminals and acquit innocent defendants more reliably.

For more information, see http://www.aviz.fr/bayes.

IMAGINE Team

6. New Results

6.1. Introduction

We are developing user-centred, knowledge-based models in three main domains: shape, motion and narrative design, leading us to three research axes. The fourth one is the combination of these models with intuitive interaction tools, in order to set up interactive creative environments dedicated to specific categories of content. The following sections describe our activities in 2012 for each axis.

6.2. High level model for shapes

Scientist in charge: Stefanie Hahmann

Other permanent researchers: Marie-Paule Cani, Jean-Claude Léon, Damien Rohmer.

6.2.1. Implicit surface modeling

Participants: Adrien Bernhardt, Marie-Paule Cani, Maxime Quiblier, Cédric Zanni.

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

In collaboration with a researcher in formal computation, Evelyne Hubert, we improved and extended the analytical methods for computing closed form solutions for convolution surfaces [6].

Within Cédric Zanni's PhD we proposed a method based on anisotropic, surface Gabor noise, for generating procedural details on skeleton-based implicit surfaces, see Figure 4 (left). The surfaces enhanced with details can still be smoothly blended, with a natural transition between the details they carry [19].

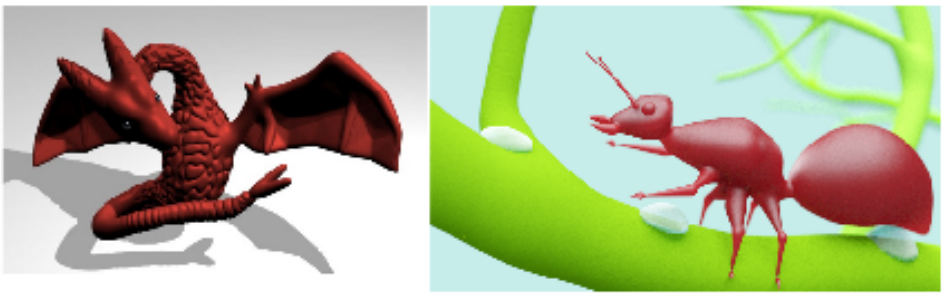


Figure 4. Left: Dragon model showing the variety of details that can be generated. Computation time was less than 2 minutes. Right: Shape obtained by the use of scale-invariant integral surfaces.

We also developed an extension to convolution surfaces, so-called scale-invariant integral surfaces, see Figure 4 (right). Thanks to blending properties that are scale invariant these surfaces have three major advantages: the radius of the surface around a skeleton can be explicitly controlled, shapes generated in blending regions are self-similar regardless of the scale of the model, and thin shape components are not smoothed-out anymore when blended into larger ones. This work has been presented at AFIG2012 [23] and submitted for international publication.

Lastly, in collaboration with Loic Barthe in Toulouse, we contributed to a new blending operator, gradient blending, which enables us to blend implicit shapes not only in function of the field values but also of their gradients. This solves a number long standing problems in implicit modeling: we can generate bulge-free blending, ensure that the topological genius of the blended shape remains the one of the union of the input one, and avoid the blur of small details. A paper has been accepted for publication in ACM ToG [4].

6.2.2. Developable surfaces

Participants: Rémi Brouet, Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer.

Figure 5. Design preserving garment transfer of a multi-layer outfit from a woman to a young girl. Middle: Automatically graded patterns shown to scale. Right: The zoomed-in source and target patterns for the back panel highlight the subtle changes in shape

A developable surface is a surface, which can be unfolded (developed) into a plane without stretching or tearing. Because of this property, developable surfaces lead to a variety of applications in manufacturing with materials that are not amenable to stretching (leather for shoes or hand bags, skins of aircrafts, sails). In computer graphics developable surfaces are very popular to model, simulate or animate clothes or folded papers in virtual environments.

In collaboration with Alla Sheffer (University of British Columbia, Canada visiting Inria) we developed a fully automatic method for design-preserving transfer of garments between characters with different body shapes. The method is able to generate design-preserving versions of existing garments for target characters whose proportions and body shape significantly differ from those of the source. The work has been presented at SIGGRAPH 2012 [1].

Folded paper exhibits very characteristic shapes, due to the presence of sharp folds and to exact isometry with a given planar pattern. In the past we proposed a purely geometric solution to generate static folded paper geometry from a 2D pattern and a 3D placement of its contour curve. Current research focuss on the interactive manipulation of the folded surface without the strong requirement of starting by an initial contour curve, but using sparser positional constraints on the surface.

Damien Rohmer joined in 2012 the Hevea project: this is a project in collaboration between Vincent Borrelli (Institut Camille Jordan, Lyon), Boris Thibert (MGMI, LJK Grenoble) and Francis Lazarus (Gipsa Lab, Grenoble) focussed on the generation and visualisation of the flat torus. The flat torus is a mathematical smooth surface with the topology of a torus but having locally the metric of the plane. In other word, this

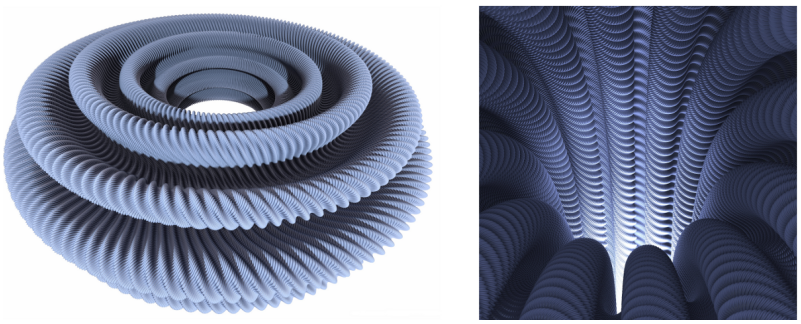


Figure 6. The first representation of a flat torus.

is a developable torus. So far, no representation of such object had ever being made. In 2012, based on a convex integration algorithm generating coherent wrinkles on the torus called *corrugations*, we generated the first representation of such object that is both C^1 while being fractal as the number of wrinkles has to tend to infinity to converge toward true developability. The rendering made by Damien Rohmer has been used for the cover image of Proceedings of the National Academy of Sciences (PNAS) (http://www.pnas.org/content/109/19.cover-expansion).

6.2.3. Parametric surfaces

Participant: Stefanie Hahmann.

We are developing new smooth parametric surface models defined on irregular quad meshes. They are in fact a powerful alternative to singularly parameterized tensor product surfaces since they combine the advantages of both, the arbitrary topology of quad meshes and the smoothness of the tensor product patches.

In collaboration with G.-P. Bonneau (Maverick team) several parametric triangular surface models for arbitrary topologies have been published in the past (CAGD, IEEE TVCG and ACM ToG). A new tensor product spline surface model has been developed this year. It solves the problem of defining a G^1-continuous surface interpolating the vertices of an irregular quad mesh with low degree polynomial tensor product patches. It further aims to produce shapes of very high visual quality while reducing the number of control points. A comparison with existing methods and a journal paper are in preparation.

6.2.4. Fibrous structures

Participant: Damien Rohmer.

Due to anisotropy, fibrous structures may exhibit complex deformation properties. These properties are of main interest to understand the behavior of some human organs such as the heart.

In collaboration with Grant Gullberg, Archontis Giannakidis from Lawrence Berkeley Laboratory, and Alexander Veress from University of Washington we developped a new visualization of heart defects based on the fibrous structure organization. Combining 3D visualization with the fiber structure analysis may help to detect heart defects such as cardiac Hypertrophy. This work as been published as a book chapter [29].

6.2.5. Virtual Prototypes

Participants: Flavien Boussuge, Francois Faure, Stefanie Hahmann, Jean-Claude Léon.

In the context of virtual prototyping (process of product development involving CAD/CAE software), a DMU (digital mock up) is the container of all the components of a 3D virtual product that be used during design and simulations.

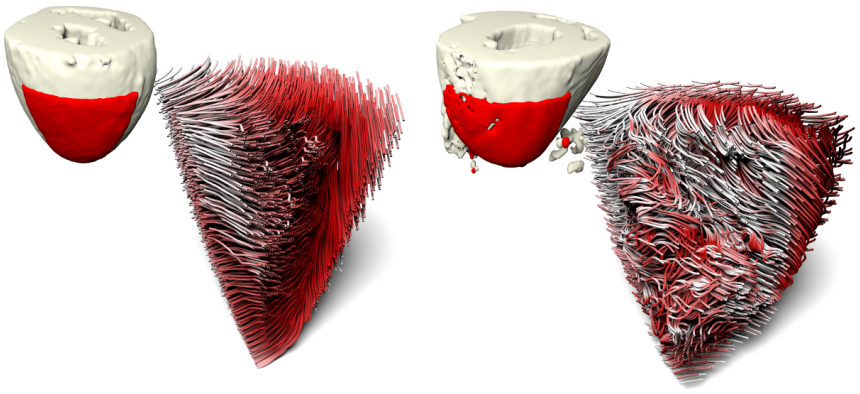


Figure 7. Visualization of fibrous structure in the heart for a normal case (left) and a defect heart (right).

Herein geometric interfaces, i.e. the imprint of a component onto each of its neighboring components, must be taken into account to generate simulation models. Indeed, a DMU does not contain these geometric interfaces. However, extensive use of CAD assemblies has led to increasingly complex DMUs with up to hundreds of thousands of components. The detection and generation of the geometric interfaces between all components with existing software is a very tedious task, which may require hours or days of user-interaction or is even not possible. As part of the ANR project ROMMA in collaboration with Georges-Pierre Bonneau and Francois Jourdes from the Maverick team, we developed a new method to rapidly detect and precisely describe the geometry of interfaces in highly complex assemblies [20].

Within the PhD of Flavien Boussuge, we take advantage of these interfaces to focus on the generation of mixed dimensional models from enriched DMUs for FE analysis of structural assemblies. The goal is to provide a methodology and operators for transforming geometries of complex assemblies so that they are directly usable for FE mesh generation. A first contribution to assembly model preparation for simulation has been presented at ECT12 [11]. Herein, a model preparation methodology has been proposed that addresses the shape transformation categories specific to assemblies. Current and future research includes the generation of construction graphs of volume models that contribute to idealization operators. These algorithms take the simulation objectives into account as part of the proposed methodology.

Another important issue connected to geometry transformation of assemblies and construction graphs of volume models relates to the global as well as partial symmetries of components and assemblies. Here, symmetry analysis is applied to B-Rep NURBS models and must be obtained within the tolerance of a geometric modeler, which differs rather significantly from approximate symmetries extracted from meshes. The symmetry analysis helps structuring the construction graphs of volume models to take into account repetitive locations of primitives. Also, symmetry properties combine with functional annotations of components to enhance their search and retrieval[16].

6.3. Models for real-time motion synthesis

Scientist in charge: François Faure

Other permanent researchers: Marie-Paule Cani, Damien Rohmer, Rémi Ronfard.

6.3.1. Interactive manipulation of folded paper surfaces

Participant: Damien Rohmer.

Although physically-based simulation has become very popular to model deformable surfaces such as cloth it is still not applicable to generate animations of creased paper. Due to the stiffness of this uncompressible material and to the complex changes of its mechanical behavior during creasing. As a result, this standard material in every-day life almost never appears in Computer Graphics applications such as movies or video games. Animating creased paper brings two main challenges: First, such surface needs to be deformed while preserving its length in every direction according to its original pattern. Secondly, sharp features, which are not commonly handled in numerical simulators, need to be generated on the surface.

With the master work of Ulysse Vimont, we developped a prototype (as seen in fig. 8) of a deformation tool enabling to interactively manipulate a virtual sheet of paper. The approach is a procedural approach based on some geometrical apriori knowledge of behavior of paper under deformation. We plan to extend this work in the next year with a new master student Camille Shreck.

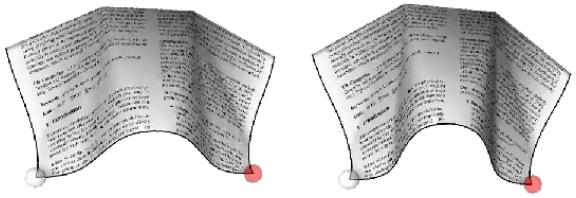


Figure 8. Example of interactive manipulation of a sheet of paper.

6.3.2. Real-time skinning deformation with contacts

Participants: Marie-Paule Cani, Damien Rohmer.

Skinning deformation based on linear blending or dual quaternion approach is a very popular technique thanks to its fast computation. However, they do not capture the complex behavior of flesh bulging and contact between body parts.

In collaboration with Loic Barthe, Rodolphe Vaillant from IRIT Toulouse, and Gael Guennebaud from LaBRI Bordeaux, we developped a skinnning deformation handling both flesh bulges and collision avoidance.

An implicit surface is first fitted onto the original mesh surface. During the animation, the mesh is deformed using a standard skinning deformation while the implicit surface follows the rigid articulation of the bone. Finaly, the mesh is projected back toward the deformed implicit surface enabling to both compensate for mesh collapse and self collision. This work has been presented in AFIG [10] conference and won the *best article* award. It has also being accepted for publication in the REFIG Journal.

6.3.3. Particle-based simulation of concrete structures

Participants: Marie Durand, François Faure.

In collaboration with the LIG and L3S-R labs, we have published results on gpu-accelerated simulation of concrete fracturation due to impacts [2], leading to a speedup factor of about 15 compared to a CPU implementation. This led us to notice that collision detection was the major bottleneck. Consequently, we investigated and published a new incremental sorting method to more efficiently cluster the particles along a Z-curve, by improving the Packed Memory Array data structure for fast updates [15], as illustrated in Figure 9 . We have proposed a new strategy to efficiently update the sorting, while maintaining a desired fill rate in each branch of the tree structure. Experiments show that our PMA can outperform a compact sorted array for up to 50% particle cell changes per time step.

6.3.4. Collision detection and response

Participant: François Faure.

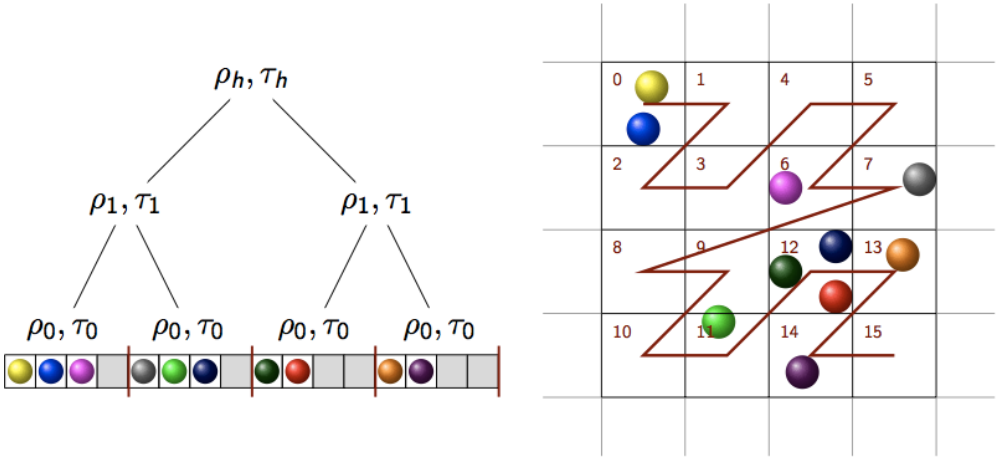


Figure 9. Z-sort using a Packed Memory Array structure. The gaps allow fast updates.

In collaboration with UBC, Vancouver, we have presented at SIGGRAPH 2012 a new method for image-based contact detection and modeling, with guaranteed precision on the intersection volume [8]. Unlike previous image-based methods, our method optimizes a nonuniform ray sampling resolution and allows precise control of the volume error. By cumulatively projecting all mesh edges into a generalized 2D texture, we construct a novel data structure, the Error Bound Polynomial Image (EBPI), which allows efficient computation of the maximum volume error as a function of ray density. Based on a precision criterion, EBPI pixels are subdivided or clustered. The rays are then cast in the projection direction according to the non-uniform resolution. The EBPI data, combined with ray-surface intersection points and normals, is also used to detect transient edges at surface intersections. This allows us to model intersection volumes at arbitrary resolution, while avoiding the geometric computation of mesh intersections. Moreover, the ray casting acceleration data structures can be reused for the generation of high quality images, as illustrated in Figure 10 .

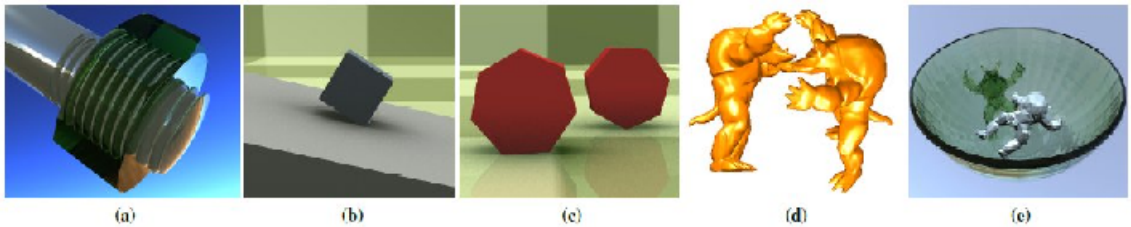


Figure 10. Examples of challenging contact scenarios handled by our method. (a) The movement of a tight fitting nut on a bolt can be simulated directly using the geometric models. (b) Very small geometric features on a flat surface can dramatically change the behavior of objects sliding on it. (c) "Ruina wheels." Two almost identical wheels have obviously different rolling behavior due to subtle features (one is slightly convex and another is slightly concave); our method can simulate this contact behavior realistically. (d) A simulation with 4:4 million triangles. (e) A snapshot of an interactive simulation with ray-traced rendering.

6.3.5. Action representation, segmentation and recognition

Participant: Remi Ronfard.

Following Daniel Weinland's PhD thesis, we published a survey of modern methods for representing, segmenting and recognizing full-body actions in video [32]. A taxonomy of methods is elaborated in that paper, where actions can be represented with local, structured or global features both in time and in space. The potential for future work in grammar-based action recognition is emphasized, with possible applications in corpus-based procedural modeling of actions.

6.3.6. Simulation software architecture

Participants: Ali-Hamadi Dicko, Guillaume Bousquet, Françcois Faure.

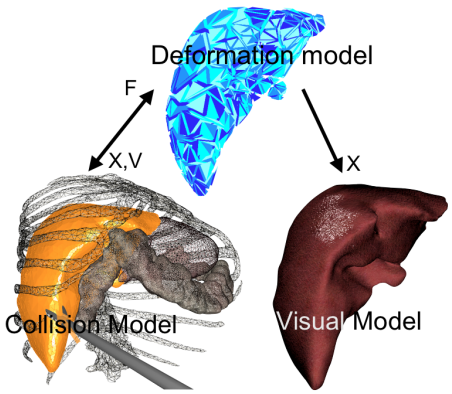


Figure 11. A simulated Liver. Three representations are used: one master model for the internal deformable mechanics, one for the collisions, and one for the visualization. Mappings (black arrows) are used to propagate positions (X) and velocities (V) from master to slaves, while forces (F) are propagated in the opposite direction

We continue the development of SOFA, the open source simulation library, which is becoming an international reference, and we have published a chapter on it in a Springer book [28]. SOFA facilitates collaborations between specialists from various domains, by decomposing complex simulators into components designed independently and organized in a scenegraph data structure. Each component encapsulates one of the aspects of a simulation, such as the degrees of freedom, the forces and constraints, the differential equations, the main loop algorithms, the linear solvers, the collision detection algorithms or the interaction devices. The simulated objects can be represented using several models, each of them optimized for a different task such as the computation of internal forces, collision detection, haptics or visual display, as illustrated in Figure 11 . These models are synchronized during the simulation using a mapping mechanism. CPU and GPU implementations can be transparently combined to exploit the computational power of modern hardware architectures. Thanks to this flexible yet efficient architecture, SOFA can be used as a test-bed to compare models and algorithms, or as a basis for the development of complex, high-performance simulators

6.3.7. Real time fluid animation on GPU

Participant: Martin Guay.

In collaboration with Manuel Vennier (Maverick, Inria), we developped a simple and fast method to animate fluids on the GPU. Inspired from the classical SPH method (Smooth Particles Hydrodynamics), we express a weekly compressible formulation for the fluid animation. Contrary to standard approaches, we fully developed the formulation on a grid, leading to an efficient GPU implementation. The method replace the implicit formulation of pressure by an explicit one based on density invariance. We therefore propose a method to simulate 3D Eulerian gaseous fluids in a single pass on the GPU. The results published in [22] are less accurate than a standard fluid simulation approach, but lead to real-time fluid-looking models (see fig. 12) which are practicable for video games or other interactive applications.

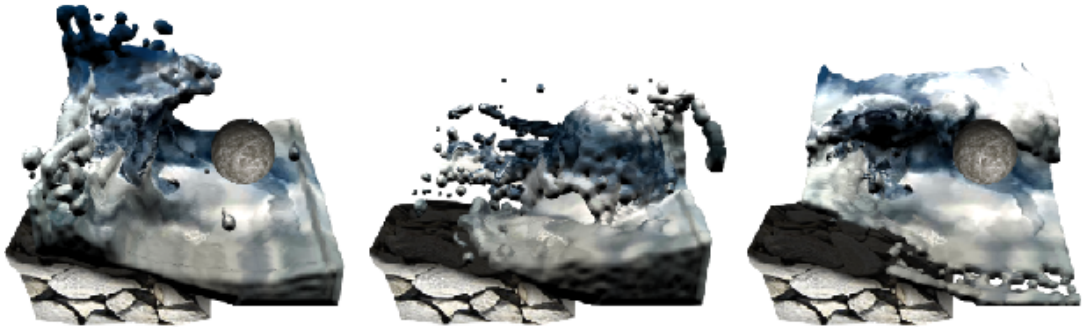


Figure 12. Example of fluid results obtained by our approach in [22].

6.4. Knowledge-based models for narrative design

Scientist in charge: Rémi Ronfard

Other permanent researchers: François Faure, Jean-Claude Léon, Olivier Palombi

6.4.1. Computational model of film editing

Participants: Remi Ronfard, Quentin Galvane.

Collaboration with the Mimetic team (Marc Christie) is continuing on this topic as part of the CINECITA (ANR jeune chercheur) and CHROME (ANR) projects.

We presented a survey of automatic video editing and new results from our ongoing collaboration at the first workshop on intelligent cinematography and editing (WICED) which took place during the Foundation of Digital Games (FDG) international conference [18], [14].

6.4.2. Stochastic Plex Grammars

Participant: Remi Ronfard.

During Quentin Doussot's master thesis, we experimented with stochastic plex grammars, which proved to be efficient for generating 3D scenes in the style of Keith Haring [17]. The model is able to generate static scenes by assembling colorful body parts into Keith Haring figures. The model is also able to simulate Markov chains of such figures by randomly changing attributes and composition of the scene.

6.4.3. Reframing theatre performances

Participants: Remi Ronfard, Vineet Gandhi.

In 2012, we made full-hd video recordings of rehearsals and performances at Celestins - Theatre de Lyon:

- A l'Ouest, directed by Nathalie Fillion. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre-Saintes. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre, Saintes.
- Lorenzaccio, directed by Claudia Stavisky, Théâtre de Saint Petersbourg.
- Mort d'un commis, directed by Claudia Stavisky, Célestins, Théâtre de Lyon.

As part of his PhD thesis, Vineet Gandhi developped novel algorithms for actor detection and naming. This has been tested on movies as well as theatre performances. Current work is focusing on automatically reframing those recordings into cinematically-valid shots focusing on one or more actors.

A related thread of work was started for semantic annotation of the recordings using the syntax and semantics of blocking notations, a symbolic notation used in North-American theatres [25].

6.4.4. *Virtual theatre rehearsals*

Participant: Remi Ronfard.

We are starting to investigate the possibility of rehearsing theatre plays with real and virtual actors, using extensions of interactive scores initially proposed for computer music. A position paper was presented to researchers in theatre studies during a seminar on the notation of theatre [24].

6.4.5. *Extracting functional information from assembly models*

Participants: Jean-Claude Léon, Ahmad Shahwan, Olivier Palombi.

Assembly models of products, as available from CAD software reduce to a set of independent geometric models of its components and a logical structure of the assembly described as a tree containing components' names. Such a model lacks of geometric connections between its components and the work performed at 6.5 contributes already to structure the geometric model of each component with its geometric interfaces. However, the assembly tree structure and components' names still have no connection with the geometric model of components and their names don't convey robust information because their are user chosen. Here, the purpose is to set tight connections between components' geometric models and their functions. Using dualities between geometric interfaces and interaction forces, it is possible to initialize qualitative mechanical values at each geometric interface, producing different possible configurations.

Then, the proposed approach builds upon relationships between function, behavior and shape to derive functional information from the geometry of component interfaces. Among these concepts, the concept of behavior is more difficult to set up and connect to the geometry of interfaces and functions. Indeed, states and design rules are introduced to express the behavior of components through a qualitative reasoning process [7]. This reasoning process, in turn, takes advantage of domain knowledge rules and facts, checking the validity of certain hypotheses that must hold true all along a specific state of the product's lifecycle, such as operational, stand-by or relaxed states. Eliminating configurations at geometric interfaces that contradict one or more of those hypotheses in their corresponding reference state reduces ambiguity, subsequently producing functional information in a bottom-up manner.
This bottom-up process starts with the generation of a Conventional Interfaces Graph (CIG) with components as nodes, and conventional interfaces (CI: the geometric interfaces) as arcs. A CI is initially defined by a geometric interaction that can be a contact or an interference between two components. CIs are then populated with Functional Interpretations (FI) according to their geometric properties, producing potentially many combinations. A first step of the reasoning process, the validation against reference states, reduces the number of FIs per CI. Then, a matching process takes place using inferences of an ontology reasoner to produce a functional designation of each component. The ontology is based on several taxonomies: conventional interfaces, functional interfaces and functional designations that are connected through the qualitative reasoning process. As a result, the geometric model of each component assigned with a functional designation becomes intrinsically structured with functional interfaces (see Figure 13). Structured models can

be used to perform high level shape transformations like virtual prototypes. MyCorporisFabrica is a software framework we started to connect to. This activity is part of the ANR ROMMA project. It is a first contribution to simulation scenarios.

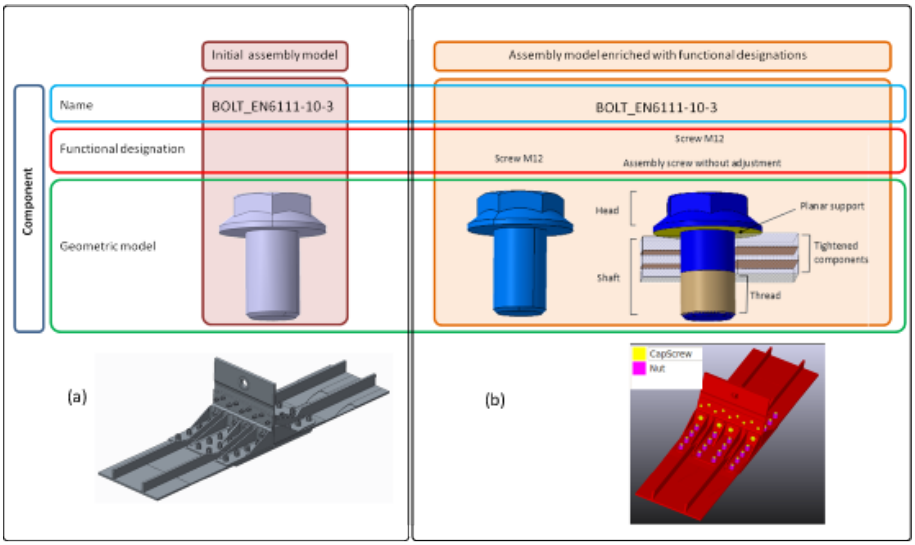


Figure 13. An example of assembly before (a) and after (b) the extraction of functional information. The upper part shows the influence of the extraction process on the structure of components' geometric models. The lower part illustrates the extraction process applied to a mechanical assembly.

6.4.6. Anatomical models

Participants: Ali-Hamadi Dicko, Olivier Palombi, François Faure.

We continue the development and the exploitation of MyCF, our ontology-centered anatomical knowledge base, in collaboration with the Grenoble University Hospital, and the DEMAR team in Montpellier (Benjamin Gilles).

We have presented a novel pipeline for the construction of biomechanical simulations by combining generic anatomical knowledge with specific data [27], [21], as illustrated in Figure 14 . Based on functional descriptors supplied by the user, the list of the involved anatomical entities (currently bones and muscles) is generated using formal knowledge stored in ontologies, as well as a physical model based on reference geometry and mechanical parameters. This simulation-ready model can then be registered to subject-specific geometry to perform customized simulations. The user can provide additional specific geometry, such as a simulation mesh, to assemble with the reference geometry. Subject-specific information can also be used to individualize each functional model. The model can then be visualized and animated. This pipeline dramatically eases the creation of biomechanical models.

6.4.7. Managing morphological and functional information of the human body

Participants: Olivier Palombi, Ali-Hamadi Dicko, François Faure, Jean-Claude Léon, Ahmad Shahwan.

My Corporis Fabrica (MyCF) is an anatomical knowledge database. During 2012, we have linked functional entities defined in MyCF to the involved anatomical structures. The scope has been limited to the musculoskeletal system. Based on this brain new formal description of the functional anatomy of limbs, we present a novel pipeline for the construction of biomechanical simulations by combining generic anatomical knowledge with specific data. This pipeline dramatically eases the creation of biomechanical models [27].

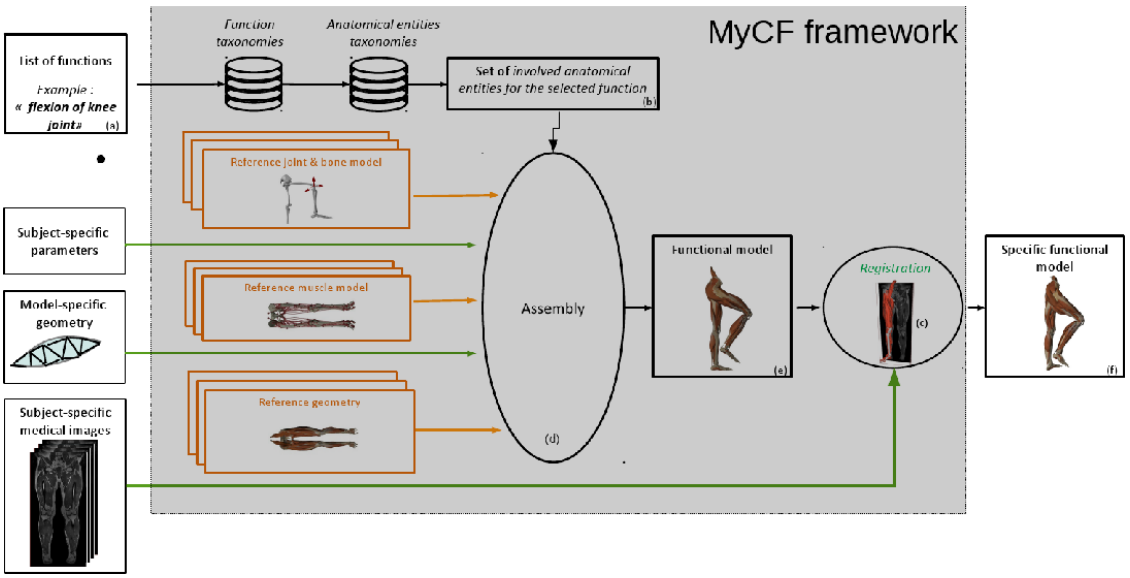


Figure 14. An overview of our modeling framework. On the left, the user input is a list of functions to simulate, optionally complemented with specific data. On the right, the output is a mechanical model ready for simulation. The modeling pipeline uses symbolic knowledge to select anatomical entities to assemble. The final model can be composed of a mix of reference and specific parameters and geometry.

MyCF-Browser which is the GUI of MyCF has been completely reviewed and rewritten. The MyCf's style software architecture is REST (Representational State Transfer) that has emerged as a predominant Web service design model. The anatomical knowledge is now available through a WEB service. The next step is to write a full web MyCF-Browser. MyCF browser is now available on line: http://www.mycorporisfabrica.org/. The MyCf's generic programming framework can be used for other domains. The link with semantic and 3D models matches research activities of IMAGINE towards interactive digital creation media. Anatomy can be seen as a study case.

6.5. Creating and interacting with virtual prototypes

Scientist in charge: Jean-Claude Léon

Other permanent researchers: Marie-Paule Cani, Rémi Ronfard, Olivier Palombi

6.5.1. Space deformations

Participant: Stefanie Hahmann.

Free Form Deformation (FFD) is a well-established technique for deforming arbitrary object shapes in space. Although more recent deformation techniques have been introduced, amongst them skeleton-based deformation and cage based deformation, the simple and versatile nature of FFD is a strong advantage, and justifies its presence in nowadays leading commercial geometric modeling and animation software systems. Several authors have addressed the problem of volume preserving FFD. These previous approaches however make either use of expensive non-linear optimization techniques, or resort to first order approximation suitable only for small-scale deformations. Our approach was to take advantage from the multi-linear nature of the volume constraint in order to derive a simple, exact and explicit solution to the problem of volume preserving FFD. Two variants of the algorithm have been developed, without and with direct shape manipulation.

Moreover, we showed that the linearity of our solution enables to implement it efficiently on GPU. This work has been done in collaboration with Gershon Elber from TECHNION, Hans Hagen from TU Kaiserslautern, Georges-Pierre Bonneau and Sébastien Barbier from Maverick Inria. It has been published in the journal The Visual Computer [5].

Figure 15. Comparison between standard FFD deformation (middle) and our method preserving the volume (right) from an initial rest shape (left).

Within Lucian Stanculescu PhD, we developped a mesh structure that dynamically adapts to the deformation defined by the user. Thanks to the quasi-uniform property of the mesh, it can be locally extended by any arbitrary deformation, and the mesh can also hanlde changes of topologies to be used as a virtual sculpting tool. This year we extend this work to handle local features such as sharp edges. In defining features (points or curves) over the surface we can interactively define meaningfull regions limiting the influence of the deformation tools, or to ease artistic decorration mapping such as textures or extra geometric layers. We aim to generate a new tool enabling to sculpt objects which blend between organic to CAD-style appearance.

6.5.2. Procedural modeling of terrains and cities

Participants: Adrien Bernhard, Marie-Paule Cani, Arnaud Emilien.

Within the PhD of Adrien Bernhard we introduced a real-time terrain modeling tool using a fast GPU-based terrain solver with a lightweight CPU-based data structure.

We then work on adding roads and settlements on this terrain within the PhD of Arnaud Emilien. We focused on the modeling of small, European villages that took benefit of terrain features to settle in safe, sunny or simply convenient places. We introduced a three step procedural method [3] for generating scattered settlements on arbitrary terrains, enabling villages and hamlets, with the associated roads, forests and fields to be built on arbitrary landscapes.

Figure 16. Fortified village at the top of a cliff, using a war-time growth scenario followed by farming style settlement.

6.5.3. Hand Navigator

Participant: Jean-Claude Léon.

The different deformation models we developed in the past few years open the problem of providing intuitive interaction tools for specifying the desired deformations in real-time. Therefore, work has focused on developing new devices to investigate interactions incorporating a rather large number of parameters. For the past three years, we focused on developing a peripheral device similar to a mouse, called the HandNavigator, enabling to control simultaneously ten or more degrees of freedom of a virtual hand. This device developed in collaboration with Jean-Rémy Chardonnet (Inst. Image, Arts et Métiers ParisTech) consists in a 3D mouse for the position and orientation of the hand in 3D space, enhanced with many sensors for moving and monitoring the virtual fingers. Thanks to a pre-industrialization project funded by the incubator GRAVIT, the first prototype, patented by Inria, has been extended with the incorporation of new sensors and new shapes to improve the device efficiency and evolve toward a passive haptic device (see Figure 17). An extension of the patent and a partnership with HAPTION company are new steps toward the industrialization of this device. The partnership with HAPTION focuses on grasping actions to use the Hand Navigator as a complement to their haptic feedback device. Publications took place after setting up the patent extension [12], [13]. The ongoing BQR INTUACTIVE funded by Grenoble-INP will lead to further scientific topics regarding interactions during grasping as well as with deformable bodies and a partnership is ongoing with GIPSA-Lab to study the muscular activity during interactions. A specific experiment has been set up to study the user's muscles activity.

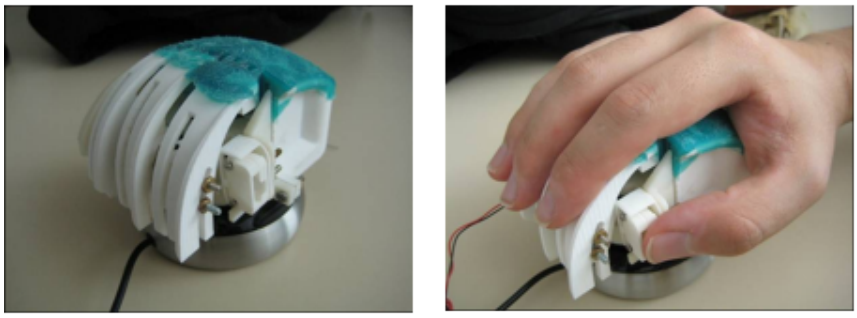


Figure 17. Current version of the HandNavigator prototype with three sensors per finger and a vibration damping structure.

IN-SITU Project-Team

6. New Results

6.1. Interaction Techniques

Participants: Caroline Appert, Michel Beaudouin-Lafon, David Bonnet, Anastasia Bezerianos, Olivier Chapuis, Emilien Ghomi, Stéphane Huot, Can Liu, Wendy Mackay [correspondant], Mathieu Nancel, Cyprien Pindat, Emmanuel Pietriga, Theophanis Tsandilas, Julie Wagner.

We explore interaction techniques in a variety of contexts, including individual interaction techniques on mobile devices, the desktop, and very large wall-sized displays, using one or both hands. We also explore interaction with physical objects and across multiple devices, to create mixed or augmented reality systems. This year, we explored interaction techniques based on time (*EWE* and *Dwell-and-Spring*), bimanual interaction on mobile devices (*Bipad*) and interaction on very large wall displays (*Jelly Lenses*, *Looking-Around-Bezels*). We also developed interactive paper systems to support early, creative design (*Pen-based Mobile Assistants*, *Paper Tonnetz*, *Paper Substrates*). We also explored augmented reality systems, using tactile feedback (*TactileSnowboard Instructions*) and tangible interaction (*Mobile AR*, *Combinatorix*) to support learning.

EWE – Although basic interaction techniques, such as multiple clicks or spring-loaded widgets, take advantage of the temporal dimension, more advanced uses of rhythmic patterns have received little attention in HCI. Using temporal structures to convey information can be particularly useful in situations where the visual channel is overloaded or even not available. We introduce Rhythmic Interaction [24] which uses rhythm as an input technique (Figure 10). Two experiments demonstrate that (i) rhythmic patterns can be efficiently reproduced by novice users and recognized by computer algorithms, and (ii) rhythmic patterns can be memorized as efficiently as traditional shortcuts when associated with visual commands. Overall, these results demonstrate the potential of Rhythmic Interaction and richer repertoire of interaction techniques. (*Best Paper award, CHI'12*)

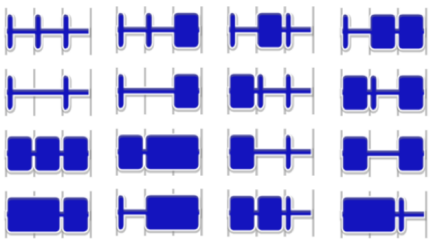


Figure 10. We defined 16 three-beat patterns: Each rectangle represents a tap, the thin gray lines show beats.

Dwell-and-Spring – Direct manipulation interfaces consist of incremental actions that should be reversible The challenge is how to provide users with an effective "undo". Actions such as manipulating geometrical shapes in a vector graphics editor, navigating a document using a scrollbar, or moving and resizing windows on the desktop rarely offer an undo mechanism. Users must manually revert to the previous state by recreating a similar sequence of direct manipulation actions, with a high level of associated motor and cognitive costs. We need a consistent mechanism that supports undo in multiple contexts. *Dwell-and-Spring* [19] uses a spring metaphor that lets users undo a variety of direct manipulations of the interface. A spring widget pops up whenever the user dwells during a press-drag-release interaction, giving her the opportunity to either cancel the

current manipulation (Figure 11) or undo the last one. The technique is generic and can easily be implemented on top of existing applications to complement the traditional undo command. A controlled experiment demonstrated that users can easily discover the technique and adopt it quickly when it is discovered.

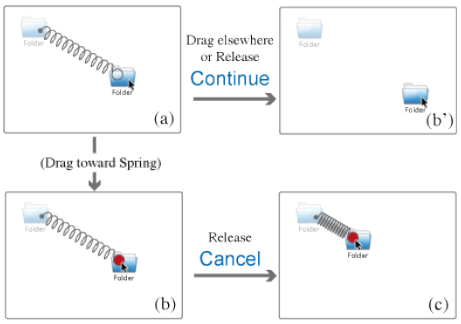


Figure 11. Cancel scenario: The user dwells while dragging an icon (a), which pops up a spring. She either (b) catches the spring's handle and releases the mouse button to cancel the current drag and drop, causing the spring to shrink smoothly (c) and returning the cursor and icon to their original locations, or she continues dragging the spring's handle any direction (b').

BiPad – Although bimanual interaction is common on large tabletops, it is rare on hand-held devices. We take advantage of the advanced multitouch input capabilities available on today's tablets to introduce new bimanual interactio techniques, under a variety of mobility conditions. We found that, when users hold a tablet, the primary function of the non-domininant hand is to provide support, which limits its potential movement. We studied how users "naturally" hold multi-touch tablets to identify comfortable holds, and then developed a set of 10 two-handed interaction techniques that accounts for the need to support the device while interacting with it. We introduced the *BiTouch* design space that extends Guiard's "Kinematic Chain Theory" [49] to account for the *support function* in bimanual interaction. We also designed and implemented the *BiPad* toolkit and set of interactions, which enables us to implement bimanual interaction on multitouch tablets (Figure 12). Finally, a controlled experiment demonstrated the benefits and trade-offs among specific techniques and offered insights for designing bimanual interaction on hand-held devices [31].

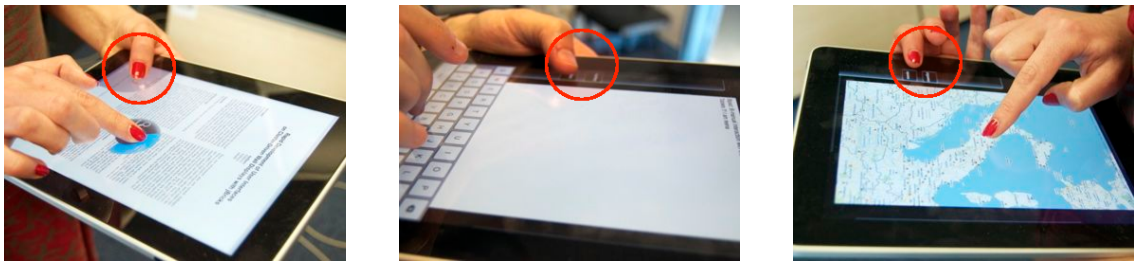


Figure 12. Bimanual interaction on a multitouch tablet with BiPad: (left) navigating in a document; (center) switching to uppercase while typing on a virtual keyboard; (right) zooming a map. The non-dominant hand is holding the device and could perform 'tapa', 'gestures' or 'chords' in order to augment dominant hand's interactions.

Jelly Lenses – Focus+context lens-based techniques smoothly integrate two levels of detail using spatial distortion to connect the magnified region and the context. Distortion guarantees visual continuity, but causes problems of interpretation and focus targeting, partly due to the fact that most techniques are based on statically-defined, regular lens shapes, that result in far-from-optimal magnification and distortion (Figure 13 left and center). JellyLenses [27] dynamically adapt to the shape of the objects of interest, providing detail-in-context visualizations of higher relevance by optimizing what regions fall into the focus, context and spatially-distorted transition regions (Figure 13 -right). A multi-scale visual search task experiment demonstrated that JellyLenses consistently perform better than regular fisheye lenses.

Figure 13. Magnifying the Lido in Venice. (left) a small fisheye magnifies one part of the island (Adriatic sea to Laguna Veneta), but requires extensive navigation to the whole island in detail; (center) a large fisheye magnifies a bigger part of the island, but severely distorts almost the entire image, hiding other islands; (right) a JellyLens automatically adapts its shape to the region of interest, with as much relevant information in the focus as (b) while better preserving context: surrounding islands are almost untouched from (a).

Looking behind Bezels – Using tiled monitors to build wall-sized displays has multiple advantages: higher pixel density, simpler setup and easier calibration. However, the resulting display walls suffer from the visual discontinuity caused by the bezels that frame each monitor. To avoid introducing distortion, the image has to be rendered as if some pixels were drawn behind the bezels. In turn, this raises the issue that a non-negligible part of the rendered image, that might contain important information, is visually occluded. We drew upon the analogy to french windows that is often used to describe this approach, and make the display really behave as if the visualization were observed through a french window [21]. We designed and evaluated two interaction techniques that let users reveal content hidden behind bezels. One enables users to offset the entire image through explicit touch gestures. The other adopts a more implicit approach: it makes the grid formed by bezels act like a true french window using head tracking to simulate motion parallax, adapting to users' physical movements in front of the display. The two techniques work for both single- and multiple-user contexts.

Pen-based Mobile Assistants – Digital pen technology allows easy transfer of pen data from paper to the computer. However, linking handwritten content with the digital world remains difficult as it requires the translation of unstructured and highly personal vocabularies into data structured so as to be understood and processed by a computer. Automatic recognition can help, but is not always reliable: it require active cooperation between users and recognition algorithms. We examined [30] the use of portable touch-screen devices in connection with pen and paper to help users direct and refine the interpretation of their strokes on paper. We explored four bimanual interaction techniques that combine touch and pen-writing, where user attention is divided between the original strokes on paper and their interpretation by the electronic device. We demonstrated these techniques through a mobile interface for writing music (Figure 14) that complements the automatic recognition with interactive user-driven interpretation. An experiment evaluated the four techniques and provided insights as to their strengths and limitations.

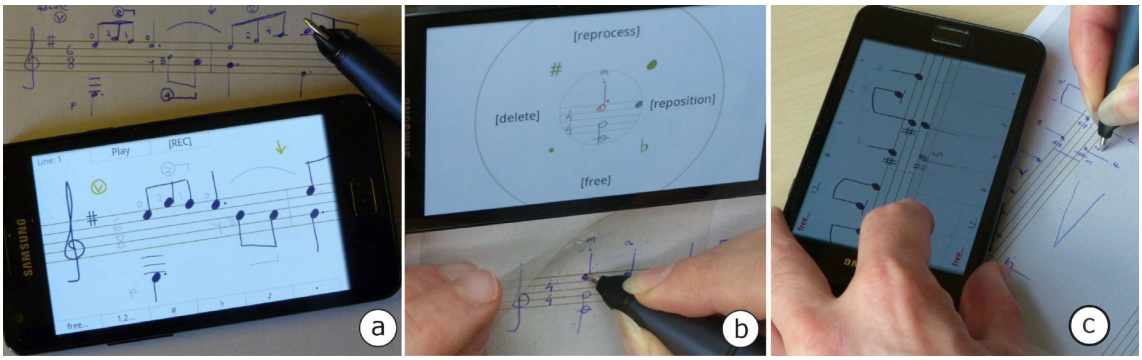


Figure 14. Writing music with a pen and a smartphone. (a) Handwritten score translated by the device. (b) Correcting the recognition of a note over a plastic sheet. (c) Guiding the interpretation of strokes with the left hand.

Paper Tonnetz – Tonnetz are space-based musical representations that lay out individual pitches in a regular structure. We investigated how properties of Tonnetz can be applied in the composition process, including how to represent pitch based on chords or scales and lay them out in a two-dimensional space (Figure 15). *PaperTonnetz*[20] is a tool that lets musicians explore and compose music with Tonnetz representations by making gestures on interactive paper, creating replayable patterns that represent pitch sequences and/or chords. An initial test in a public setting demonstrated significant differences between novice and experienced musicians and led to a revised version that explicitly supports discovering, improvising and assembling musical sequences in a Tonnetz.

Paper Substrates – Our goal is to design novel interactive paper interfaces that support the creative process. We ran a series of participatory design sessions with music composers to explore the concept of "paper substrates" [23]. Substrates are ordinary pieces of paper, printed with an Anoto dot pattern, that suppport a variety of advanced forms of interaction (Figure 15). Each substrate is strongly typed, such as a musical score or a graph, which faciliates interpretation by the computer. The composers were able to create, manipulate and combine layers of data, rearranging them in time and space as an integral part of the creative process. Moreover, the substrates approach fully supported an iterative process in which templates can evolve and be reused, resulting in highly personal and powerful interfaces. We found that paper substrates take on different roles, serving as data containers, data filters, and selectors. The design sessions resulted in several pen interactions and tangible manipulations of paper components to support these roles: drawing and modifying specialized data over formatted paper, exploring variations by superimposing handwritten data, defining programmable modules, aligning movable substrates, linking them together, overlaying them, and archiving them into physical folders.

Tactile Snowboard Instructions – Beginnning snowboarders have difficulty getting instructions and feedback on their performance if they are separated spatially from their coach. Snowboarders can learn correct technique by wearing a system with actuators (vibration motors) attached to the thighs and shoulders, which reminds them to shift their weight and to turn their upper body in the correct direction (Figure 16). A field study with amateur snowboarders demonstrated that these "tactile instructions" are effective for learning basic turns and offered recommendations for incorporating tactile instructions into sports training. *Best Paper award, Mobile HCI'12*

Mobile AR – We examined how new capabilities of hand-held devices, specifically higher resolution screens, camera and localization, can be used to create mobile *Augmented Reality (AR)* to help users learn and manage their interactions with everyday physical objects, such as door codes and home appliances. We explored AR-based mobile note-taking [50] to provide real-time on-screen feedback of physical objects that the user must manipulate, such as entering a door code. Here, the user uses the device to identify the required values of sliders

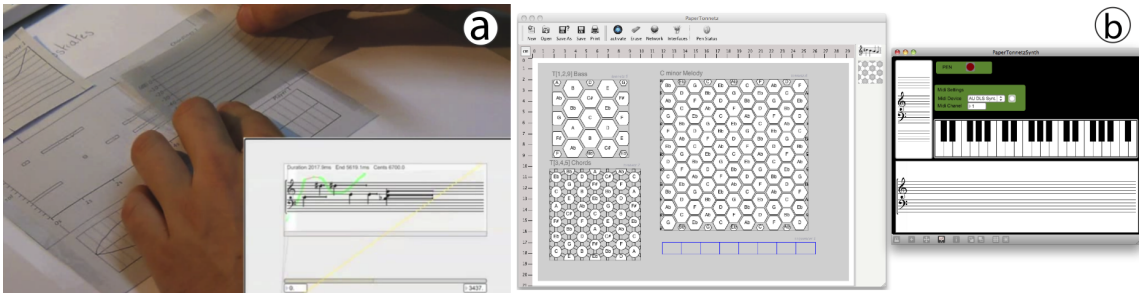


Figure 15. Paper-based interfaces for musical creation. (a) Paper substrates are interactive paper components for working with musical data. (b) PaperTonnetz main interface representing the virtual page with three Tonnetz and one sequencer (left). The Max/MSP patch to play and visualize created sequences (right).

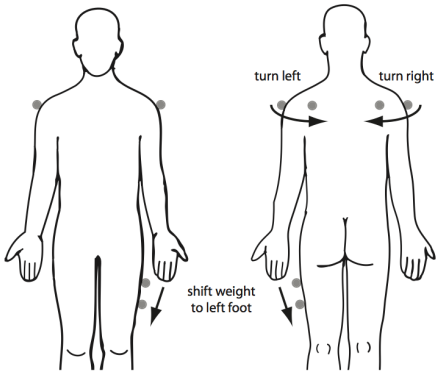


Figure 16. Two vibration motors are placed at each shoulder and laterally at the thigh that points forward during the ride. Arrows illustrate the direction of the stimuli on the skin, labels show the corresponding messages.

and buttons (Figure 17). A controlled experiment showed that mobile AR improved both speed and accuracy over traditional text or picture-based instructions. We also demonstrated that adding real-time feedback in the AR layer that shows the user's actions with respect to the physical controls further increases performance [25]. (*Honorable Mention award, CHI 2012*)

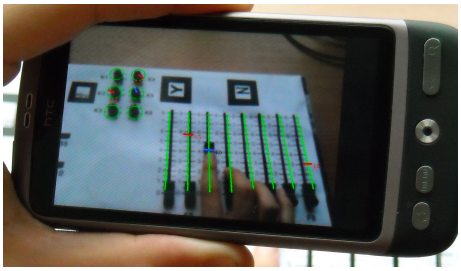


Figure 17. Mobile augmented reality for setting physical controls. Required values are displayed in red and turn blue when set correctly.

Combinatorix – We developed Combinatorix [28], a mixed tabletop system to help groups of students work collaboratively to solve probability problems. Users combine tangible objects in different orders and watch the effects of various constraints on the problem space (Figure 18). A second screen displays an abstract representation, such as a probability tree, to show how their actions influenced the total number of combinations. We followed an iterative participatory design process with college students taking a combinatorics class and demonstrated the benefits of using a tangible approach to facilitate learning abstract concepts.

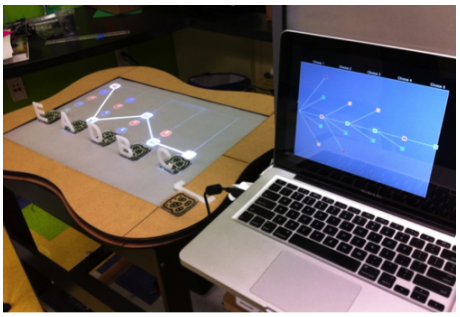


Figure 18. Combinatorix uses tangible objects on an interactive tabletop to control the tabletop display and associated screen, to help users explore and understand complex problems in combinatorial statistics.

6.2. Research Methods

Participants: Caroline Appert, Michel Beaudouin-Lafon, Anastasia Bezerianos, Olivier Chapuis, Jérémie Garcia, Stéphane Huot, Ilaria Liccardi, Wendy Mackay [correspondant], Emmanuel Pietriga.

Human-Computer Interaction is a multi-disciplinary field, with elements of computer science, software engineering, experimental psychology and anthropology. More recently, designers have joined the CHI community, offering an important perspective, but also a different fundamental research paradigm, which

differs from the value systems of engineering and the natural sciences. We explored the paradigm of *Research through Design* [34], which we differentiate from traditional epistemologies in the human sciences. We distinguish design from research-through-design: the end goal is not to produce an artifact, but rather to frame an alternative future and uncover unmet human needs, desires, emotions, and aspirations. We identified three research perspectives that have been adopted within the HCI community: *Projection* explores possible future states, *Place* specifies the context in which design artifacts presented to gather data, and *Point-of-View* identifies the philosophical perspective imposed by researchers. Our goal is to understand what it means to conduct research through design and how to value research-through-design contributions.

In addition to exploring general questions about research paradigms, we also explore more focused questions that apply new research methods to the design of multi-surface environments. Large interactive surfaces (like the WILD platform) are interesting collaborative settings that allow viewing of large amounts of visual information. Before we are ready to use these platforms in real visual analysis situations, where people can place themselves at different positions around the display, we must first understand how the perception of visual information is affected by perspective distortion introduced by varying viewing distances and angles. A deeper understanding of such distortion effects can help visualization researchers design effective visualizations for these spaces and implement interaction techniques to aid in extreme distortion situations. We conducted [16] two studies on the perception of visual variables that encode information, such as Angle, Area, and Length, and found that perception is impacted differently at different locations on the screen, depending on the vertical and horizontal positioning of this information. The visual variable of Length was the most accurately perceived across the display. Our second study examined the effect of perception when participants can move freely in such situations, compared to when they have a static viewpoint, and found that a far but static viewpoint was as accurate but less time consuming than one that included free motion. But we observed that in free motion participants often chose non-optimal walking strategies that can increased perception errors, thus we provide precise recommendations on where and how to move in such environments. This work is a first step towards understanding and predicting the impact of large display environments to people's understanding and tasks.

Annotations play an important role in visual analysis and record-keeping. We discuss the use of annotations on visualization dashboards citeelias:hal-00719221, collections of linked visualizations, focusing on business intelligence analysis through a user centered design process from expert analysts in this domain. The first contribution bridges the gap between expert analyst needs and designers, when it comes to visualization annotations. The second offers a novel approach to annotating of visualizations, "context aware annotations": We annotate data queries directly, rather than image/chart locations. Annotations are present irrespective of the visual data representations users select (different charts, numeric tabular views of their data, etc). We focus particularly on novel annotation aspects made possible by this approach, such as multi-visualization annotations, annotation transparency across charts and data hierarchy levels, and the possibility of recommending to users annotations done to similar data to enable annotation re-use. We also consider new challenges that arise from such our approach, such as what happens to annotations when the underlying data is changed, and provide recommendations and design solutions.

6.3. Engineering of interactive systems

Participants: Caroline Appert, Michel Beaudouin-Lafon [correspondant], Olivier Chapuis, Stéphane Huot, Wendy Mackay, Emmanuel Pietriga, Clément Pillias, Romain Primet.

INSITU continues to develop and apply toolkits to explore and implement interactive systems. Most of the projects listed in the *Interaction Techniques* section either build upon existing toolkits, e.g., *Jelly Lenses* to improve management of focus+context on a wall-sized display, or created new ones, e.g., *BiPad* to create various bimanual interaction techniques for hand-held tablets.

INSITU's primary testbed for exploring multi-surface interaction is the WILD Room [15] (Wall-sized Interaction with Large Datasets), a multisurface environment featuring a wall-sized display, a multitouch table, and various mobile devices. Our goal is to explore the next generation of interactive systems by distributing interaction across these diverse computing devices, enabling multiple users to easily and seamlessly create,

share, and manipulate digital content. Our research strategy is to design an extreme environment that pushes the limits of technology – both hardware and software. To ground the design process, we work with extreme users – scientists whose daily work both inspires and stress-tests the environment as they seek to understand exceptionally large and complex datasets. The WILD room, and the soon-to-be-built WILDER room are part of *DigiScope*, a 22.5 Meuro "Equipement d'Excellence" project led by INSITU.

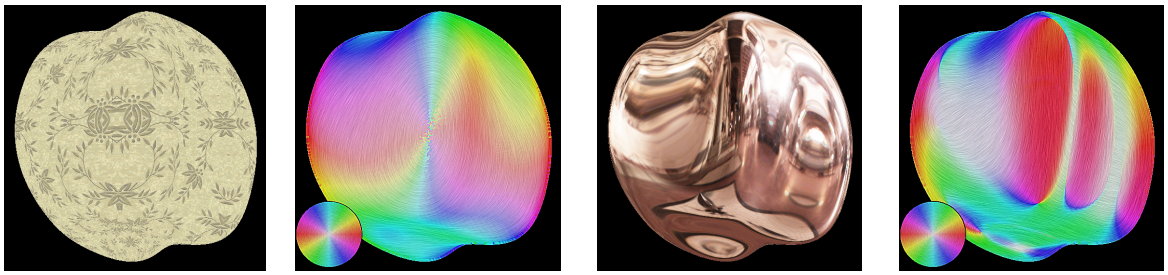
INSITU's collaboration with the ALMA radio-telescope on the design and implementation of user interfaces for operations monitoring and control continued this year [26], and was eventually transferred to Inria Chile in July (see Section 7.4.2). The ALMA radio-telescope, currently under construction in northern Chile, is a very advanced instrument that presents numerous challenges. From a software perspective, one critical issue is the design of graphical user interfaces for operations monitoring and control that scale to the complexity of the system and to the massive amounts of data users are faced with. Early experience operating the telescope with only a few antennas showed that conventional, WIMP-based user interfaces are not adequate in this context. They consume too much screen real-estate, require many unnecessary interactions to access relevant information, and fail to provide operators and astronomers with a clear mental map of the instrument. They increase extraneous cognitive load, impeding tasks that call for quick diagnosis and action. To address this challenge, the ALMA software division adopted a user-centered design approach in collaboration with members of INSIUT. For the last two years, astronomers, operators, software engineers and human-computer interaction researchers from INSITU have been working on the design of better user interfaces based on state-of-the-art visualization techniques. This eventually led to the joint development of those interface components using various software toolkits, some of them developed at INSITU (Section 5.2).

MANAO Team

5. New Results

5.1. Axis 1:Analysis and Simulation

5.1.1. First Order Analysis of Shading



Texuring | 1st order gradient field | Environment reflection | 2st order gradient field

Figure 8. First-oder analysis [21] 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.

We introduced [21] a novel method for producing convincing pictures of shaded objects based entirely on 2D image operations. This approach, which we call image-based shading design, offers direct artistic control in the picture plane by deforming image primitives so that they appear to conform to specific 3D shapes. Using a differential analysis of reflected radiance, we have identified the two types of surface flows involved in the depiction of shaded objects, which are consistent with recent perceptual studies. We have also introduced two novel deformation operators that closely mimic surface flows while providing direct artistic controls in real-time.

5.1.2. Rational BRDF

Over the last two decades, much effort has been devoted to accurately measuring Bidirectional Reflectance Distribution Functions (BRDFs) of real-world materials and to use efficiently the resulting data for rendering. Because of their large size, it is difficult to use directly measured BRDFs for real-time applications, and fitting the most sophisticated analytical BRDF models is still a complex task. In this paper, we introduce Rational BRDF [19], a general-purpose and efficient representation for arbitrary BRDFs, based on Rational Functions (RFs). Using an adapted parametrization, we demonstrate how Rational BRDFs offer 1) a more compact and efficient representation using low-degree RFs, 2) an accurate fitting of measured materials with guaranteed control of the residual error, and 3) efficient importance sampling by applying the same fitting process to determine the inverse of the Cumulative Distribution Function (CDF) generated from the BRDF for use in Monte-Carlo rendering.

5.2. Axis 2: From Acquisition to Display

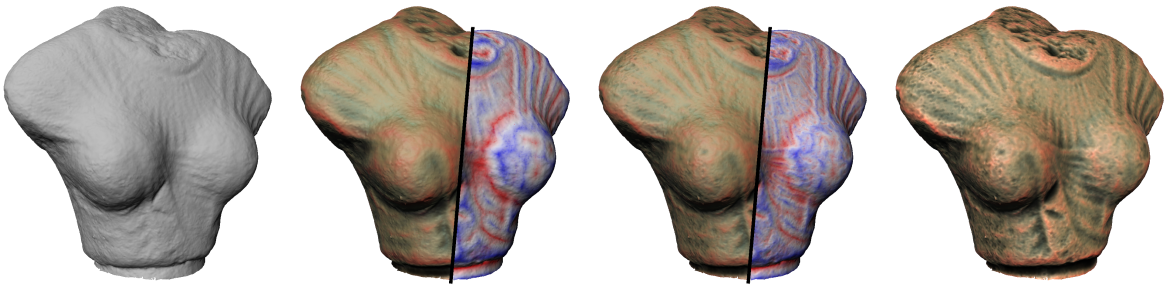
5.2.1. Outdoor Lighting for Augmented Reality

Figure 9. Consistent illumination of a virtual car in real outdoor lighting.

In augmented reality, one of the key tasks to achieve a convincing visual appearance consistency between virtual objects and video scenes is to have a coherent illumination along the whole sequence. As outdoor illumination is largely dependent on the weather, the lighting condition may change from frame to frame. We have proposed [17] a full image-based approach for online tracking of outdoor illumination variations from videos captured with moving cameras. Our key idea is to estimate the relative intensities of sunlight and skylight via a sparse set of planar feature-points extracted from each frame. To address the inevitable feature misalignments, a set of constraints are introduced to select the most reliable ones. Exploiting the spatial and temporal coherence of illumination, the relative intensities of sunlight and skylight are finally estimated by using an optimization process. We have validated our technique on a set of real-life videos and show that the results with our estimations are visually coherent along the video sequences (cf. Figure 9).

5.3. Axis 3: Rendering, Visualization and Illustration

5.3.1. Surface Relief Analysis for Illustrative Shading

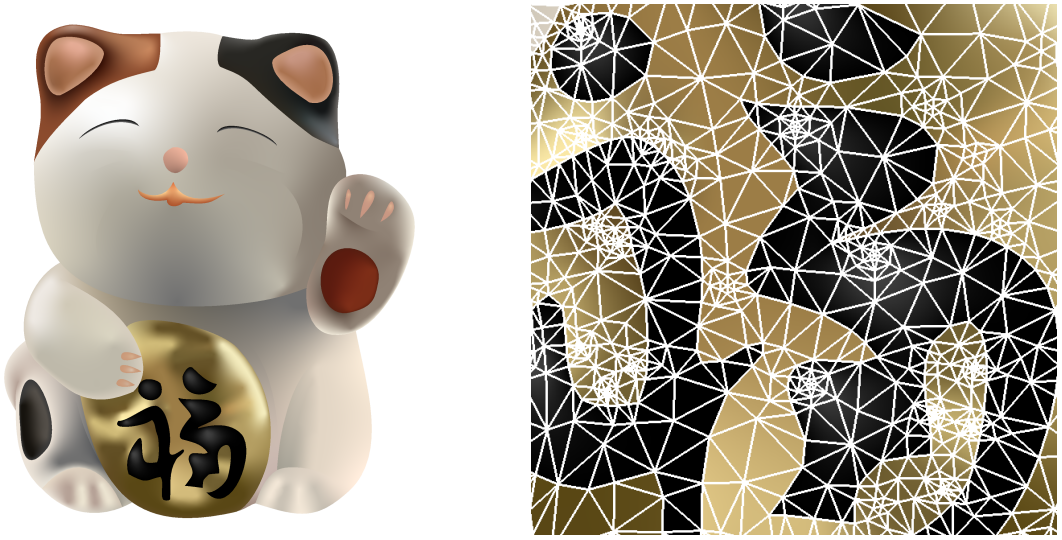


(a) (b) (c) (d)

Figure 10. Given a detailed surface (a), we analyze its relief to locate relief features in the neighborhood of each surface point (b). We focus on three types of features: convexities, concavities, and inflexions, shown on the right half with blue, red and white colors respectivelly. Extracted information is used to assign them different shading functions: here we use three different lit-spheres, shown on the left half. An additional accessibility shading effect helps convey relief cavities. Features are extracted and combined at multiple scales to depict relevant relief details (c). Finally, radiance scaling is added to enhance the relief based on the curvature at each feature (d).

Rendering techniques are often used to convey shape in scientific illustrations. We present an analysis technique that leverages the complexity found in detailed 3D models for illustrative shading purposes. Given a smooth base surface with relief, it locates relief features (concavities, convexities and inflections) around each surface point and at multiple scales, using cubic-polynomial fitting. This object-space, per-vertex information is then used to guide a variety of shading techniques including normal enhancement, feature visualization, accessibility shading and radiance scaling. Thanks to this approach, features at multiple scales are easily combined, filtered and shaded, allowing users to explore surface relief in real-time (cf. Figure 10).

5.4. Axis 4: Editing and Modeling



(a) (b)

Figure 11. A complex image obtained using our vectorial solver (a), with a close-up view showing the automatically generated intermediate triangle mesh (b).

5.4.1. Free form vector gradients

The creation of free-form vector drawings as been greatly improved in recent years with techniques based on harmonic or bi-harmonic interpolation. Such methods offer the best trade-off between sparsity (keeping the number of control points small) and expressivity (achieving complex shapes and gradients). Unfortunately, the lack of a robust and versatile method to compute such images still limits their use in real-world applications. We developed a vectorial solver for the computation of free-form vector gradients based on a non-conform Finite Element Methods (FEM). Its key feature is to output a low-level vector representation suitable for very fast GPU accelerated rasterization and close-form evaluation (fig. 11). This intermediate representation is hidden from the user: it is dynamically updated using FEM during drawing when control points are edited. We demonstrated novel usages of vector drawings such as instancing, layering, deformation, texture and environment mapping. Finally, we also generalized and extended the set of drawing possibilities, in particular, by showing how to locally control vector gradients. This work has been published at SIGGRAPH Asia [16] and featured by the 3DFV website [24].

5.4.2. Growing Least Squares (GLS) for the Analysis of Manifolds in Scale-Space

We created a novel approach for the multi-scale analysis of point-sampled manifolds of co-dimension 1. It is based on a variant of Moving Least Squares, whereby the evolution of a geometric descriptor at increasing scales is used to locate pertinent locations in scale-space, hence the name “Growing Least Squares (GLS)”. Compared to existing scale-space analysis methods, our approach is the first to provide a continuous solution in space and scale dimensions, without requiring any parametrization, connectivity or uniform sampling. An important implication is that we identify multiple pertinent scales for any point on a manifold, a property that had not yet been demonstrated in the literature. In practice, our approach exhibits an improved robustness to change of input, and is easily implemented in a parallel fashion on the GPU, and it can be used in a wide variety of applications. For example, the GLS can be used for the detection of similarity, according to a given scale range (see Figure 12). This work has been published at the Symposium of Geometry Processing [18].

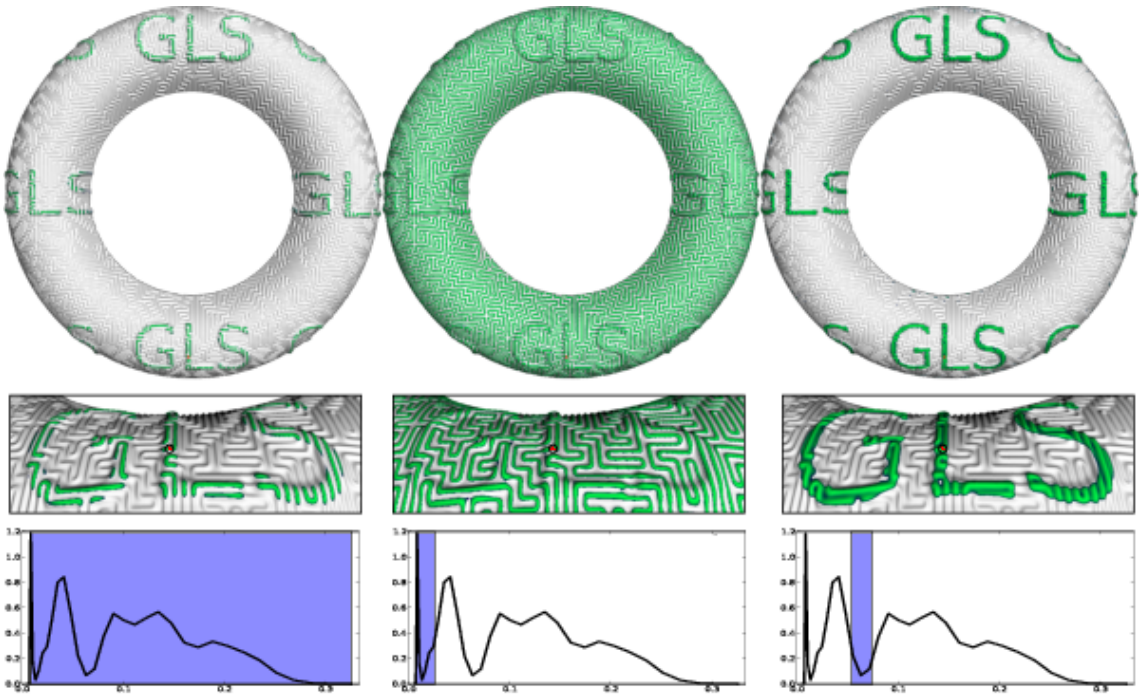


Figure 12. GLS Multi-scale similarity. Top and middle rows: For a selected point (in red), similar points are selected (in green) via our dissimilarity measure. The similarity is computed for each vertex and interpolated per fragment during the rendering. Bottom row: the type of selected feature depends on a user-controlled global prior (shown as a blue box), which is locally refined by our geometric variation. In (a), all scales are selected. In (b), only the fine displacement pattern emerges. In (c), the large-scale GLS letters are properly segmented.

MAVERICK Team

6. New Results

6.1. Computer visualization

6.1.1. Immersive Virtual Environment for Visuo-Vestibular Therapy: Preliminary Results

Participants: Jean-Dominique Gascuel, Henri Payno, Sébastien Schmerber, Olivier Martin.

The sense of equilibrium aggregates several interacting cues. On patients with vestibular loss, vision plays a major role. In this study, the goal is to propose a new immersive therapy based on 3D opto-kinetic stimulation. We propose to demonstrate that 3D monoscopic optical flows are an efficient tool to stimulate adaptive postural adjustment. We developed an immersive therapeutic platform that enables to tune the balance task difficulty by managing optic flow speed and gaze anchoring (Figure 5). METHODOLOGY: the immersive sessions proposed to vestibular areflexic patients are composed of a repetition of dynamic optic flows, with varying speed and presence or not of a gaze anchor. The balance adjustments are recorded by a force plate, and quantified by the length of the center of pressure trajectory. RESULTS: Preliminary analysis shows that (i) Patients report a strong immersion felling in the motion flow, triggering more intense motor response to "fight against fall" than in standard opto-kinetic protocols; (ii) An ANOVA factorial design shows a significant effect of flow speed, session number and gaze anchor impact. CONCLUSION: This study shows that 3D immersive stimulation removes essential limits of traditional opto-kinetic stimulators (limited 2D motions and remaining fixed background cues). Moreover, the immersive optic flow stimulation is an efficient tool to induce balance adaptive reactions in vestibular patients. Hence, such a platform appears to be a powerful therapeutic tool for training and relearning of balance control processes.

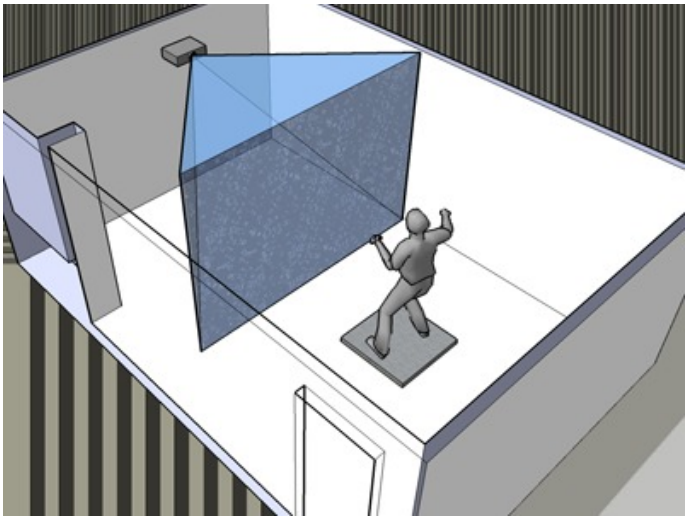


Figure 5. The immersive platform, installed in an available room of the hospital. The large retro projected screen is at 60 cm of the patient, covering most of its visual field. The patient is standing on a force plate, recording CoP.

6.1.2. Evaluation of Depth of Field for Depth Perception in DVR

Participants: Pascal Grosset, Mathias Schott, Georges-Pierre Bonneau, Hansen Charles.

we present a user study on the use of Depth of Field for depth perception in Direct Volume Rendering (Figure 6). Direct Volume Rendering with Phong shading and perspective projection is used as the baseline. Depth of Field is then added to see its impact on the correct perception of ordinal depth. Accuracy and response time are used as the metrics to evaluate the usefulness of Depth of Field. The on site user study has two parts: static and dynamic. Eye tracking is used to monitor the gaze of the subjects. From our results we see that though Depth of Field does not act as a proper depth cue in all conditions, it can be used to reinforce the perception of which feature is in front of the other. The best results (high accuracy & fast response time) for correct perception of ordinal depth is when the front feature (out of the users were to choose from) is in focus and perspective projection is used.

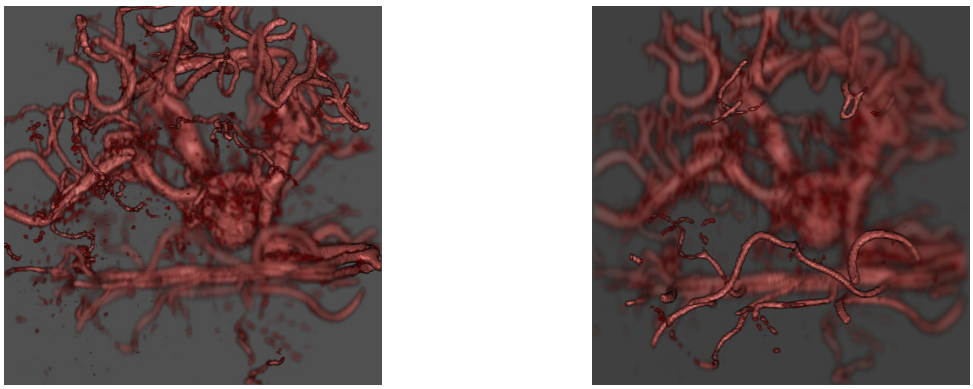


Figure 6. Aneurism. Depth of Field reinforces the perception of which feature is in front of the other.

6.1.3. Volume Preserving FFD for Programmable Graphics Hardware

Participants: Stefanie Hahmann, Georges-Pierre Bonneau, Sébastien Barbier, Gershon Elber, Hans Hagen.

Free Form Deformation (FFD) is a well established technique for deforming arbitrary object shapes in space. Although more recent deformation techniques have been introduced, amongst them skeleton-based deformation and cage based deformation, the simple and versatile nature of FFD is a strong advantage, and justifies its presence in nowadays leading commercial geometric modeling and animation software systems. Since its introduction in the late 80's, many improvements have been proposed to the FFD paradigm, including control lattices of arbitrary topology, direct shape manipulation and GPU implementation. Several authors have addressed the problem of volume preserving FFD. These previous approaches either make use of expensive non-linear optimization techniques, or resort to first order approximation suitable only for small-scale deformations. In this paper we take advantage from the multi-linear nature of the volume constraint in order to derive a simple, exact and explicit solution to the problem of volume preserving FFD. Two variants of the algorithm are given, without and with direct shape manipulation. Moreover, the linearity of our solution enables to implement it efficiently on GPU (Figure 7).

6.1.4. Sharp feature preserving MLS surface reconstruction based on local feature line approximations

Participants: Christopher Weber, Stefanie Hahmann, Hans Hagen, Georges-Pierre Bonneau.

Sharp features in manufactured and designed objects require particular attention when reconstructing surfaces from unorganized scan point sets using moving least squares (MLS) fitting. It's an inherent property of MLS fitting that sharp features are smoothed out. Instead of searching for appropriate new fitting functions our approach computes a modified local point neighborhood so that a standard MLS fitting can be applied

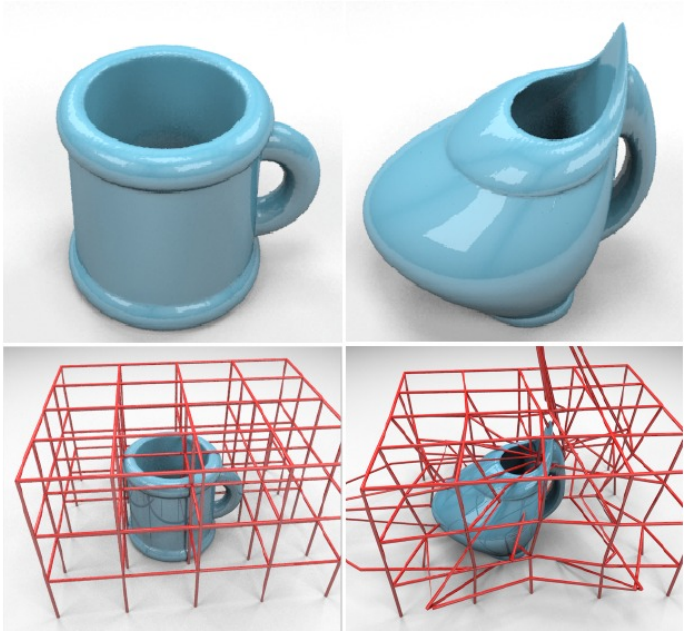


Figure 7. Sculptured cup with volume preservation.

enhanced by sharp features reconstruction. We present a two-stage algorithm. In a pre-processing step sharp feature points are marked first. This algorithm is robust to noise since it is based on Gauss map clustering. In the main phase, the selected feature points are used to locally approximate the feature curve and to segment and enhance the local point neighborhood. The MLS projection thus leads to a piecewise smooth surface preserving all sharp features. The method is simple to implement and able to preserve line-type features as well as corner-type features during reconstruction (Figure 8).

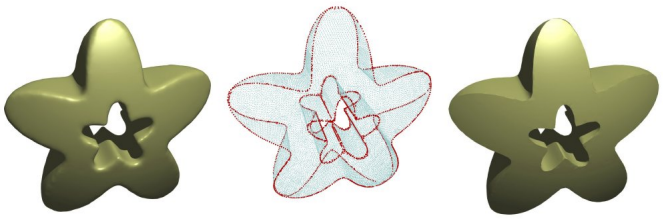


Figure 8. Left: standard MLS surface. Middle: feature point detection in point cloud. Right: sharp feature preserving MLS.

6.2. Expressive rendering

6.2.1. Active Strokes: Coherent Line Stylization for Animated 3D Models

Participants: Pierre Bénard, Lu Jingwan, Forrester Cole, Adam Finkelstein, Joëlle Thollot.

We present a method for creating coherently animated line drawings that include strong abstraction and stylization effects (Figure 9). These effects are achieved with active strokes: 2D contours that approximate and track the lines of an animated 3D scene. Active strokes perform two functions: they connect and smooth unorganized line samples, and they carry coherent parameterization to support stylized rendering. Line samples are approximated and tracked using active contours ("snakes") that automatically update their arrangment and topology to match the animation. Parameterization is maintained by brush paths that follow the snakes but are independent, permitting substantial shape abstraction without compromising fidelity in tracking. This approach renders complex models in a wide range of styles at interactive rates, making it suitable for applications like games and interactive illustrations.

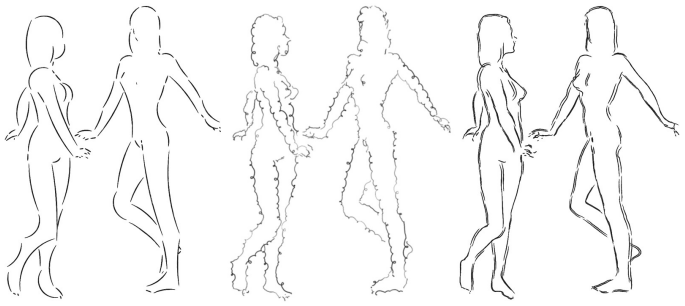


Figure 9. Stylization examples. Woman in two poses and three styles: arcs, loopy offsets, and overdrawn.

6.2.2. Temporally Coherent Video Stylization

Participants: Pierre Bénard, Joëlle Thollot, John Collomosse.

The transformation of video clips into stylized animations remains an active research topic in Computer Graphics. A key challenge is to reproduce the look of traditional artistic styles whilst minimizing distracting flickering and sliding artifacts; i.e. with temporal coherence. This chapter surveys the spectrum of available video stylization techniques, focusing on algorithms encouraging the temporally coherent placement of rendering marks, and discusses the trade-offs necessary to achieve coherence. We begin with flow-based adaptations of stroke based rendering (SBR) and texture advection capable of painting video. We then chart the development of the field, and its fusion with Computer Vision, to deliver coherent mid-level scene representations. These representations enable the rotoscoping of rendering marks on to temporally coherent video regions, enhancing the diversity and temporal coherence of stylization. In discussing coherence, we formalize the problem of temporal coherence in terms of three defined criteria, and compare and contrast video stylization using these.

6.3. Illumination simulation

6.3.1. Accurate fitting of measured reflectances using a Shifted Gamma micro-facet distribution

Participants: Mahdi M. Bagher, Cyril Soler, Nicolas Holzschuch.

Material models are essential to the production of photo-realistic images. Measured BRDFs provide accurate representation with complex visual appearance, but have larger storage cost. Analytical BRDFs such as Cook- Torrance provide a compact representation but fail to represent the effects we observe with measured appearance. Accurately fitting an analytical BRDF to measured data remains a challenging problem. In this paper we introduce the SGD micro-facet distribution for Cook-Torrance BRDF. This distribution accurately

models the behavior of most materials. As a consequence, we accurately represent all measured BRDFs using a single lobe. Our fitting procedure is stable and robust, and does not require manual tweaking of the parameters (Figure 10).

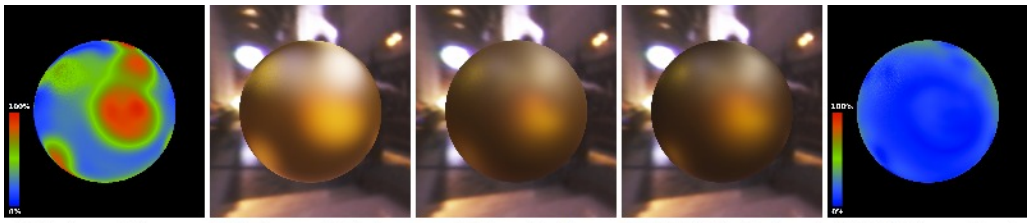


Figure 10. fitting of measured reflectances: comparison between ground truth and our approach.

6.3.2. Interactive rendering of acquired materials on dynamic geometry using bandwidth prediction

Participants: Mahdi M. Bagher, Cyril Soler, Kartic Subr, Laurent Belcour, Nicolas Holzschuch.

Shading complex materials such as acquired reflectances in multi-light environments is computationally expensive. Estimating the shading integral involves sampling the incident illumination independently at several pixels. The number of samples required for this integration varies across the image, depending on an intricate combination of several factors. Adaptively distributing computational budget across the pixels for shading is therefore a challenging problem. In this paper we depict complex materials such as acquired reflectances, interactively, without any precomputation based on geometry. We first estimate the approximate spatial and angular variation in the local light field arriving at each pixel. This *local bandwidth* accounts for combinations of a variety of factors: the reflectance of the object projecting to the pixel, the nature of the illumination, the local geometry and the camera position relative to the geometry and lighting. We then exploit this bandwidth information to adaptively sample for reconstruction and integration. For example, fewer pixels per area are shaded for pixels projecting onto diffuse objects, and fewer samples are used for integrating illumination incident on specular objects (Figure 11).

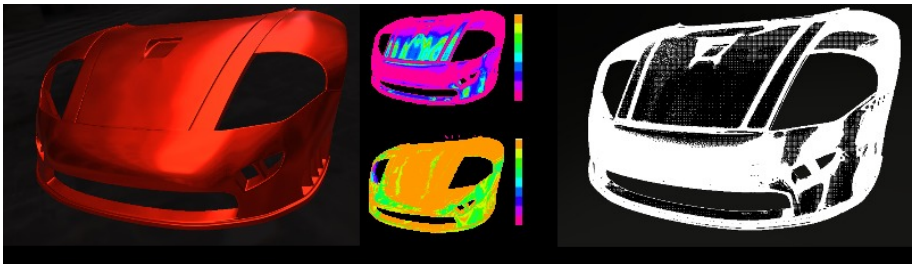


Figure 11. Interactive rendering of aquired materials. Center: predicted bandwidth and variance. Right: sample points where we compute illumination. Left: rendered result.

6.3.3. Real-Time Rendering of Rough Refraction

Participants: Charles De Rousiers, Adrien Bousseau, Kartic Subr, Nicolas Holzschuch, Ravi Ramamoorthi.

We present an algorithm to render objects made of transparent materials with rough surfaces in real-time, under all-frequency distant illumination (Figure 12). Rough surfaces cause wide scattering as light enters and exits objects, which significantly complicates the rendering of such materials. We present two contributions to approximate the successive scattering events at interfaces, due to rough refraction: First, an approximation of the Bidirectional Transmittance Distribution Function (BTDF), using spherical Gaussians, suitable for real-time estimation of environment lighting using pre-convolution; second, a combination of cone tracing and macro-geometry filtering to efficiently integrate the scattered rays at the exiting interface of the object. We demonstrate the quality of our approximation by comparison against stochastic ray-tracing. Furthermore we propose two extensions to our method for supporting spatially varying roughness on object surfaces and local lighting for thin objects.

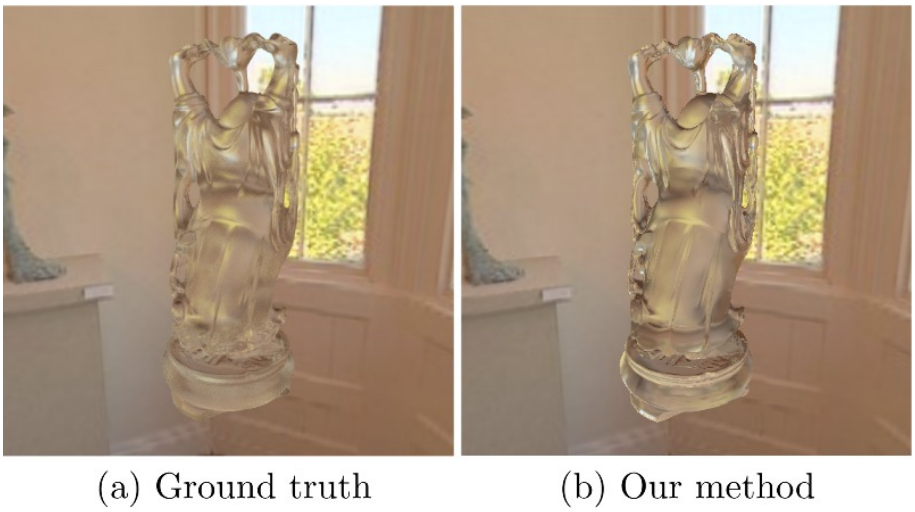


Figure 12. Real-Time Rendering of Rough Refraction

6.3.4. Multiple-scattering and double-scattering effects in translucent materials

Participants: Jean-Dominique Gascuel, Nicolas Holzschuch.

Some materials, such as coffee, milk or marble, have a soft translucent aspect because of sub-surface scattering: light enters them, is scattered several times inside before leaving in a different place. A full representation of sub-surface scattering effects in illumination simulation is computationally expensive. The main difficulty comes from multiple scattering events: the high number of events increases the incertainty on the result, forcing us to allocate more time for the computations. In this paper, we show that there is a strong correlation between the surface effects of multiple scattering inside the material and the effects after just two scatter events. This knowledge will help for accelerating multiple scattering effects.We also provide a model for fast computation of double-scattering events, using a precomputed density function we store in a compact way (Figure 13).

6.3.5. Frequency analysis of participating media

Participants: Laurent Belcour, Cyril Soler, Kavita Bala.

Computing global illumination in participating media is frustratingly expensive: while the computation itself is long and complicated, the result involve very smooth regions of illumination. This motivates an a priori approach to find out how fast the resulting image will vary in space (i.e. it's spatial frequency) to adapt computation effort to reach the maximal efficiency. For this we are extending the theory of Fourier Analysis of

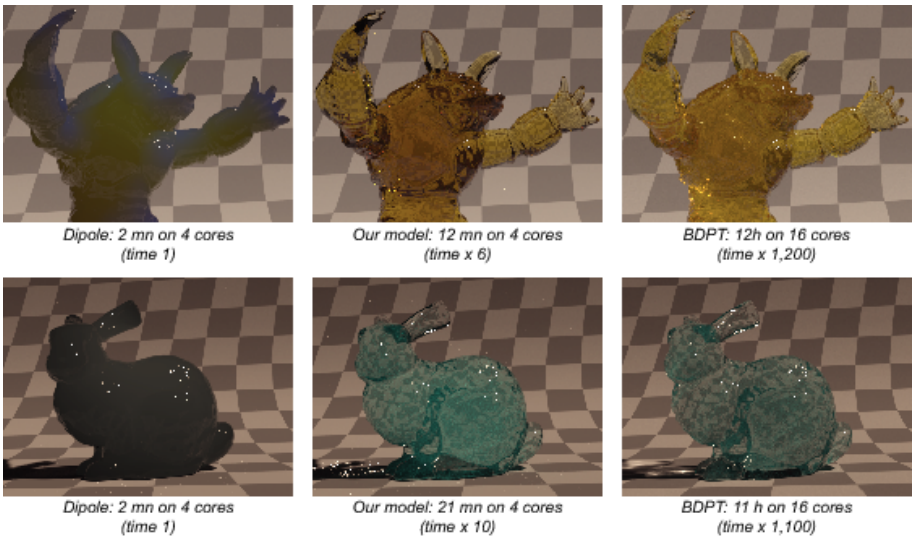


Figure 13. Multiple-scattering and double-scattering effects in translucent materials

light transport to participating media. Our work builds on the covariance analysis of light transport developed by Laurent Belcour in his PhD Thesis. It offers the possibility to drastically accelerate the algorithms involved in the computation of the illumination in scenes with participating media (Figure 14).

6.4. Complex scenes

6.4.1. A Survey of Non-linear Pre-filtering Methods for Efficient and Accurate Surface Shading

Participants: Eric Bruneton, Fabrice Neyret.

Rendering a complex surface accurately and without aliasing requires the evaluation of an integral for each pixel, namely a weighted average of the outgoing radiance over the pixel footprint on the surface. The outgoing radiance is itself given by a local illumination equation as a function of the incident radiance and of the surface properties. Computing all this numerically during rendering can be extremely costly. For efficiency, especially for real-time rendering, it is necessary to use precomputations. When the fine scale surface geometry, reflectance and illumination properties are specified with maps on a coarse mesh (such as color maps, normal maps, horizon maps or shadow maps), a frequently used simple idea is to pre-filter each map linearly and separately. The averaged outgoing radiance, i.e., the average of the values given by the local illumination equation is then estimated by applying this equation to the averaged surface parameters. But this is really not accurate because this equation is non-linear, due to self-occlusions, self-shadowing, non-linear reflectance functions, etc. Some methods use more complex pre-filtering algorithms to cope with these non-linear effects. This paper is a survey of these methods. We start with a general presentation of the problem of pre-filtering complex surfaces. We then present and classify the existing methods according to the approximations they make to tackle this difficult problem. Finally, an analysis of these methods allows us to highlight some generic tools to pre-filter maps used in non-linear functions, and to identify open issues to address the general problem.

6.4.2. Real-time Realistic Rendering and Lighting of Forests

Participants: Eric Bruneton, Fabrice Neyret.

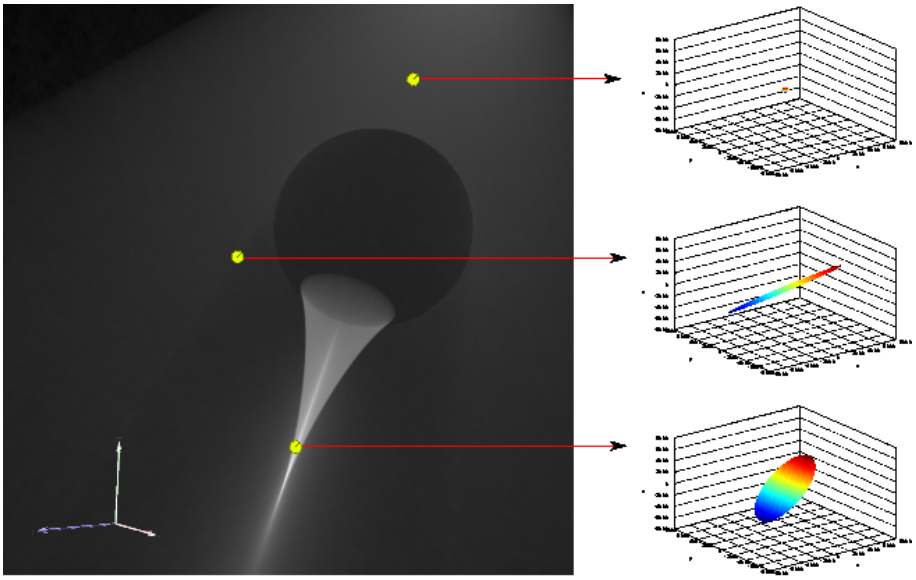


Figure 14. A glass sphere castinc a volumetric caustic in participating media with multiple scattering. We can predict the covariance of the spectrum of the illumination locally everywhere in the volume so as to adapt the computation effort.

Realistic real-time rendering and lighting of forests is an important aspect for simulators and video games. This is a difficult problem, due to the massive amount of geometry: aerial forest views display millions of trees on a wide range of distances, from the camera to the horizon. Light interactions, whose effects are visible at all scales, are also a problem: sun and sky dome contributions, shadows between trees, inside trees, on the ground, and view-light masking correlations. In this paper we present a method to render very large forest scenes in realtime, with realistic lighting at all scales, and without popping nor aliasing (Figure 15). Our method is based on two new forest representations, z-fields and shader-maps, with a seamless transition between them. Our first model builds on light fields and height fields to represent and render the nearest trees individually, accounting for all lighting effects. Our second model is a location, view and light dependent shader mapped on the terrain, accounting for the cumulated subpixel effects. Qualitative comparisons with photos show that our method produces realistic results.

Figure 15. Some real-time results obtained with our method, showing large forest scenes with a wide range of view distances, various tree densities and lighting conditions.

6.4.3. Representing Appearance and Pre-filtering Subpixel Data in Sparse Voxel Octrees

Participants: Eric Heitz, Fabrice Neyret.

Sparse Voxel Octrees (SVOs) represent efficiently complex geometry on current GPUs. Despite the fact that LoDs come naturally with octrees, interpolating and filtering SVOs are still issues in current approaches. In this paper, we propose a representation for the appearance of a detailed surface with associated attributes stored within a voxel octree. We store macro- and micro-descriptors of the surface shape and associated attributes in each voxel. We represent the surface macroscopically with a signed distance field and we encode subvoxel microdetails with Gaussian descriptors of the surface and attributes within the voxel. Our voxels form a continuous field interpolated through space and scales, through which we cast conic rays. Within the ray marching steps, we compute the occlusion distribution produced by the macro-surface inside a pixel footprint, we use the microdescriptors to reconstruct light- and view-dependent shading, and we combine fragments in an A-buffer way. Our representation efficiently accounts for various subpixel effects. It can be continuously interpolated and filtered, it is scalable, and it allows for efficient depth-of-field. We illustrate the quality of these various effects by displaying surfaces at different scales, and we show that the timings per pixel are scale-independent (Figure 16).

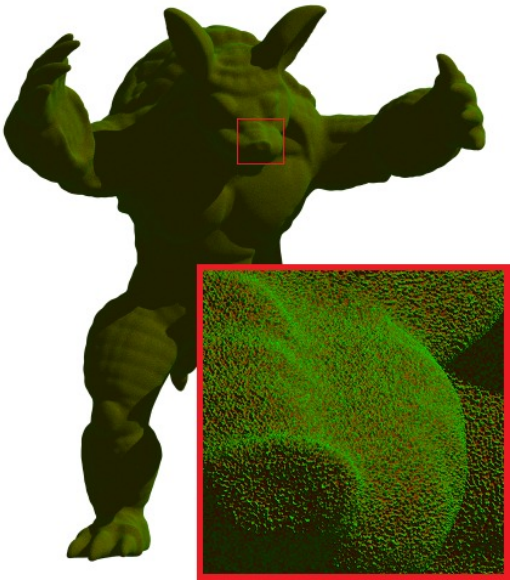


Figure 16. Our method allows for correct filtering of color variations, like anti-aliasing demonstrated here.

MIMETIC Team

6. New Results

6.1. Motion Sensing and analysis

Participants: Franck Multon [contact], Richard Kulpa, Anthony Sorel, Edouard Auvinet.

Sensing human activity is a very active field of research, with a wide range of applications ranging from entertainment and serious games to personal ambient living assistance. MimeTIC aims at proposing original methods to process raw motion capture data in order to compute relevant information according to the application.

In personal ambient living monitoring, we have collaborated with University of Montreal, Department of Computer Science and Operations Research (DIRO) which main activity is biomedical engineering. A co-supervised student is addressing two complementary problems: detecting people falling in everyday environment and providing easy-to-use clinical gait analysis systems for early detection of potential risks of falling. In the last decade, gait analysis has become one of the most active research topics in biomedical research engineering partly due to recent developpement of sensors and signal processing devices and more recently depth cameras. The latters can provide real-time distance measurements of moving objects. In this context, we present a new way to reconstruct body volume in motion using multiple active cameras from the depth maps they provide. A first contribution of this paper is a new and simple external camera calibration method based on several plane intersections observed with a low-cost depth camera which is experimentally validated. A second contribution consists in a body volume reconstruction method based on visual hull that is adapted and enhanced with the use of depth information. Preliminary results based on simulations are presented and compared with classical visual hull reconstruction. These results show that as little as three low-cost depth cameras can recover a more accurate 3D body shape than twenty regular cameras (see figure 4).

In entertainment and serious games, the problem is different as we need to accurately now the action performed by the user in order to react in a convenient manner. Collaboration with Artefacto Company enabled us to develop such motion recognition methods in serious games scenarios. Given motion capture data provided by an optical motion capture system lead to large state vectors in which the relevant information is hidden. Mixture of Gaussians is generally used as an input of Hidden Markov Models to recognize a motion according to this raw data. To simplify, features are generally introduced in order to capture the relevant geometrical property of the motion with either general information (such as joint angles or Cartesian positions) or application-specific information. The former type of information has the advantage to be generic but leads to recognizers that are very sensitive to style and morphology variations. Previously, we have proposed a new generic feature based on morphology-independent representation that enables to tackle this problem [28]. We now have explored the robustness of this type of features for early recognition, when using mixture of Gaussians instead of Hidden Markov Models. We have shown that a motion can be recognized when only 50% of the motion is performed. The recognition rate is especially high with this type of feature compared to classical Euler angles and Cartesian data, especially when a new user is performing the motion [6].

6.2. VR and Sports

Participants: Richard Kulpa [contact], Benoit Bideau, Sébastien Brault, Anne-Marie Burns.

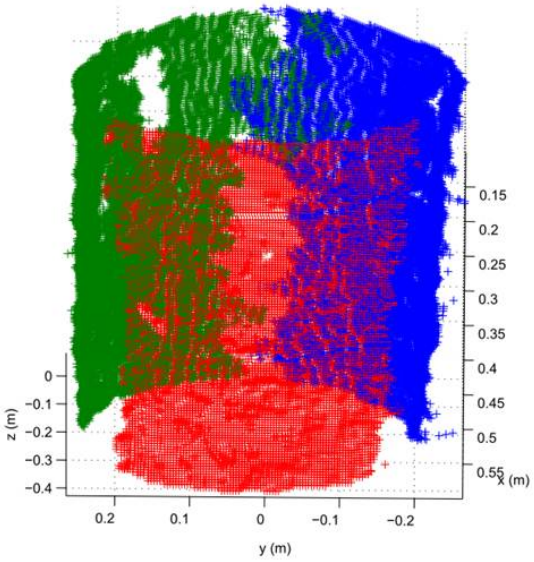


Figure 4. 3D silouhettes reconstructed with three depth-cameras - reconstructed points of a reference cylinder. Each color corresponds to one of the depth camera.

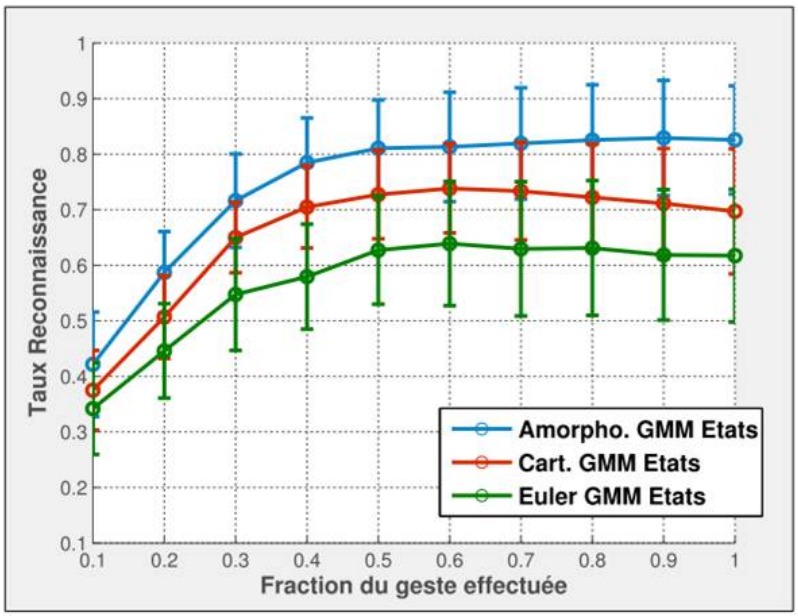


Figure 5. Early recognition of a motion performed by a new user with three different features: Cartesian, Euler and the proposed amorphological features.

In the past, we have worked on the interaction between two opponents in virtual environment. These duels were between a handball goalkeeper and a thrower; and between a rugby defender and an attacker performing deceptive movements. Even if these sports applications are different in terms of kinematic parameters, information picked-up and type of interaction, we have designed a unique framework to simulate such duels in a reality center and to analyze the gestures of real athletes immersed in this environment. This VR framework was validated by showing that behaviors in real and virtual environments were similar. These works have been extended by using perception-action coupling and perception-only studies to evaluate the anticipation of opponents. In order to evaluate the importance of perceived parameters, the ball and/or the character animation was successively hidden to determine their importance and the same kind of study was done on the graphical level of details.

This year, we have addressed the problem of the tennis serve. The first step is the PhD of Caroline Martin who will end next year. This work provides biomechanical analysis of the serve and the influence of the kinematical and dynamic parameters on performance. Thanks to an accepted project funded by the INSEP institute, we are importing this biomechanical model to virtual environment to make perceptual analysis. This work is based on the same methodology used for the detection of deceptive movements in rugby. The next step is to combine the use of cutoffs with biomechanical analysis to extract important kinematic information that could explain differences between experts and novices. This information is then correlated to kinematical parameters of this player. Concurrently, we are working on the creation of models of rugby defenders based on the results of the previous perceptual analyses

Finally, we have worked on the use of virtual environments to train athletes. The first step was to evaluate if a better score in the virtual environment implied only an improvement of the athlete in the virtual game or also a better performance back on the field. The PhD of Anne-Marie Burns has demonstrated that the improvement of training based on virtual environment was similar to training with a real teacher or based on videos. The use of VR for sports training, at least by imitation, is thus possible. Furthermore, we have explored the influence of the self-representation of the immersed learner by displaying his avatar as if he was in front of a virtual mirror. We made both kinematical and evocation analyses. The results do not show significant difference with or without the use of the mirror and it is confirmed by the subjective analysis that shows that the use of the virtual mirror by immersed athletes was limited. This work was partially funded by the Biofeedback project.

6.3. Biomechanics and Motion Analysis

6.3.1. Interaction strategies between two walkers to avoid collision

Participants: Armel Crétual, Julien Pettré, Anne-Hélène Olivier, Antoine Marin.

Walkers are extremely efficient in avoiding collisions, even in relatively condition of density. We experimentally addressed two questions. What are the conditions for walkers to perform adaptations to their trajectory, and second, how avoidance performed in time. We checked several hypothesis, that led to two contributions, as presented in [15]. First, human are able to anticipate the future conditions of an interactions and the distance they would meet. They react accordingly, i.e., if and only if a future risk of collision can be predicted. Second, we demonstrated that the avoidance is performed with anticipation, i.e., avoidance maneuvers are over before walkers get at closest distance.

6.3.2. Quantification of pathological motion

Participant: Armel Crétual.

In clinical routine, precise quantification of patients' gesture remains a challenge. Several simple means are daily used by practicians in physical medecine. Their main drawback is often a large inter-operator variability and even sometimes an intra-operator one. To overcome this, we have developed and validated still simple to remain usable) but much more objective tools in two different fields: gait and shoulder laxity.

First, we have proposed a new index of gait quantification based on EMG profiles called KeR-EGI (for Kerpape-Rennes EMG-based Gait Index). Our recent works allowed us to demonstrate its reproducibility even in patients with severe troubles. Moreover, we have also demonstrated the complementarity of this index based on muscular activation and an index based on kinematics, the Edinburgh Visual Gait Score (EVGS) that can be computed easily from a simple video recordings of the patient's gait. Indeed, we have shown that the relationship between these indices depends on the fact that pathology is congenital or acquired. Using both indices at the time, allows to evaluate the potential kinematics compensation the patient does to improve his/her gait despite a damaged motor control.

Secondly, in shoulder surgery, the surgeon has to choose between different protocols depending on whether the patient is hyperlax or not. Until now, shoulder laxity is very roughly evaluated without actual measurement and above all mobilizing only one axis (external rotation) of this complex joint. By measuring precisely the whole Range Of Motion of 28 subjects recruited to ensure a large spectrum of laxity (from hypo to hyper-laxity), we have shown that the usual clinical indices fail to actually classify subjects, as they do focus on only one dimension of mobility. From, that result, we have then proposed a new method to evaluate laxity that remains simple and usable in daily routine but that takes into account all dimensions of shoulder's mobility.

6.3.3. Modeling gesture in sports: fin swimming

Participants: Nicolas Bideau, Guillaume Nicolas, Benoit Bideau, Richard Kulpa.

In swimming, experimental approaches are commonly used to analyze performance. However, due to obvious limitations in experimental approaches (impoossibility to standardize any situations etc.), it is difficult to characterize surrounding fluid. To overcome this limitation, we currently develop analysis, modeling and simulation of aquatic locomotion, using CFD computer simulation and new methods based on animation of virtual characters.

- A first application of this topic enables to evaluate the influence of swim fin flexibility on efficiency during swimming based on a CFD structure interaction model. Finite elements simulations are carried out for various material properties and various prescribed kinematics. Besides the significant effect of flexibility on propulsive forces, the results indicate that the propulsive efficiency is greatly influenced by the stroke frequency and the initial angle of attack. For the selected material properties, the results show that efficiency increases from 3.6 percents to 11.9 percents when the stroke frequency is increased from 0 to 1.7 Hz. Moreover efficiency is clearly increased from 5.0 percents to 24.2 percents when increasing the angle of attack from 0 to 45 degrees. Therefore, an interesting prospect of the present work could be an enhancement of the design of better performing swim fins.
- A second application of this topic related to aquatic propulsion deals with a new method to evaluate cross-sectional area based on computer animation of swimming. Indeed, reducing cross sectional area (CSA) during starts and turns is a key part of performance optimisation. Different methods have been used to obtain this parameter without any standard: total human body volume to the power 2/3, wetted area or frontal area based on planimetry technique (PT). These different methods can lead to discrepancies in drag values. Recently, we used two synchronized camcorders to evaluate drag parameters during the different phases of an undulatory stroke cycle. However, such a technique needs accurate synchronization and calibration of the different camcorders views. The aim of this study is to provide a new method based on animation of virtual characters to obtain instantaneous cross-sectional area in an undulatory stroke cycle. Its main advantage is to obtain cross-sectional area as well as biomechanical analysis with a single camcorder in a sagittal plan and without space calibration. A camcorder placed side-on to the swimmer recorded the undulatory movements in the sagittal plane of eight swimmers. This information provided the angles between limbs. These data were then used by our animation engine to animate a virtual swimmer whose anthropometric data came from the real swimmer. A specific algorithm has been developed to automatically obtain the CSA using body outlines. In order to validate our method, we also calculated the CSA using PT with a frontal camcorder view of the same undulatory movements. Our results show similar values of maximum CSA using PT and the frontal camcorder view and our algorithm based on 3D animation. The mean coefficient of variation between the results obtained from the two methods is

7.3 percents. This difference could be related to the level of details of the mesh used to model the avatar. One prospect to this work is to take resistive and propulsive body segments into account in CSA calculation. From this method, we intend to better understand swimming hydrodynamics and the way CSA influences active drag. More generally, this approach has been designed to provide new practical insights into swimming analysis protocols.

6.4. Crowds

Participants: Julien Pettré [contact], Richard Kulpa, Anne-Hélène Olivier, Samuel Lemercier, Jonathan Perrinet, Kevin Jordao.

6.4.1. A realistic model of following behaviors in crowds

Following is an important type of interactions between individuals in crowds. In uni- or bidirectionnal pedestrian traffic, density prevent people from overtaking and going through the crowd: they just start following each other. Based on some experiments performed in the frame of the national project ANR-PEDIGREE, we elaborated a model for simulating following behavior with a very high level of realism. Contributions were presented in [9]. Especially, realism was evaluated both at the microscopic scale and at the macroscopic scale. At the microscopic scale, we carefully reproduce how human do control their motion to follow another walker. At the macroscopic scale, we focused on the emergence of stop-and-go waves that emerge from such traffic. Detailed analysis of experimental data analysis is described in 2 papers in Physical Review E: [15] and [35].

6.5. Interactive Virtual Cinematography

Participants: Marc Christie [contact], Christophe Lino.

The domain of Virtual Cinematography explores the operationalization of rules and conventions pertaining to camera placement, light placement and staging in virtual environments. In 2012, we have tackled two key issues in relation to the reactive control of virtual cameras: (i) the design of an efficient occlusion-free target tracking technique in dynamic environments and (ii) the design of a novel composition technique based on a 2D-manifold representation of search space.

The first issue is related to maintaining the visibility of target objects, a fundamental problem in automatic camera control for 3D graphics applications. Practical real-time camera control algorithms generally only incorporate mechanisms for the evaluation of the visibility of target objects from a single viewpoint, and idealize the geometric complexity of target objects. Drawing on work in soft shadow generation, we perform low resolution projections, from target objects to rapidly compute their visibility for a sample of locations around the current camera position. This computation is extended to aggregate visibility in a temporal window to improve camera stability in the face of partial and sudden onset occlusion. To capture the full spatial extent of target objects we use a stochastic approximation of their surface area. Our implementation is the first practical occlusion-free real-time camera control framework for multiple target objects. The result is a robust component that can be integrated to any virtual camera control system that requires the precise computation of visibility for multiple target (see [20]).

The second challenge is related to the automatic positioning a virtual camera in a 3D environment given the specification of visual properties to be satisfied (on-screen layout of subjects, vantage angles, visibility) is a complex and challenging problem. Most approaches tackle the problem by expressing visual properties as constraints or functions to optimize, and rely on computationally expensive search techniques to explore the solution space. We have shwon how to express and solve the exact on-screen positioning of two or three subjects by expressing the solution space for each couple of subjects as a 2D manifold surface [23]. We demonstrate how to use this manifold surface to solve Blinn's spacecraft problem with a straightforward algebraic approach. We extend the solution to three subjects and we show how to cast the complex 6D optimization problem tackled by most contributions in the field in a simple 2D optimization on the manifold surface by pruning large portions of the search space. The result is a robust and very efficient technique which finds a wide range of applications in virtual camera control and more generally in computer graphics.

We have also explored the application of automated editing techniques to Machinema [19].

Besides we have been involved in the process of rendering camera motions (from real movies) using haptic devices (a joint work with Technicolor and VR4i, accepted at VRST 2012 [21]), and have authored a state of the art report on Haptic Audiovisual (published in Transactions on Haptics [8]).

6.6. Autonomous Virtual Humans

6.6.1. Unifying activity scheduling and path-planning

Participants: Carl-Johan Jorgensen, Fabrice Lamarche [contact].

Crowd distribution in cities highly depends on how people schedule their daily activities. This schedule depends on temporal constraints like appointments or shops opening times. It also relies on the city structure and the locations of the places where activities can be achieved. Personal preferences also affect this schedule: choosing favorite shops or paths for instance.

Within the framework of iSpace&Time project, we are currently working on a model that unifies activity scheduling and path planning into a single process. This process takes city topological configuration into account, as well as time constraints and personal preferences. Applied to thousands of agents, his approach allows us to credibly populate cities. Credible flows of people automatically emerge depending on the time of the day and the city topology.

6.6.2. Long term planning and opportunism

Participants: Philippe Rannou, Fabrice Lamarche [contact].

Autonomous virtual characters evolve in dynamic virtual environments in which changes may be unpredictable. However, they need to behave properly and adapt their behavior to perceived changes while fulfilling their goals. We propose a system that combines long term action planning with failure anticipation and opportunism [27]. The system is based on a modified version of an HTN planning algorithm. It generates plans enriched with information that enable a monitor to detect relevant changes of the environment. Once those changes are detected, a plan adaptation is triggered. Such adaptations include modifying the plan to react to a predicted failure and more importantly to exploit opportunities offered by the environment.

6.6.3. Space-Time planning in dynamic environments

Participants: Thomas Lopez [contact], Fabrice Lamarche [contact].

When automatically populating 3D geometric databases with virtual humanoids, modeling the navigation behavior is essential since navigation is used in most exhibited behaviors. In many application fields, the need to manage navigation in dynamic environments arises (virtual worlds taking physics laws into account, numerical plants in which step stools can be moved,...). This study focuses on the following issue: how to manage the navigation of virtual entities in such dynamic environments where topology may change at any time i.e. where unpredictable accessibility changes can arise at runtime. In opposition to current algorithms, movable items are not only considered as obstacles in the environment but can also help virtual entities in their navigation.

The proposed algorithm [10] splits that problem into two complementary processes: a topology tracking algorithm and a path planning algorithm. The aim of the topology tracking algorithm is to continuously detect and update topological relations between moving objects i.e. accessibility or obstruction, while storing temporal information when recurring relations are observed. The path planning algorithm uses this information to plan a path inside the dynamic environment. The coupling of those algorithms endows a virtual character with the ability to immediately use inserted / moved object to reach previously unreachable locations. Moreover, this algorithm is able to find a path through moving platforms to reach a target located on a surface that is never directly accessible.

MINT Project-Team

6. New Results

6.1. Noisy input filtering for interactive systems

Participants: Géry Casiez [correspondant], Nicolas Roussel.

Noisy signals occur when an original time varying value undergoes undesirable and unpredictable perturbations. These may be caused by things like heat and magnetic fields affecting hardware circuitry, the limits of sensor resolution, or even unstable numerical computation. Noisy signals are a common problem when tracking human motion, particularly with custom sensing hardware and inexpensive input devices like the Kinect or Wiimote.

We developed the 1€ filter ("one Euro filter") is a simple algorithm to filter noisy signals for high precision and responsiveness. It uses a first order low-pass filter with an adaptive cutoff frequency: at low speeds, a low cutoff stabilizes the signal by reducing jitter, but as speed increases, the cutoff is increased to reduce lag. The algorithm is easy to implement, uses very few resources, and with two easily understood parameters, it is easy to tune. When compared with other filters, the 1€ filter shows less lag for a reference amount of jitter reduction [15].

The 1€ filter is already used on a daily basis by many other researchers and companies.

6.2. Transfer functions for subpixel interaction

Participants: Jonathan Aceituno, Géry Casiez [correspondant], Nicolas Roussel.

The current practice of using integer positions for pointing events artificially constrains human precision capabilities (Figure 1). The high sensitivity of current input devices can be harnessed to enable precise direct manipulation "in between" pixels, called subpixel interaction. In [23], we provide a detailed analysis of subpixel theory and implementation, including the critical component of revised control-display gain transfer functions. A prototype implementation is described with several illustrative examples. Guidelines for subpixel domain applicability are provided and an overview of required changes to operating systems and graphical user interface frameworks are discussed.

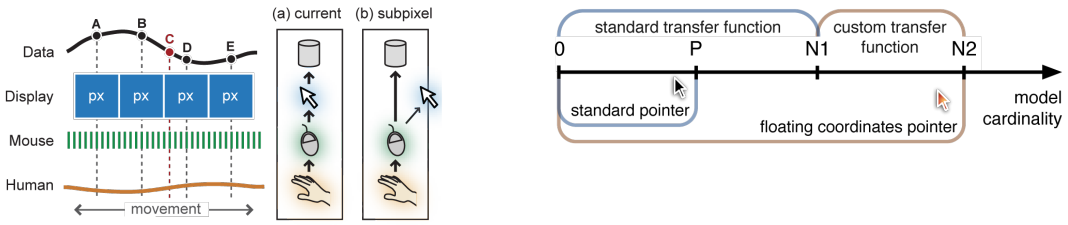
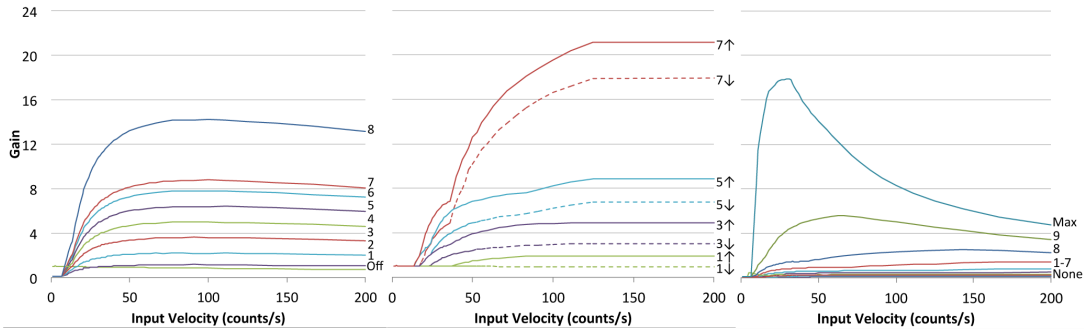


Figure 1. Input mappings: (a) currently, human movements are discretized by mouse sensitivity, then again by display density: data points "in between" pixels like 'C' are unreachable; (b) a subpixel mapping discretizes human movements by mouse sensitivity only, for precise data manipulation (left). Four zones of applicability for subpixel and custom transfer functions (see text for description) (right).

6.3. Transfer functions for scrolling tasks

Participants: Géry Casiez [correspondant], Nicolas Roussel.

Scrolling is controlled through many forms of input devices, such as mouse wheels, trackpad gestures, arrow keys, and joysticks. Performance with these devices can be adjusted by introducing variable transfer functions to alter the range of expressible speed, precision, and sensitivity. However, existing transfer functions are typically "black boxes" bundled into proprietary operating systems and drivers. This presents three problems for researchers: (1) a lack of knowledge about the current state of the field; (2) a difficulty in replicating research that uses scrolling devices; and (3) a potential experimental confound when evaluating scrolling devices and techniques. These three problems are caused by gaps in researchers' knowledge about what device and movement factors are important for scrolling transfer functions, and about how existing devices and drivers use these factors (Figure 2). We fill these knowledge gaps with a framework of transfer function factors for scrolling, and a method for analysing proprietary transfer functions demonstrating how state of the art commercial devices accommodate some of the human control phenomena observed in prior studies [22].



(a) Mac OS X (Generic Device). (b) Microsoft IntelliPoint: up (solid) and down (dashed). (c) Logitech Control Center.

Figure 2. Gain scale factors across input velocity (counts per second) with Mac OS X, Microsoft IntelliPoint (under Windows 7), and Logitech drivers under Mac OS X. Gain is measured as the level of amplification in the system's base unit (pixels per count for Mac OS X and Logitech; lines per count for Microsoft IntelliPoint), and is plotted at varying levels of each driver's respective UI sliders for acceleration.

6.4. Design of transparent tactile stimulators

Participants: Michel Amberg, Frédéric Giraud, Betty Lemaire-Semail [correspondant].

Friction reduction based tactile devices are able to modulate the friction between the fingertip and the active touched surface as a function of fingertip's position. This type of tactile stimulator is thus based on two main components: an active area which vibrates and produces a squeezed air film bearing and a position sensor. Our previous design was made up with a copper plate fully covered by piezo cells, a material which bent when energized by a voltage.

However, this design no longer makes sense when we look forward using tactile feedback on a transparent display. Indeed, for co-localized operation, we can't place piezo cells on the bottom surface of a touch screen since the touched surface would not be transparent; moreover, glass is a non conductive material which complicates the electrical connection.

To cope with these problems, a new design has been introduced. Two copper exciters are firmly bonded on the touch screen to obtain the vibration. These exciters vibrate and propagate their vibration to the glass touch screen. To be efficient, the size of the exciters has to be perfectly adapted to the glass plate. This is why, we not only propose a new way to obtain the vibration of the active area, but we also provide the key design rules of the exciters[19].

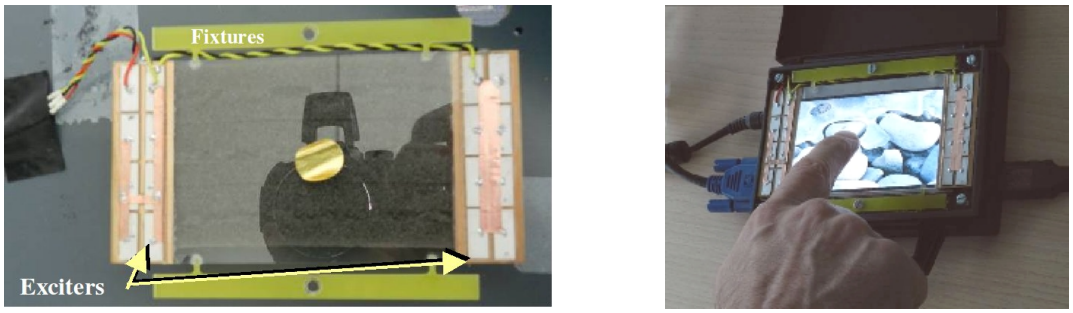


Figure 3. The transparent tactile display, during test procedure (left) and in a co-localized operation (right).

6.5. Methodology for developing textures on friction based interfaces

Participants: Géry Casiez, Thomas Pietrzak, Ludovic Potier, Nicolas Roussel [correspondant], Ibrahim Yapici.

The design of textures for so-called variable friction technologies requires multiple perspectives, which this paper aims to outline and discuss. We first propose a definition of texture and describe the current state of knowledge on their perception. After presenting two technologies for variable friction and comparing them to other tactile interfaces, we describe several particular uses for these devices (Figure 4). We then discuss psychophysical methods for signal perception evaluation and finally discuss methodologies for creating multidimensional tactile content [26].

Figure 4. Examples of textures with increasing complexity in one dimension.

6.6. Hand occlusion on mutitouch surfaces

Participant: Géry Casiez [correspondant].

Operating a computer by directly touching the display surface has many benefits, and in tabletop computing, multi-touch is arguably the most natural form of input. However, with any form of direct input, where the input device and the output display are coincident, the hand and arm cover - or occlude part - of the display. This can be a problem, because compared to manipulating objects on a real tabletop, a tabletop computer is dynamic and can display relevant information, sequential widgets, and system messages in occluded areas.

We examined the shape of hand and forearm occlusion on a multi-touch table for different touch contact types and tasks. Individuals have characteristic occlusion shapes, but with commonalities across tasks, postures, and handedness. Based on this, we create templates for designers to justify occlusion-related decisions and we propose geometric models capturing the shape of occlusion. A model using diffused illumination captures performed well when augmented with a forearm rectangle, as did a modified circle and rectangle model with

ellipse "fingers" suitable when only X-Y contact positions are available (Figure 5). Finally, we describe the corpus of detailed multi-touch input data we generated which is available to the community [24].

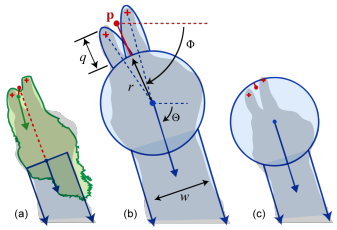


Figure 5. Three occlusion shape models: (a) DI and rectangle; (b) multi-touch circle and rectangle; (c) Vogel et al.

6.7. Indirect multitouch interaction on large screens

Participants: Géry Casiez [correspondant], Jérémie Gilliot, Nicolas Roussel.

Multitouch interaction shows its limits with large display surfaces. Indirect interaction allows to use control surfaces that are much smaller than display surfaces. Absolute indirect interaction raises accuracy problems and relative indirect interaction only allows to interact with a single cursor. We present a relative indirect multitouch interaction technique allowing to create, control, delete several cursors without sacrifying precision for interacting with small objects (Figure 6) [25].

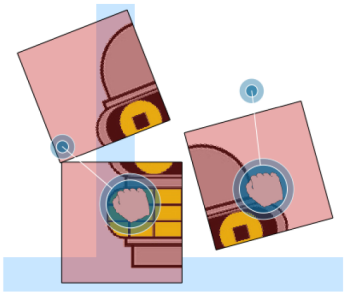


Figure 6. Overview of cursors and cursels used to manipulate two objects.

6.8. Pseudo-rigid movements for flexible multi-finger interactions

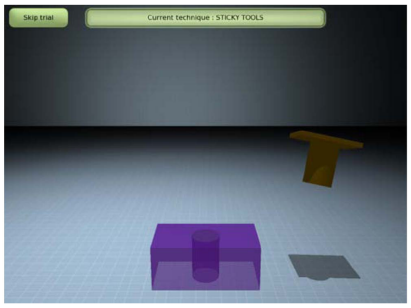
Participants: Laurent Grisoni [correspondant], Yosra Rekik, Nicolas Roussel.

Multi-touch interaction requires a trade-off between users' desires and capabilities and gesture recognition constraints. Current approaches to that problem lack flexibility. The number of fingers used for a gesture usually plays a key part in the recognition process, for example. To increase the flexibility of this process, we proposed the use of *pseudo-rigid movements* [27]. We showed how these movements can be determined in real time from the contact information usually available. We explained how they allow to free the recognition process from the number of fingers used and to move towards multi-movement gestures, independent or coordinated. We also presented an interaction technique that takes advantage of this increased flexibility.

6.9. 3D manipulation on multitouch displays

Participants: Anthony Martinet, Géry Casiez [correspondant], Laurent Grisoni.

Multitouch displays represent a promising technology for the display and manipulation of data. While the manipulation of 2D data has been widely explored, 3D manipulation with multitouch displays remains largely unexplored. Based on an analysis of the integration and separation of degrees of freedom, we propose a taxonomy for 3D manipulation techniques with multitouch displays. Using that taxonomy, we introduce Depth-Separated Screen-Space (DS3), a new 3D manipulation technique based on the separation of translation and rotation. In a controlled experiment, we compared DS3 with Sticky Tools and Screen-Space. Results show that separating the control of translation and rotation significantly affects performance for 3D manipulation, with DS3 performing faster than the two other techniques [13].



		Translation			Rotation		
	Mode	Tx	Ty	Tz	Rx	Ry	Rz
DS3	1*d*	○	○				
	1*d* + 1*i*	○	○	*i*	Z-Technique		
	≥ 2*d*				○	○	○
	≥ 2*d* + 1*i*			*i*	○	○	○

Figure 7. Screen capture of the peg-in-hole task used in the experiment (left). Description of the DS3 technique using the proposed taxonomy (right).

6.10. 3D navigation on multitouch displays

Participants: Clément Moerman, Damien Marchal [correspondant], Nicolas Roussel.

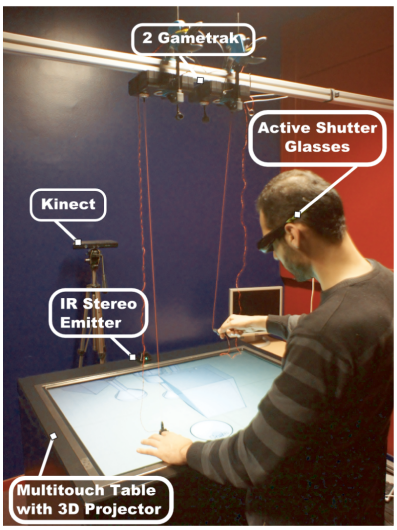
Navigation is one of the elementary tasks of 3d virtual environment. It is composed of two parts: locomotion where there is a physical control of the camera and the wayfinding where a path is found through the environment. Despite being widely studied, there is still need for more efficient and intuitive techniques especially for novice users. Within the context the I-Lab, we worked on a new locomotion technique that combines the advantages of multi-scale navigation and of direct manipulation (Figure 8). The technique, called *Drag'n Go*, was evaluated with a comparative experiment against three other techniques. The results show that *Drag'n Go*: improves performances, reduces learning time and get good user satisfaction either from novice and expert users. The approach and the associated experiment are published in [20].

6.11. Modeling on and above a multitouch surface

Participants: Géry Casiez [correspondant], Bruno De Araujo.

We introduced a semi-immersive environment for conceptual design where virtual mockups are obtained from gestures we aim to get closer to the way people conceive, create and manipulate three- dimensional shapes. We presented on-and-above-the-surface interaction techniques following Guiard's asymmetric bimanual model to take advantage of the continuous interaction space for creating and editing 3D models in a stereoscopic environment. To allow for more expressive interactions, our approach continuously combines hand and finger tracking in the space above the table with multi-touch on its surface. This combination brings forth an alternative design environment where users can seamlessly switch between interacting on the surface or in the space above it depending on the task (Figure 9). Our approach integrates continuous space usage with bimanual interaction to provide an expressive set of 3D modeling operations. Preliminary trials with our experimental setup show this as a very promising avenue for further work [17], [16].

Figure 8. With the Drag'n Go method user can navigate in a 3D virtual environment with a multi-touch device. The movement speed is calculated using perspective based progression scale and it is let under the user's control.



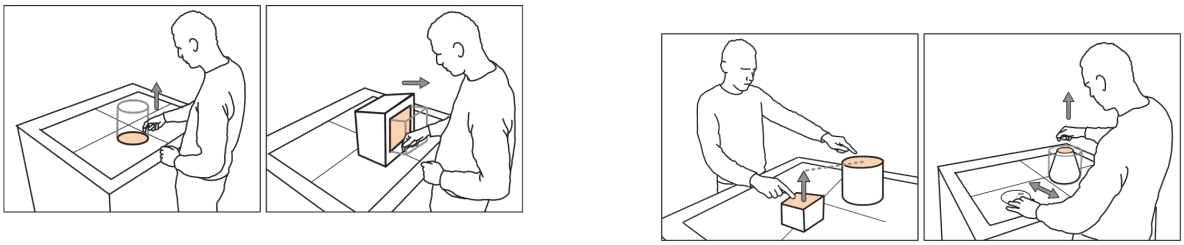


Figure 9. Overview of MockupBuilder setup (left). Examples of face straight extrusion, height constraint and scaling (right).

6.12. Paper-based annotation of digital content from a mobile device

Participant: Thomas Pietrzak [correspondant].

S-Notebook is a hybrid system that makes it possible to take notes on paper about digital content one is exploring on a mobile device (Figure 10). The user can link notes on paper with the content he is currently interacting with, making it possible to reopen it at a later time by tapping the note on his notebook with the digital pen. Therefore he can create bookmarks and hyperlinks on paper notes [21].

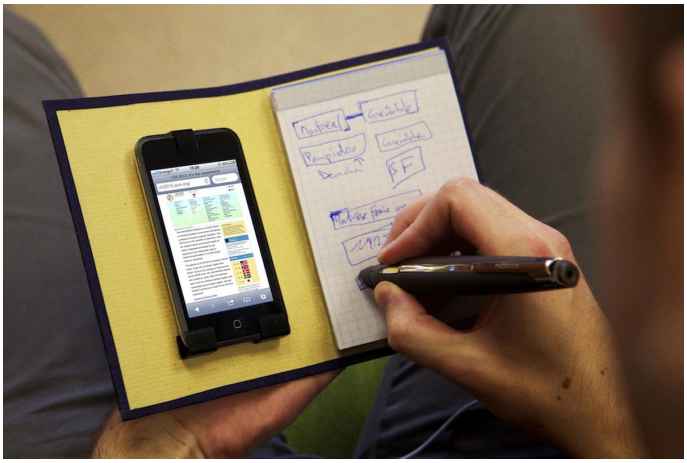


Figure 10. Annotation of digital content on paper.

POTIOC Team

6. New Results

6.1. Spatial augmented reality for physical drawing

Participants: Jérémy Laviole, Martin Hachet.

Spatial augmented reality (SAR) promises the integration of digital information in the real (physical) world through projection. We proposed different tools to improve speed or ease the drawing by projecting photos, virtual construction lines and interactive 3D scenes (published in the 3DUI symposium [16]). We explored the creation of tools which help to create drawings that are "difficult" to achieve for a human being, such as stereographic drawings (published in the 3DCHI CHI workshop [18]). Through these tools, we want to apply existing computer graphics techniques to enhance existing drawing tools, and to use it to teach how to draw. Furthermore, we proposed some insights for the creation of digital games and programs which can take full advantages of physical drawings (published in the UIST doctoral symposium [17]).

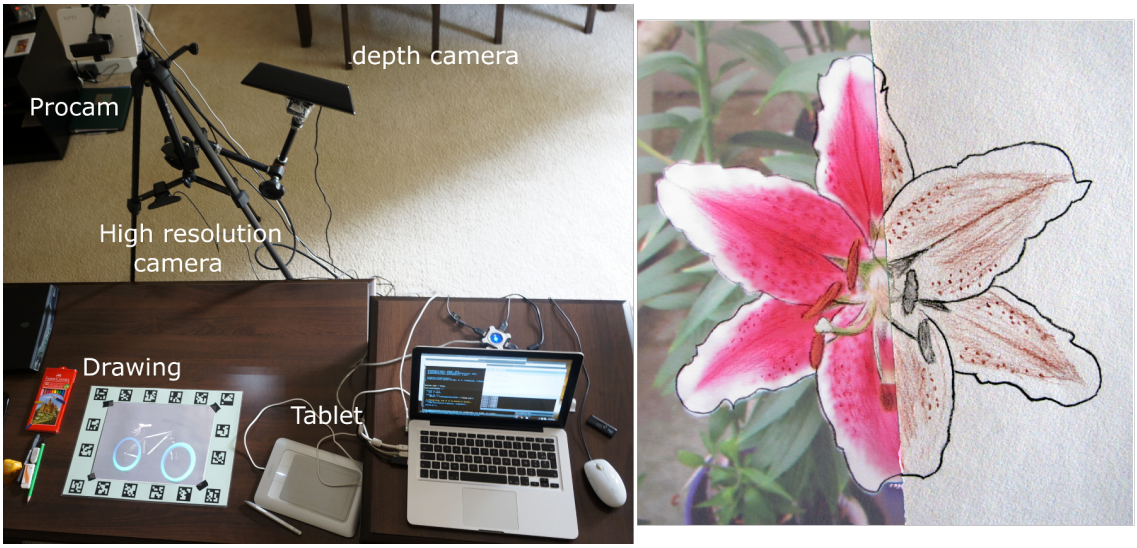


Figure 3. Left: Spatial augmented reality system for physical drawing. Right: Projection of a source image on overlay of an actual drawing, to teach drawing thanks to computer graphic tools [16].

6.2. Brain-Computer Interfaces

Participants: Fabien Lotte, Florian Larrue, Martin Hachet.

As part of our research on Brain-Computer Interfaces (BCI), our contributions addressed two different levels: 1) the brain signal processing level, in order to design more efficient BCI systems and 2) the applications level, in order to propose and explore new BCI applications.

At the signal processing level, we explored and designed new features to represent ElectroEncephaloGraphic (EEG) signals. In particular we explored multifractal cumulants and predictive complexity features (which we published in the Neurocomputing journal [5]), as well as waveform length features together with an optimal spatial filter that we designed for such features (which we published in the ICPR international conference [19]). All these features proved useful to classify EEG signals, and, more importantly, increased the classification performances of the system when combined together with the gold standard features, namely, band power features. Thus, this contributed to extending the repertoire of features available to BCI designers as well as increasing BCI performances. Nevertheless, our studies of BCI and educational research led us to the conclusion that current BCI feedback training approaches (which aimed at teaching people how to use a BCI and how to control their own brain activity), are most probably highly inappropriate and one of the major causes for the limited performances of current BCI - maybe more than signal processing methods. We therefore stressed the need for alternative feedback training approaches for BCI in a publication at the international BBCI workshop [20].

At the application level, we mostly focused on Virtual Reality (VR) related applications. Indeed, together with other groups in the field, we reviewed how BCI and VR could be combined in order to give rise to new applications and to improve BCI designs. This was published in a book chapter dedicated to BCI [22]. Similarly, with international colleagues, we reviewed and envisioned new applications of BCI outside the medical domain, and proposed guidelines to move towards these new applications. This notably includes VR and game applications, user-state monitoring, neuro-evaluation, training and education, cognitive improvement as well as safety and security. This was published in the IEEE Computer journal [8]. Finally, we proposed a new and innovative application of BCI: using it as a tool to study spatial cognition and transfer from VR to real environments. In particular, since BCI can be used to navigate a Virtual Environment (VE) without any motor activity, BCI can be used to assess how much motor activity is really needed to transfer spatial knowledge from a VE to a real one. This is what we did by comparing a BCI and a treadmill in order to teach users a path in a VE and then asking them to retrieve this path in the real world. Contrary to what was believed before, our results showed that motor activity is not necessary to learn a path in VR. We showed that what is really necessary is performing an action, but that this action does not have to be motor, and can be, for instance, cognitive (e.g., imagining hand movements), with a BCI. This was published in the VRST international conference [14].

6.3. Understanding user gestures for touch screen-based 3D User Interfaces

Participants: Aurélie Cohé, Martin Hachet.

In the scope of the ANR project Instinct, we studied how users tend to interact with a touchscreen for interacting with 3D content. Our main contributions were to study user behaviors with a standard touchscreen on the one hand, and with a pressure sensitive touchscreen on the other hand.

Multi-touch interfaces have emerged with the widespread use of smartphones. Although a lot of people interact with 2D applications through touchscreens, interaction with 3D applications remains little explored. Most 3D object manipulation techniques have been created by designers who have generally put users aside from the design creation process. We conducted a user study to better understand how non-technical users tend to interact with a 3D object from touchscreen inputs. The experiment has been conducted while users were manipulating a 3D cube with three viewpoints for rotations, scaling and translations (RST). Sixteen users participated and 432 gestures were analyzed. To classify data, we introduce a taxonomy for 3D manipulation gestures with touchscreens. Then, we identify a set of strategies employed by users to perform the proposed cube transformations. Our findings suggest that each participant uses several strategies with a predominant one. Furthermore, we conducted a study to compare touchscreen and mouse interaction for 3D object manipulations. The results suggest that gestures are different according to the device, and touchscreens are preferred for the proposed tasks. Finally, we propose some guidelines to help designers in the creation of more user friendly tools. This work was published in the Graphics Interface (GI) conference [12] as well as in the Computers and Graphics journal [6].

Figure 4. A user navigating a virtual model of the city of Bordeaux with a BCI, in order to learn a specific path [14].

Moreover, few works have focused on the relation between the manipulated data and the quantity of force applied with the fingers sliding on a touch sensor. In another work, we conducted two user studies to better understand how users manage to control pressure, and how they tend to use this input modality. A first set of experiments allows us to characterize pressure in relation to finger motions. Based on the results of this study, we designed a second set of experiments focusing on the completion of 3D manipulation tasks from 2D gestures. The results indicate that a strong relationship exists between the actions the participants intend to perform, and the quantity of force they apply for 3D object manipulations. This finding opens new promising perspectives to enhance user interfaces dedicated to force-based touch sensors.

All these works were published in the PhD thesis of Aurélie Cohé [4], which was defended on December 13th, 2012.

6.4. Virtual reality for Musical Performance

Participants: Florent Berthaut, Martin Hachet.

Immersive virtual environments open new perspectives for music interaction, notably for the visualization of sound processes and of musical structures, for the navigation in musical compositions, for the manipulation of sound parameters and for musical collaboration. Research conducted by Florent Berthaut and Martin Hachet, in collaboration with Myriam Desainte-Catherine from the SCRIME/LaBRI, explore these new possibilities.

Among the current projects, development of the Drile immersive virtual musical instrument was pursued in order to enable various scenographic setups that will be evaluated in the context of public performance. New perspectives for the Tunnels, 3D widgets for musical modulation (see Figure 6), were published as a Poster in the Proceedings of the Symposium on 3D User Interfaces (3DUI) [10]. Novel 3D selection techniques that take music interaction constraints into account are also being designed.

Another project was conducted with David Janin and Benjamin Martin from the LaBRI on new musical models that will be used to improve the hierarchical musical structures manipulated with Drile. It was published in the International Conference on Semantic Computing [11].

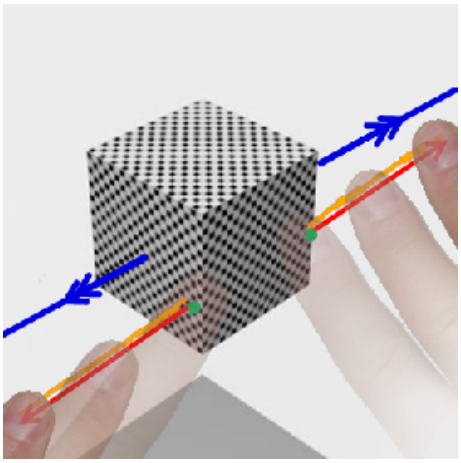


Figure 5. Analysis of users' gestures on touch screen to manipulate 3D content [6].

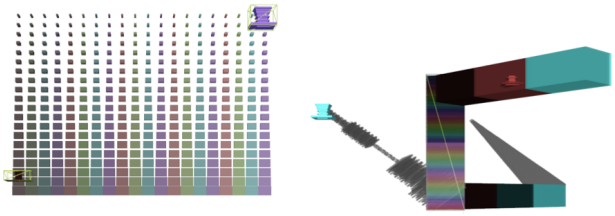


Figure 6. The Tunnels 3D widgets for musical modulation.

A collaboration was started with researchers of the Center for Computer Research on Music and Acoustics (CCRMA) of Stanford University. Florent Berthaut was invited for two months at CCRMA, where he worked with Luke Dahl and Chris Chafe on the implementation of musical collaboration modes in immersive virtual environments. A first result is the design of 3D musical collaboration widgets for Drile, which will be evaluated with musicians.

Another project was initiated with researchers of the Bristol Interaction and Graphics group of the University of Bristol. This project aims at improving the audience experience with Digital Musical Instruments (DMIs). These instruments are often confusing for spectators because of the variety of used components and because of the lack of physical continuity between musicians gestures and the resulting sound. A novel approach was implemented using a mixed-reality system in order to reveal the mechanisms of DMIs (see Figure 7). A description of this approach and of the first prototype will be submitted to the conference on New Interfaces for Musical Expression.

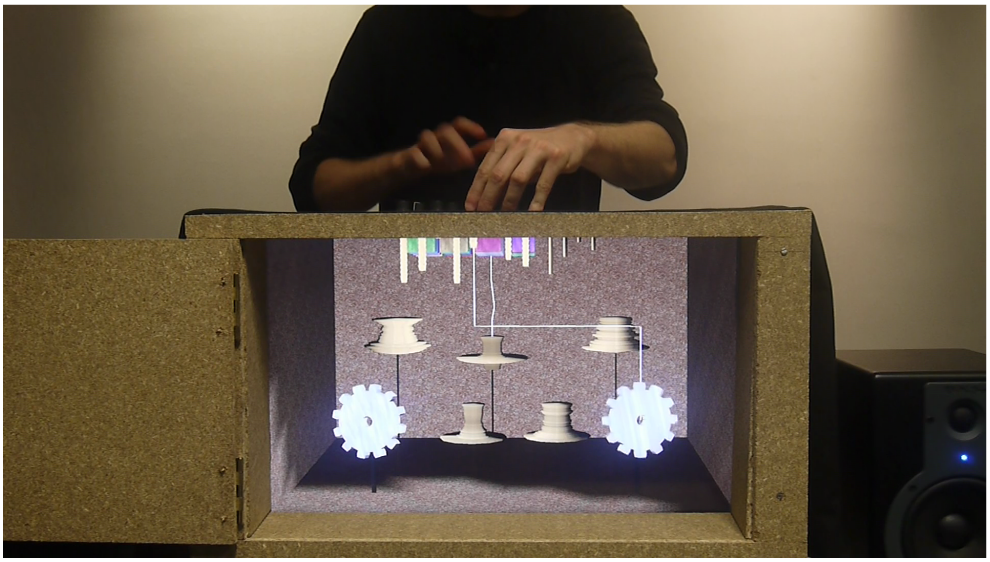


Figure 7. Rouages: a mixed-reality system that reveals the mechanisms of digital musical instruments to the audience.

6.5. Gateway driving simulator

Participants: Florian Larrue, Pauline Davignon, Pierre-Alexandre Favier, Martin Hachet.

As part of the SIMCA FUI project, the POTIOC team focuses on the design and evaluation of a gateway driving simulator, to teach drivers how to drive an airport gateway in virtual reality, i.e., in a safe and cost-effective environment. Gateways are the means to transfer passengers between the airport and the plane, for departures and arrivals. We have developped 3 simulators with different immersion levels (small, medium and immersive simulators, see, e.g., Figure 8). For each immersion level, we developped protocols in order to evaluate the impact of 3D technologies such as stereoscopy and head tracking on users' performances and preferences. Experimentations and evaluations are currently in progress.

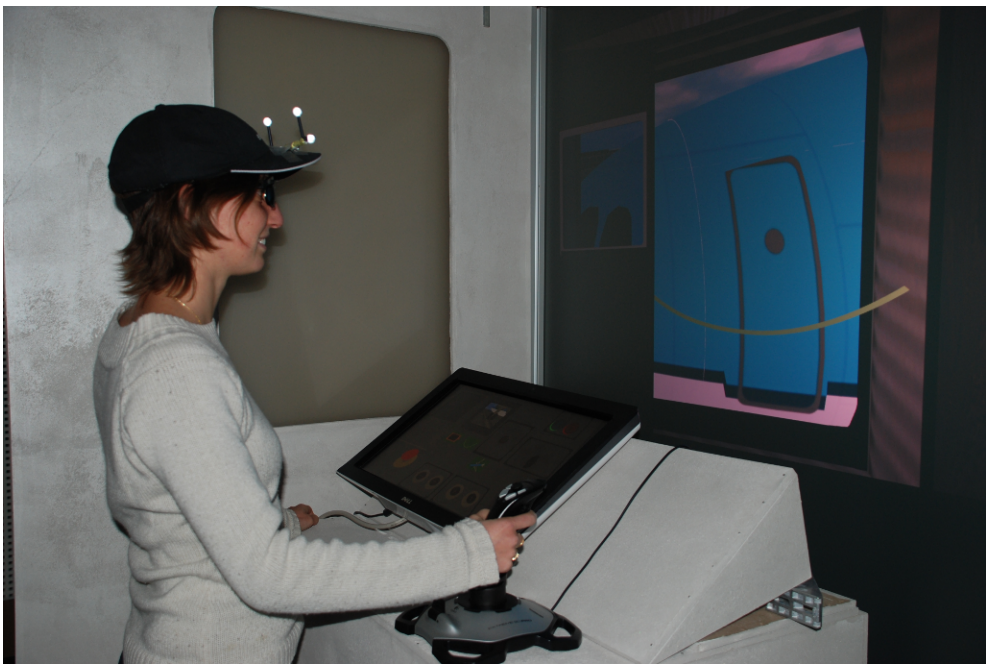


Figure 8. A user, equipped with head tracking and stereoscopic glasses, using the gateway driving simulator.

REVES Project-Team

5. New Results

5.1. Plausible Image Rendering

5.1.1. Rich Intrinsic Image Decomposition of Outdoor Scenes from Multiple Views

Participants: Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

Intrinsic image decomposition aims at separating photographs into independent reflectance and illumination layers. We show that this ill-posed problem can be solved by using multiple views of the scene from which we derive additional constraints on the decomposition.

Our first method uses pictures from multiple views at a *single time of day* to automatically reconstruct a 3D point cloud of an outdoor scene. Although this point cloud is sparse and incomplete, it is sufficient to compute plausible sky and indirect illumination at each oriented 3D point, given an environment map that represents incoming distant radiance. We introduce an optimization method to estimate sun visibility over the point cloud, which compensates for the lack of accurate geometry and allows the extraction of precise cast shadows. We finally use image-guided propagation algorithms to propagate the illumination computed over the sparse point cloud to every pixel, and to separate the illumination into distinct sun, sky, and indirect components. This *rich intrinsic image decomposition* enables advanced image manipulations, illustrated in Figure 3 .

This work has led to the RID software (Section 4.1) and to a technology transfer agreement with Autodesk (Section 6.1.1.1). A paper will be published in the IEEE Transactions on Visualization and Computer Graphics journal [18] (in press). It has also been presented at SIGGRAPH 2012 in the Poster and Talk sessions [22].

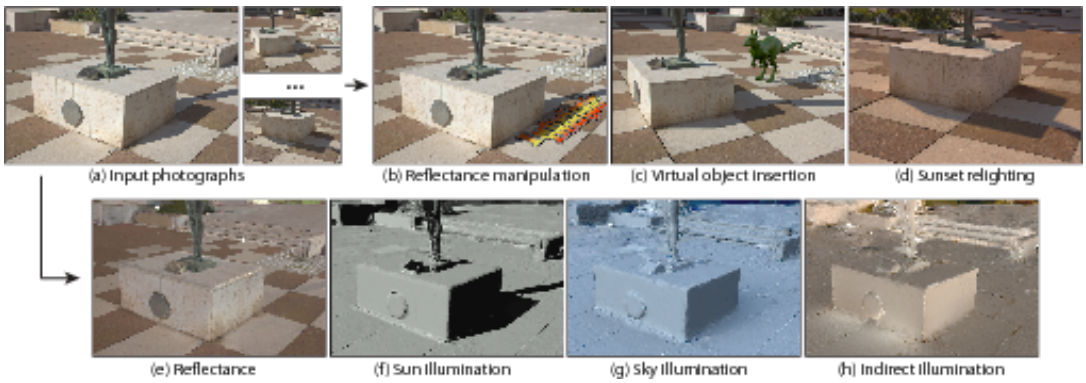


Figure 3. Starting from multiple views of the scene (a), our method decomposes photographs into four intrinsic layers — the reflectance (e), the illumination due to sun (f), the illumination due to sky (g) and the indirect illumination (h). Each layer can then be manipulated independently for advanced image editing applications (b-d).

5.1.2. Coherent Intrinsic Images from Photo Collections

Participants: Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

We propose a second method to compute intrinsic images in the presence of varying lighting conditions. Our method exploits the rich information provided by *multiple viewpoints and illuminations* in an image collection to process complex scenes without user assistance, nor precise and complete geometry. Such collections can be gathered from photo-sharing websites, or captured indoors with a light source which is moved around the scene.

We use multi-view stereo to automatically reconstruct 3D points and normals, from which we derive relationships between reflectance values at different locations, across multiple views, and consequently across different lighting conditions. In addition, we propose an optimization approach which enforces coherent reflectance in all views of a scene.

The resulting *coherent intrinsic images* enable image-based illumination transfer between photographs of the collection, as illustrated in Figure 4 .

This work is a collaboration with Frédo Durand (MIT) and Sylvain Paris (Adobe), and started with a visit of Pierre-Yves Laffont at MIT during Summer 2011. It has been published in the ACM Transactions on Graphics journal [19], and has been presented at SIGGRAPH Asia 2012.

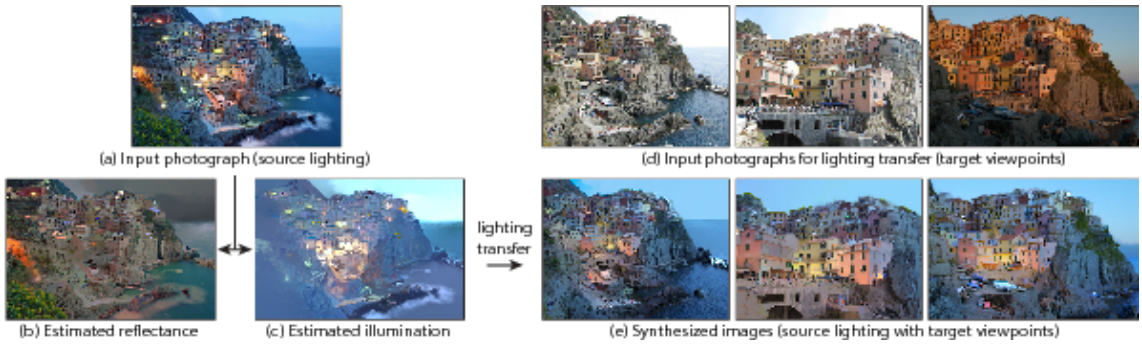


Figure 4. Our method automatically decomposes each image of a photo collection into reflectance and illumination (a-c). Transferring the illumination layer (c) to other viewpoints (d) yields synthetic images with novel viewpoint/lighting combinations (e).

5.1.3. Intrinsic Images by Clustering

Participant: Jorge Lopez Moreno.

Decomposing an input image into its intrinsic illumination and reflectance components is a long-standing ill-posed problem. We present a novel algorithm that requires no user strokes and works on a single image. Based on simple assumptions about its reflectance and luminance, we first find clusters of similar reflectance in the image, and build a linear system describing the connections and relations between them. Our assumptions are less restrictive than widely-adopted Retinex-based approaches, and can be further relaxed in conflicting situations. The resulting system is robust even in the presence of areas where our assumptions do not hold. We show a wide variety of results, including natural images, objects from the MIT dataset and texture images, along with several applications, proving the versatility of our method (see Figure 5).

This work is a collaboration with Elena Garces, Adolfo Munoz and Diego Gutierrez from University of Zaragoza (Spain). The work was published in an special issue of the journal Computer Graphics Forum and presented at the Eurographics Symposium on Rendering 2012 [16].

5.1.4. Relighting for Image Based Rendering

Participants: Sylvain Duchêne, Jorge Lopez Moreno, Stefan Popov, George Drettakis.

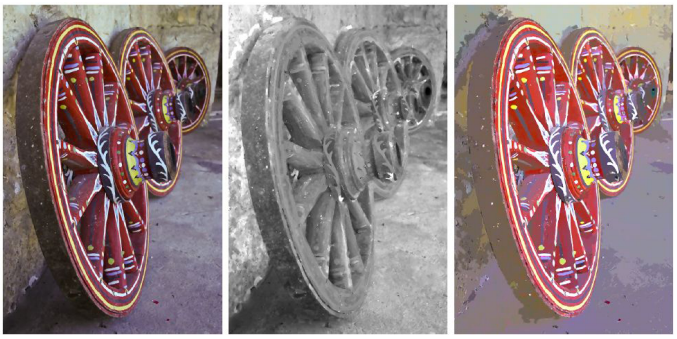


Figure 5. Decomposition by our method of the input image (left) into illumination (center) and reflectance (right) components.

Image-based rendering generates realistic virtual images from a small set of photographs. However, while current methods can simulate novel viewpoints from the input pictures, they cannot produce novel illumination conditions that differ from the lighting at the time of capture. The goal of this project is to provide such relighting capabilities. Our method first rely on multi-view stereo algorithms to estimate a coarse geometry of the scene. This geometry is often innacurate and incomplete. We complement it with image-based propagation algorithms that fill-in the missing data using the high-resolution input pictures. This combination of geometric and image-based cues allows us to generate plausible shadow motion and simulate novel sun directions.

5.1.5. Depth Synthesis and Local Warps for Plausible Image-based Navigation

Participants: Gaurav Chaurasia, Sylvain Duchêne, George Drettakis.

Modern multi-view stereo algorithms can estimate 3D geometry from a small set of unstrcutured photographs. However, the 3D reconstruction often fails on vegetation, vehicles and other complex geometry present in everyday urban scenes. We introduce a new Image-Based Rendering algorithm that is robust to unreliable geometry. Our algorithm segments the image into superpixels, *synthesizes* depth in superpixels with missing depth, warps them using a shape-preserving warp and blends them to create real-time plausible novel views for challenging target scenes, resulting in convincing immersive navigation experience.

This work is in collaboration with Dr. Olga Sorkine at ETH Zürich. and has been submitted to ACM Transactions on Graphics.

5.1.6. Perception of Slant for Image-Based Rendering

Participants: Christian Richardt, Peter Vangorp, George Drettakis.

Image-based rendering can create images with a high level of realism using simple geometry. However, as soon as the viewer moves away from the correct viewpoint, the image appears deformed. This work investigates the parameters which influence the perception of these image deformations. We propose a novel model of slant perception, which we validate using psychophysical experiments.

This work is a collaboration with Peter Vangorp at MPI Informatik, and Emily Cooper and Martin Banks from the University of California, Berkeley; in the context of the Associate Team CRISP (see also Section 7.3.1.1).

5.1.7. Lightfield Editing

Participant: Adrien Bousseau.

Lightfields capture multiple nearby views of a scene and are consolidating themselves as the successors of conventional photographs. As the field grows and evolves, the need for tools to process and manipulate lightfields arises. However, traditional image manipulation software such as Adobe Photoshop are designed to handle single views and their interfaces cannot cope with multiple views coherently. In this work we evaluate different user interface designs for lightfield editing. Our interfaces differ mainly in the way depth is presented to the user and build uppon different depth perception cues.

This work is a collaboration with Adrian Jarabo, Belen Masia and Diego Gutierrez from Universidad de Zaragoza and Fabio Pellacini from Sapienza Università di Roma.

5.1.8. Example-Based Fractured Appearance

Participants: Carles Bosch, George Drettakis.

A common weathering effect is the appearance of cracks due to material fractures. Previous exemplar-based aging and weathering methods have either reused images or sought to replicate observed patterns exactly. We propose an approach to exemplar-based modeling that creates weathered patterns by matching the statistics of fracture patterns in a photograph. We conducted a user study to determine which statistics are correlated to visual similarity and how they are perceived by the user. We describe a physically-based fracture model capable of producing similar crack patterns at interactive rates and an optimization method to determine its parameters based on key statistics of the exemplar. Our approach is able to produce a variety of fracture effects from simple crack photographs at interactive rates, as shown in Figure 6 .

Figure 6. Application of our example-based fracturing method on different scenes. Photographs of input fracture patterns are shown in the insets.

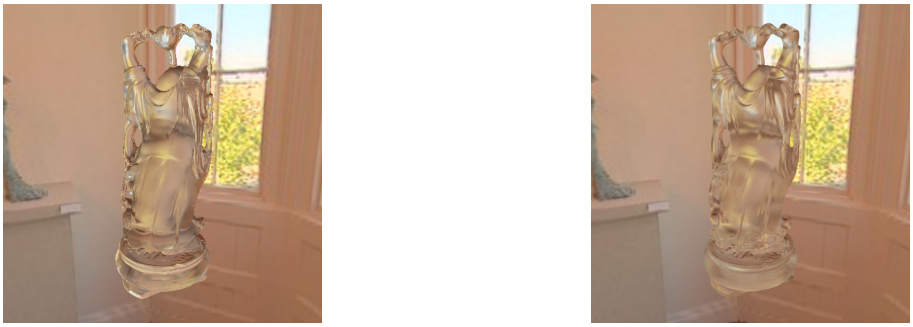
This work is a collaboration with Loeiz Glondu, Maud Marchal and George Dumont from IRISA-INSA/Inria Rennes - Bretagne Atlantique, Lien Muguercia from the University of Girona, and Holly Rushmeier from Yale University. The work was published in the Computer Graphics Forum journal and presented at the 23rd Eurographics Symposium on Rendering [17].

5.1.9. Real-Time Rendering of Rough Refraction

Participant: Adrien Bousseau.

We propose an algorithm to render objects made of transparent materials with rough surfaces in real-time, under all-frequency distant illumination. Rough surfaces cause wide scattering as light enters and exits objects, which significantly complicates the rendering of such materials. We present two contributions to approximate the successive scattering events at interfaces, due to rough refraction: First, an approximation of the Bidirectional Transmittance Distribution Function (BTDF), using spherical Gaussians, suitable for real-time estimation of environment lighting using pre-convolution; second, a combination of cone tracing and macro-geometry filtering to efficiently integrate the scattered rays at the exiting interface of the object. We demonstrate the quality of our approximation by comparison against stochastic ray-tracing (see Figure 7).

Furthermore we propose two extensions to our method for supporting spatially varying roughness on object surfaces and local lighting for thin objects.



(a) Ground truth (b) Our method

Figure 7. Compared to an expensive ray-traced reference (a), our method produces plausible results in real-time (b).

This work is a collaboration with Charles De Rousiers , Kartic Subr, Nicolas Holzschuch from Inria Grenoble, and Ravi Ramamoorthi from UC Berkeley in the context of the Associate Team CRISP (see also Section 7.3.1.1). A paper describing the method was published in the IEEE Transactions on Visualization and Computer Graphics journal [14].

5.1.10. Gabor Noise by Example

Participants: Ares Lagae, George Drettakis.

Procedural noise is a fundamental tool in Computer Graphics. However, designing noise patterns is hard. In this project, we propose *Gabor noise by example*, a method to estimate the parameters of bandwidth-quantized Gabor noise, a procedural noise function that can generate noise with an arbitrary power spectrum, from exemplar Gaussian textures, a class of textures that is completely characterized by their power spectrum (see Figure 8).

More specifically, we introduce (i) bandwidth-quantized Gabor noise, a generalization of Gabor noise to arbitrary power spectra that enables robust parameter estimation and efficient procedural evaluation; (ii) a robust parameter estimation technique for quantized-bandwidth Gabor noise, that automatically decomposes the noisy power spectrum estimate of an exemplar into a sparse sum of Gaussians using non-negative basis pursuit denoising; and (iii) an efficient procedural evaluation scheme for bandwidth-quantized Gabor noise, that uses multi-grid evaluation and importance sampling of the kernel parameters. Gabor noise by example preserves the traditional advantages of procedural noise, including a compact representation and a fast on-the-fly evaluation, and is mathematically well-founded.

This work is a collaboration with Bruno Galerne from MAP5, Université Paris Descartes and CNRS, Sorbonne Paris Cité; Ares Lagae from KU Leuven; and Sylvain Lefebvre from the ALICE project team, Inria Nancy - Grand Est. This work was presented at SIGGRAPH 2012 and published in ACM Transactions on Graphics [15].

5.1.11. Structured Gabor noise

Participants: Gaurav Chaurasia, Ares Lagae, George Drettakis.

Current procedural noise synthesis techniques [15] are limited to Gaussian random field textures. This project aims to generalize procedural noise to a broader class of structured textures.

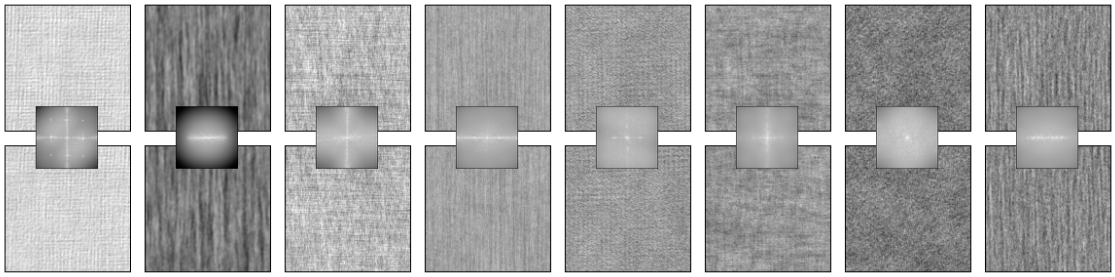


Figure 8. Gabor noise by example is a method to estimate the parameters of bandwidth-quantized Gabor noise, a procedural noise function that can generate noise with an arbitrary power spectrum, from exemplar Gaussian textures, a class of textures that is completely characterized by their power spectrum. (row 1) Gaussian texture. (row 2) Procedural noise. (insets) Estimated power spectrum.

This work is in collaboration with Dr. Ares Lagae (Katholieke Universiteit Leuven, Belgium), Dr. Bruno Galerne (Université Paris Descartes) and Prof. Ravi Ramamoorthi (UC Berkeley), in the contect of the Associate Team CRISP (Section 7.3.1.1).

5.1.12. Gloss Perception in Painterly and Cartoon Rendering

Participant: Adrien Bousseau.

Depictions with traditional media such as painting and drawing represent scene content in a stylized manner. It is unclear however how well stylized images depict scene properties like shape, material and lighting. In this project, we use non photorealistic rendering algorithms to evaluate how stylization alters the perception of gloss (see Figure 9). Our study reveals a compression of the range of representable gloss in stylized images so that shiny materials appear more diffuse in painterly rendering, while diffuse materials appear shinier in cartoon images.

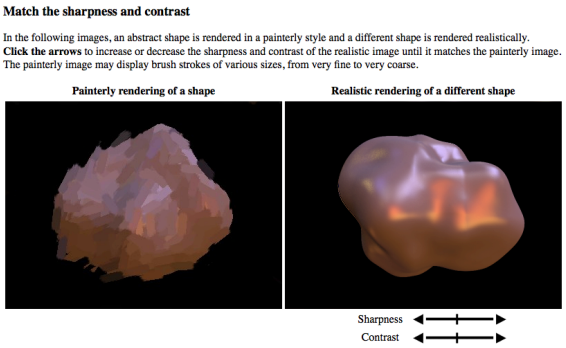


Figure 9. The experimental task used for studying gloss perception in stylized images.

From our measurements we estimate the function that maps realistic gloss parameters to their perception in a stylized rendering. This mapping allows users of NPR algorithms to predict the perception of gloss in their images. The inverse of this function exaggerates gloss properties to make the contrast between materials in

a stylized image more faithful. We have conducted our experiment both in a lab and on a crowdsourcing website. While crowdsourcing allows us to quickly design our pilot study, a lab experiment provides more control on how subjects perform the task. We provide a detailed comparison of the results obtained with the two approaches and discuss their advantages and drawbacks for studies like ours.

This work is a collaboration with James O'Shea, Ravi Ramamoorthi and Maneesh Agrawala from UC Berkeley in the context of the Associate Team CRISP (see also Section 7.3.1.1) and Frédo Durand from MIT. It will be published in ACM Transactions on Graphics 2013 [12] (in press).

5.2. Interaction and Design for Audiovisual Virtual Environments

5.2.1. Auditory-visual integration of emotional signals in a virtual environment for cynophobia

Participants: Emmanuelle Chapoulie, Adrien David, Rachid Guerchouche, George Drettakis.

Cynophobia (dog phobia) has both visual and auditory relevant components. In order to investigate the efficacy of virtual reality exposure-based treatment for cynophobia, we studied the efficiency of auditory-visual environments in generating presence and emotion. We conducted an evaluation test with healthy participants sensitive to cynophobia in order to assess the capacity of auditory-visual virtual environments to generate fear reactions. Our application involves both high fidelity visual stimulation displayed in an immersive space and 3D sound. This specificity enables us to present and spatially manipulate fearful stimuli in the auditory modality, the visual modality and both.

We conducted a study where participants were presented with virtual dogs in realistic environments. Dogs were presented in a progressive manner, from unimodal and static to audiovisual and dynamic. Participants were also submitted a Behavioral Assessment Test at the beginning and end of the experiment where they were presented a virtual dog walking towards them step by step until it was extremely close. Finally, they completed several questionnaires and were asked to comment on their experience. The participants reported higher anxiety levels in response to auditory-visual stimuli compared to unimodal stimuli. Our results strongly suggest that manipulating auditory-visual integration might be a good way to modulate affective reactions and that auditory-visual VR are a promising tool for the treatment of cynophobia.

This work is a collaboration with Marine TAFFOU and Isabelle VIAUD-DELMON from IRCAM, in the context of ARC NIEVE (see also Section 7.1.4). The work was published in the Annual Review of Cybertherapy and Telemedicine in 2012.

5.2.2. Procedural audio modeling for particle-based environmental effects

Participants: Charles Verron, George Drettakis.

In this project we proposed a sound synthesizer dedicated to particle-based environmental effects, for use in interactive virtual environments. The synthesis engine is based on five physically-inspired basic elements which we call sound atoms, that can be parameterized and stochastically distributed in time and space. Based on this set of atomic elements, models are presented for reproducing several environmental sound sources. Compared to pre-recorded sound samples, procedural synthesis provides extra flexibility to manipulate and control the sound source properties with physically-inspired parameters. The controls are used to simultaneously modify particle-based graphical models, resulting in synchronous audio/graphics environmental effects. The approach is illustrated with three models, that are commonly used in video games: fire, wind, and rain. The physically-inspired controls simultaneously drive graphical parameters (e.g., distribution of particles, average particles velocity) and sound parameters (e.g., distribution of sound atoms, spectral modifications) as illustrated on Figure 10 for fire. The joint audio/graphics control results in a tightly-coupled interaction between the two modalities that enhances the naturalness of the scene.

The work was presented at the 133rd AES convention in October 2012 [23].

5.2.3. Perception of crowd sounds

Participants: Charles Verron, George Drettakis.

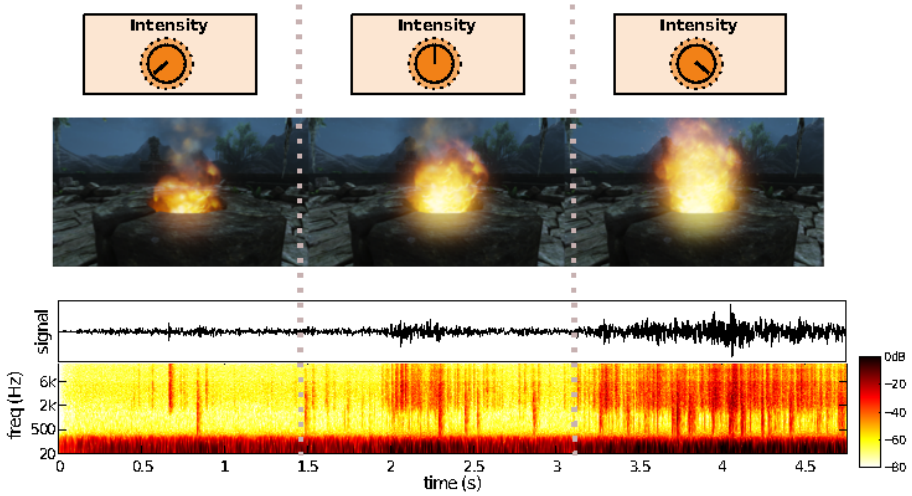


Figure 10. Audio/graphics high-level control of a fire. The control Intensity changes the rate and gain of noisy impacts, and the combustion noise of the fire sound model. Simultaneously, it controls the flame/smoke particle spawn rate for the graphics simulation.

Simulating realistic crowd scenes is an important challenge for virtual reality and games. Motion capture techniques allow to reproduce efficiently characters that look, move and sound realistic in virtual environments. However a huge amount of data is required to ensure that all agents behave differently in a big crowd. A common approach to solve this issue is to “clone" the same appearance, motion or sound several times, which can lead to perceived repetitions and break the realism of the scene. In this study we further investigate our perception of crowd scenes. Using a database of motions and sounds captured for 40 actors, along with a database of 40 different appearance templates, we propose an experimental framework to evaluate the perceptual degradations caused by clones. A particular attention is given to evaluate the influence of appearance, motion and sound, either separately or in multimodal conditions. This study aims at providing useful insights on our perception of crowd scenes, and guidelines to designers in order to reduce the amount of resources to produce convincing crowd scenes.

This ongoing project is a collaboration between Inria, CNRS-LMA (Marseille, France) and Trinity College (Dublin, Ireland).

5.2.4. Walking in a Cube: Novel Metaphors for Safely Navigating Large Virtual Environments in Restricted Real Workspaces

Participants: Peter Vangorp, Emmanuelle Chapoulie, George Drettakis.

Immersive spaces such as 4-sided displays with stereo viewing and high-quality tracking provide a very engaging and realistic virtual experience. However, walking is inherently limited by the restricted physical space, both due to the screens (limited translation) and the missing back screen (limited rotation). Locomotion techniques for such restricted workspaces should satisfy three concurrent goals: keep the user safe from reaching the translational and rotational boundaries; increase the amount of real walking; and finally, provide a more enjoyable and ecological interaction paradigm compared to traditional controller-based approaches.

We have proposed three novel locomotion techniques that attempt to satisfy these goals in innovative ways. We constrain traditional Wand locomotion by turning off the Wand controls for directions that can be reached by real walking instead, and we display warning signs when the user approaches the limits of the real workspace (Figure 11 (a)). We also extend the Magic Barrier Tape paradigm with “blinders” to avoid rotation towards the

missing back screen (Figure 11 (b)). Finally, we introduce the "Virtual Companion", which uses a small bird to guide the user through virtual environments larger than the physical space (Figure 11 (c,d)).

We evaluate the three new techniques through a user study with travel-to-target and path following tasks. The study provides insight into the relative strengths of each new technique for the three aforementioned goals. Specifically, if speed and accuracy are paramount, traditional controller interfaces augmented with our novel warning techniques may be more appropriate; if physical walking is more important, two of our paradigms, the extended Magic Barrier Tape and the Constrained Wand, should be preferred; and finally, fun and ecological criteria would favor the Virtual Companion.

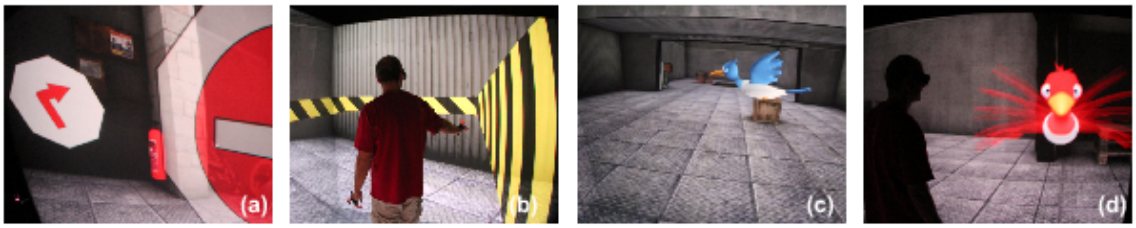


Figure 11. Screenshots illustrating the three novel locomotion techniques. From left to right: (a) Constrained Wand and signs: the "no-way" and "turn right" signs. (b) Extended Magic Barrier Tape: the tape and blinders. (c,d) Virtual Companion: the bird in "rest mode" (c) and "protection mode" (d).

This work is a collaboration with Gabriel Cirio, Maud Marchal and Anatole Lécuyer (VR4I project team, IRISA-INSA/Inria Rennes - Bretagne Atlantique) in the context of ARC NIEVE (see Section 7.1.4). The work was published in the special issue of the journal IEEE Transactions on Visualization and Computer Graphics (TVCG) [13], and presented at the IEEE Virtual Reality conference 2012.

5.2.5. *Natural Gesture-based Interaction for Complex Tasks in an Immersive Cube*

Participants: Emmanuelle Chapoulie, Jean-Christophe Lombardo, George Drettakis.

We present a solution for natural gesture interaction in an immersive cube in which users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks. To do this, we develop a solution using finger tracking coupled with a real-time physics engine, combined with a comprehensive approach for hand gestures, which is robust to tracker noise and simulation instabilities. To determine if our natural gestures are a feasible interface in an immersive cube, we perform an exploratory study for tasks involving the user walking in the cube while performing complex manipulations such as balancing objects. We compare gestures to a traditional 6-DOF Wand, and we also compare both gestures and Wand with the same task, faithfully reproduced in the real world. Users are also asked to perform a free task, allowing us to observe their perceived level of presence in the scene. Our results show that our robust approach provides a feasible natural gesture interface for immersive cube-like environments and is perceived by users as being closer to the real experience compared to the Wand.

This work is a collaboration with Evanthia Dimara and Maria Roussou from the University of Athens and with Maud Marchal from IRISA-INSA/Inria Rennes - Bretagne Atlantique. The work has been submitted to 3DUI 2013.

5.2.6. *CrossShade: Shading Concept Sketches Using Cross-Section Curves*

Participant: Adrien Bousseau.

We facilitate the creation of 3D-looking shaded production drawings from concept sketches. The key to our approach is a class of commonly used construction curves known as cross-sections, that function as an aid to both sketch creation and viewer understanding of the depicted 3D shape. In particular, intersections of these curves, or cross-hairs, convey valuable 3D information, that viewers compose into a mental model of the overall sketch. We use the artist-drawn cross-sections to automatically infer the 3D normals across the sketch, enabling 3D-like rendering (see Figure 12).

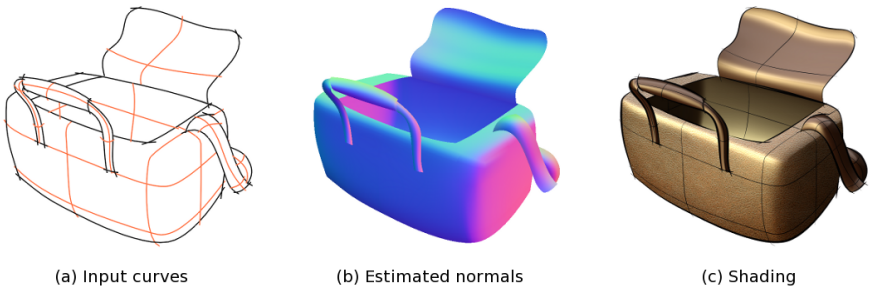


Figure 12. Concept sketches (a) frequently use cross-sections (drawn in orange) to convey 3D shape with just a handful of lines. We derive the mathematical properties of cross-section curves and leverage them to automatically estimate surface normals across the drawn objects (b). The resulting normal field allow users to shade the objects using a variety of shading styles and setups (c).

The technical contribution of our work is twofold. First, we distill artistic guidelines for drawing cross-sections and insights from perception literature to introduce an explicit mathematical formulation of the relationships between cross-section curves and the ge- ometry they aim to convey. We then use these relationships to develop an algorithm for estimating a normal field from cross-section curve networks and other curves present in concept sketches. We validate our formulation and algorithm through a user study and a ground truth normal comparison. These contributions enable us to shade a wide range of concept sketches with a variety of rendering styles.

This work is a collaboration with Cloud Shao and Karan Singh from the University of Toronto and Alla Sheffer from the University of British Columbia. It has been published at ACM Transactions on Graphics, proceedings of the SIGGRAPH 2012 conference.

5.2.7. CrossShape

Participant: Adrien Bousseau.

We facilitate the automatic creation of surfaced 3D models from design sketches that employ a commonly drawn network of cross-section curves. Our previous method generates 3D renderings of input sketches by creating a 3D surface normal field that interpolates the sketched cross-sections. This normal field however, incorporates the inevitable inaccuracy of sketched curves, making it inappropriate for 3D surface construction.

Successful construction of the 3D surface perceived from sketches requires cross-section properties and other perceived curve relationships such as symmetry and parallelism, to be met precisely. We present a novel formulation where these geometric constraints are satisfied while minimizing the difference between the sketch and the 3D cross-sections projected on it. We validate our approach by producing accurate surface reconstructions of existing 3D models represented using a network of cross-sections as well on a variety of sketch input. Finally we illustrate our surfacing solution within an interactive sketch based modeling framework.

This work is a collaboration with James McCrae and Karan Singh from the University of Toronto and Xu Baoxuan and Alla Sheffer from the University of British Columbia.

5.2.8. Computer-assisted drawing

Participants: Emmanuel Iarussi, Adrien Bousseau.

A major challenge in drawing from observation is to trust what we *see* rather than what we *know*. Drawing books and tutorials provide simple techniques to gain consciousness of the shapes that we observe and their relationships. Common techniques include drawing simple geometrical shapes first – also known as *blocking in* – and checking for alignments and equal proportions. While very effective, these techniques are usually illustrated on few examples and it takes significant effort to generalize them to an arbitrary model. In addition, books and tutorials only contain static instructions and cannot provide feedback to people willing to practice drawing.

In this project, we develop an interactive drawing tool that assists users in their practice of common drawing techniques. Our *drawing assistant* helps users to draw from any model photograph and provides corrective feedback interactively.

This work is a collaboration with Theophanis Tsandilas from the InSitu project team, Inria Saclay - Ile de France, in the context of the ANR DRAO project (see Section 7.1.2).

5.2.9. Depicting materials in vector graphics

Participants: Jorge Lopez-Moreno, Adrien Bousseau, Stefan Popov, George Drettakis.

Vector drawing tools like Illustrator and InkScape enjoy great popularity in illustration and design because of their flexibility, directness and distinctive look. Within such tools, skillful artists depict convincing material and lighting effects using 2D vector primitives like gradients and paths. However, it takes significant expertise to convey plausible material appearance in vector drawings. Instead, novice users often fill-in regions with a constant color, sacrifying plausibility for simplicity. In this project we present the first vector drawing tool that automates the depiction of material appearance. Users can use our tool to either fill-in regions automatically, or to generate an initial set of vector primitives that they can refine at will.

This work is a collaboration with Maneesh Agrawala from the University of Berkeley in the context of the Associate Team CRISP (see Section 7.3.1.1).

5.2.10. Gradient Art: Creation and Vectorization (survey)

Participant: Adrien Bousseau.

We survey the main two categories of methods for producing vector gradients. One is mainly interested in converting existing photographs into dense vector representations. By vector it is meant that one can zoom infinitely inside images, and that control values do not have to lie onto a grid but must represent subtle color gradients found in input images. The other category is tailored to the creation of images from scratch, using a sparse set of vector primitives. In this case, we still have the infinite zoom property, but also an advanced model of how space should be filled in-between primitives, since there is no input photograph to rely on. These two categories are actually extreme cases, and seem to exclude each other: a dense representation is difficult to manipulate, especially when one wants to modify topology; a sparse representation is hardly adapted to photo vectorization, especially in the presence of texture. Very few methods lie in the middle, and the ones that do require user assistance.

We published our survey in the book *Image and Video based Artistic Stylization* [25] editied by Springer. The survey was written in Collaboration with Pascal Barla from the MANAO project team, Inria Bordeaux - Sud Ouest, in the context of the ANR DRAO project (see Section 7.1.2).

VR4I Team

6. New Results

6.1. Physical modelling and simulation

6.1.1. Real-time mechanical simulation of brittle fracture

Participants: Loeïz Glondu, Georges Dumont [contact], Maud Marchal [contact].

Simulating brittle fracture of stiff bodies is now commonplace in computer graphics. However, simulating the deformations undergone by the bodies in a realistic way remains computationaly expensive. Thus, physically-based simulation of brittle fracture in real-time is still challenging for interactive applications. We have worked on a physically-based approach for simulating realistic brittle fracture in real-time.Our method is mainly composed of two parts: (1) a fracture initiation method based on modal analysis, (2) a fast energy-based fracture propagation algorithm. Results that emphasize the "real-time" part of this method have been published in [9]. Collision detection plays a key role in simulation performance. This is particularly true for fracture simulation, where multiple new objects are dynamically created. We proposed algorithms and data structures for collision detection in real-time brittle fracture simulations. We build on a combination of well-known efficient data structures, namely distance fields and sphere trees, making our algorithm easy to integrate on existing simulation engines. We proposed novel methods to construct these data structures, such that they can be efficiently updated upon fracture events and integrated in a simple yet effective self-adapting contact selection algorithm. Altogether, we drastically reduce the cost of both collision detection and collision response. We have evaluated our global solution for collision detection on challenging scenarios, achieving high frame rates suited for hard real-time applications such as video games or haptics [23]. Moreover, a common weathering effect is the appearance of cracks due to material fractures. We introduced a method to exemplar-based modeling that creates weathered patterns on synthetic objects by matching the statistics of fracture patterns in a photograph. A user study was proposed to determine which statistics are correlated to visual similarity and how they are perceived by the user. A revised physically-based fracture model capable of producing a wide range of crack patterns at interactive rates has been proposed whose parameter can be determined by a Bayesian optimization to produce a pattern with the same key statistics as an exemplar [10]. This work was the subject of the PhD thesis of Loeïz Glondu that has been successfuly defensed [3].

6.1.2. Collision detection in large scale environments with High Performance Computing

Participants: Bruno Arnaldi, Quentin Avril, Valérie Gouranton [contact].

We propose [14] a novel and efficient GPU-based parallel algorithm to cull non-colliding objects pairs in very large scale dynamic simulations. It allows to cull objects in less than 25ms with more than 100K objects. It is designed for many-core GPU and fully exploits multi-threaded capabilities and data-parallelism. In order to take advantage of the high number of cores, a new mapping function is defined that enables GPU threads to determine the objects pair to compute without any global memory access. These new optimized GPU kernel functions use the thread indexes and turn them into a unique pair of objects to test. A square root approximation technique is used based on Newton's estimation, enabling the threads to only perform a few atomic operations to cull non-colliding objects. We present a first characterization of the approximation errors that enables the fixing of incorrect computations. Input and output GPU streams are optimized using binary masks. The implementation and evaluation is made on large-scale dynamic rigid body simulations. The increase in speed is highlighted over other recently proposed CPU and GPU-based techniques. The comparison shows that our system is, in most cases, faster than previous approaches.

6.1.3. Simulation evaluations for ergonomics in VR

Participants: Georges Dumont [contact], Charles Pontonnier.

The use of virtual reality tools for ergonomics applications is a very important challenge.

In order to improve the design of workstations, an estimation of the muscle forces involved in the work tasks has to be done.

For example, one of our study assessed the level of confidence for results obtained with an inverse dynamics method from real captured work tasks. The chosen tasks are meat cutting tasks, well known to be highly correlated to musculoskeletal troubles appearance in the slaughter industry.

The experimental protocol consists in recording three main data during meat cutting tasks, and analyze their variation when some of the workstation design parameters are changing.

1. External (cutting)force data : for this purpose, a 3D instrumented knife has been designed in order to record the force applied by the subject during the task;
2. Motion Capture data : for this purpose, we use a motion capture system with active markers (Visualeyez II, Phoenix Technologies, Canada);
3. EMG data : several muscle activities are recorded using electromyographic electrodes, in order to compare these activities to the ones obtained from the inverse dynamics method.

With regard to the design parameters, that are the table height and the cutting direction, trends of recorded muscles activations were defined in order to be compared to computed ones issued from a musculoskeletal simulation performed with the AnyBody modeling system (AnyBody, Aalborg, Denmark). Results showed that an optimal set of design parameters can be obtained [27], whereas motor control strategies are highly dependent to the subject's experience and morphology.

This work has been done in collaboration with the Center for Sensory-motor Interaction (SMI, Aalborg University, Aalborg, Denmark), particularly Mark de Zee (Associate Professor) and Pascal Madeleine (Professor).

Furthermore, the fidelity of the VR simulator has to be evaluated (see Figure 2). For example, a simulator for assembly task has been evaluated in comparing different types of interaction : real, virtual and virtual + force feedback [28]. Objective and subjective metrics of discomfort led to highlight the influence of the environment on motor control and sensory feedback, changing more or less deeply the way the task is performed. Those change have to be taken into account to enable the use of such simulators for ergonomics purposes.

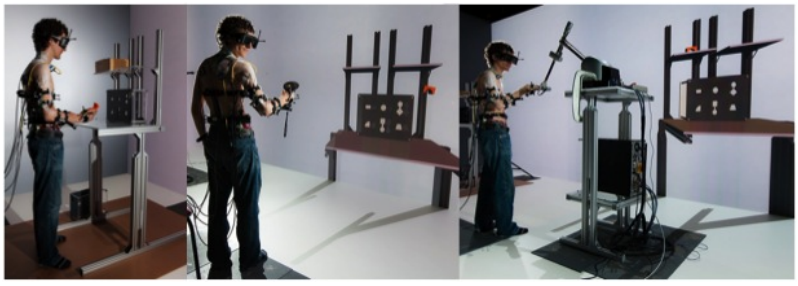


Figure 2. Simulation of an assembly task (Left in real, center in virtual, right in virtual with force-feedback)

6.2. Multimodal immersive interaction

6.2.1. Immersive Archaeology

Participants: Bruno Arnaldi, Georges Dumont, Ronan Gaugne [contact], Valérie Gouranton [contact].

We propose a workflow of tools and procedures to reconstruct an existing archaeological site as a virtual 3D reconstitution in a large scale immersive system [35] . This interdisciplinary endeavor, gathering archaeologists and virtual reality computer scientists, is the first step of a joint research project with three objectives: (i) propose a common workflow to reconstruct archaeological sites as 3D models in fully immersive systems, (ii) provide archaeologists with tools and interaction metaphors to exploit immersive reconstitutions, and (iii) develop the use and access of immersive systems to archaeologists. In this context, we present [21] results from the immersive reconstitution of Carn's monument central chamber, in Finistere, France, a site currently studied by the Creaah archaeology laboratory. The results rely on a detailed workflow we propose, which uses efficient solutions to enable archaeologists to work with immersive systems. In particular, we proposed a procedure to model the central chamber of the Carn monument, and compare several softwares to deploy it in an immersive structure. We then proposed two immersive implementations of the central chamber, with simple interaction tools.

6.2.2. Novel 3D displays and user interfaces

Participants: Anatole Lécuyer [contact], David Gomez, Fernando Argelaguet, Maud Marchal, Jerome Ardouin.

We describe hereafter our recent results in the field of novel 3D User Interfaces and, more specifically, novel displays and interactive techniques to better perceive and interact in 3D. This encloses: (1) Novel interactive techniques for interaction with 3D web content, and (2) A novel display for augmented 3D vision.

6.2.2.1. Novel interactive techniques for 3D web content

The selection and manipulation of 3D content in desktop virtual environments is commonly achieved with 2D mouse cursor-based interaction. However, by interacting with image-based techniques we introduce a conflict between the 2D space in which the 2D cursor lays and the 3D content. For example, the 2D mouse cursor does not provide any information about the depth of the selected objects. In this situation, the user has to rely on the depth cues provided by the virtual environment, such as perspective deformation, shading and shadows.

In [24], we have explored new metaphors to improve the depth perception when interacting with 3D content. Our approach focus on the usage of 3D cursors controlled with 2D input devices (the Hand Avatar and the Torch) and a pseudo-motion parallax effect. The additional depth cues provided by the visual feedback of the 3D cursors and the motion parallax are expected to increase the users' depth perception of the environment.

The evaluation of proposed techniques showed that users depth perception was significantly increased. Users were able to better judge the depth ordering of virtual environment. Although 3D cursors showed a decrease of selection performance, it is compensated by the increased depth perception.

6.2.2.2. FLyVIZ : A novel display for providing humans with panoramic vision

Have you ever dreamed of having eyes in the back of your head? In [12], we have presented a novel display device called FlyVIZ which enables humans to experience a real-time 360-degree vision of their surroundings for the first time.

To do so, we combined a panoramic image acquisition system (positioned on top of the user's head) with a Head-Mounted Display (HMD). The omnidirectional images are transformed to fit the characteristics of HMD screens. As a result, the user can see his/her surroundings, in real-time, with 360 degree images mapped into the HMD field of view.

We foresee potential applications in different fields where augmented human capacity (an extended field-of-view) could benefit, such as surveillance, security, or entertainment. FlyVIZ could also be used in novel perception and neuroscience studies.

6.2.3. Brain-Computer Interfaces

Participants: Anatole Lécuyer [contact], Laurent George, Laurent Bonnet, Jozef Legeny.

Brain-computer interfaces (BCI) are communication systems that enable to send commands to a computer using only the brain activity. Cerebral activity is generally sensed with electroencephalography (or EEG). We describe hereafter our recent results in the field of brain-computer interfaces and virtual environments: (1) Novel signal processing techniques for EEG-based Brain-Computer Interfaces, and (2) Design and study of Brain-Computer Interaction with real and virtual environments.

6.2.3.1. Novel signal processing techniques for EEG-based Brain-Computer Interfaces

A first part of the BCI research conducted in the team is dedicated to EEG signal processing and classification techniques applied to cerebral EEG data.

To properly and efficiently decode brain signals into computer commands the application of efficient machine-learning techniques is required.

In [5] we could introduce two new features for the design of electroencephalography (EEG) based Brain-Computer Interfaces (BCI): one feature based on multifractal cumulants, and one feature based on the predictive complexity of the EEG time series. The multifractal cumulants feature measures the signal regularity, while the predictive complexity measures the difficulty to predict the future of the signal based on its past, hence a degree of how complex it is. We have conducted an evaluation of the performance of these two novel features on EEG data corresponding to motor-imagery. We also compared them to the gold standard features used in the BCI field, namely the Band-Power features. We evaluated these three kinds of features and their combinations on EEG signals from 13 subjects. Results obtained show that our novel features can lead to BCI designs with improved classification performance, notably when using and combining the three kinds of feature (band-power, multifractal cumulants, predictive complexity) together.
Evolutionary algorithms have also been increasingly applied in different steps of BCI implementations. In [29], we could then introduce the use of the covariance matrix adaptation evolution strategy (CMA-ES) for BCI systems based on motor imagery. The optimization algorithm was used to evolve linear classifiers able to outperform other traditional classifiers. We could also analyze the role of modeling variables interactions for additional insight in the understanding of the BCI paradigms.

6.2.3.2. Brain-Computer Interaction with real and virtual environments

A second part of our BCI research is dedicated to the improvement of BCI-based interaction with real and virtual environments. We have first initiated research on **Combining Haptic and Brain-Computer Interfaces**.

In [22], we have introduced the combined use of Brain-Computer Interfaces (BCI) and Haptic interfaces. We proposed to adapt haptic guides based on the mental activity measured by a BCI system. This novel approach has been illustrated within a proof-of-concept system: haptic guides were toggled during a path-following task thanks to a mental workload index provided by a BCI. The aim of this system was to provide haptic assistance only when the user's brain activity reflects a high mental workload.

A user study conducted with 8 participants showed that our proof-of-concept is operational and exploitable. Results showed that activation of haptic guides occurs in the most difficult part of the path-following task. Moreover it allowed to increase task performance by activating assistance only 59 percents of the time. Taken together, these results suggest that BCI could be used to determine when the user needs assistance during haptic interaction and to enable haptic guides accordingly.

This work paves the way to novel passive BCI applications such as medical training simulators based on passive BCI and smart guides. It has received the Best Paper Award of Eurohaptics 2012 conference, and was nominated for the BCI Award 2012.

6.2.4. Natural Interactive Walking in Virtual Environments

Participants: Anatole Lécuyer [contact], Maud Marchal [contact], Gabriel Cirio, Tony Regia Corte, Sébastien Hillaire, Léo Terziman.

Figure 3. Proof-of-concept system combining Haptic and a Brain-Computer Interface (haptic guides are toggled based on a mental workload index computed by the BCI)

We describe hereafter our recent results obtained in the field of "augmented" or "natural interactive" walking in virtual environments. Our first objective is to better understand the properties of human perception and human locomotion when walking in virtual worlds. Then, we intend to design advanced interactive techniques and interaction metaphors to enhance, in a general manner, the navigation possibilities in VR systems. Last, our intention is to improve the multisensory rendering of human locomotion and human walk in virtual environments, making full use of both haptic and visual feedback.

6.2.4.1. Perception of ground affordances in virtual environments

We have evaluated the perception of **ground affordances in virtual environments** (VE).

In [11], we considered the affordances for standing on a virtual slanted surface. Participants were asked to judge whether a virtual slanted surface supported upright stance. The objective was to evaluate if this perception was possible in virtual reality (VR) and comparable to previous works conducted in real environments. We found that the perception of affordances for standing on a slanted surface in virtual reality is possible and comparable (with an underestimation) to previous studies conducted in real environments. We also found that participants were able to extract and to use virtual information about friction in order to judge whether a slanted surface supported an upright stance. Finally, results revealed that the person's position on the slanted surface is involved in the perception of affordances for standing on virtual grounds. Taken together, our results show quantitatively that the perception of affordances can be effective in virtual environments, and influenced by both environmental and person properties. Such a perceptual evaluation of affordances in VR could guide VE designers to improve their designs and to better understand the effect of these designs on VE users.

6.2.4.2. Novel metaphors for navigating virtual environments

Immersive spaces such as 4-sided displays with stereo viewing and high-quality tracking provide a very engaging and realistic virtual experience. However, walking is inherently limited by the restricted physical space, both due to the screens (limited translation) and the missing back screen (limited rotation).

In [7], we proposed three novel locomotion techniques that have three concurrent goals: keep the user safe from reaching the translational and rotational boundaries; increase the amount of real walking and finally, provide a more enjoyable and ecological interaction paradigm compared to traditional controller-based approaches.

We notably introduced the "Virtual Companion", which uses a small bird to guide the user through VEs larger than the physical space. We evaluated the three new techniques through a user study with travel-to-target and path following tasks. The study provided insight into the relative strengths of each new technique for the three aforementioned goals. Specifically, if speed and accuracy are paramount, traditional controller interfaces augmented with our novel warning techniques may be more appropriate; if physical walking is more important, two of our paradigms (extended Magic Barrier Tape and Constrained Wand) should be preferred; last, fun and ecological criteria would favor the Virtual Companion.

6.2.4.3. Novel sensory feedback for improving sensation of walking in VR: the King-Kong Effects

Third, we have designed novel sensory feedbacks named "King-Kong Effects" to enhance the sensation of walking in virtual environments [33].

King Kong Effects are inspired by special effects in movies in which the incoming of a gigantic creature is suggested by adding visual vibrations/pulses to the camera at each of its steps (Figure 4).

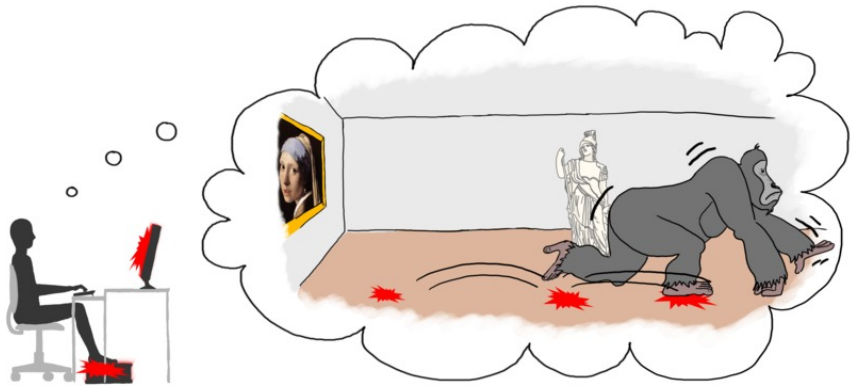


Figure 4. Concept of the King Kong Effects: Visual and Tactile vibrations inspired by special effects in movies enhance the sensation of walking in VE. Visual and Tactile feedbacks are generated at each step made in the VE.

We thus proposed to add artificial visual or tactile vibrations (King-Kong Effects or KKE) at each footstep detected (or simulated) during the virtual walk of the user. The user can be seated, and our system proposes to use vibrotactile tiles located under his/her feet for tactile rendering, in addition to the visual display. We have designed different kinds of KKE based on vertical or lateral oscillations, physical or metaphorical patterns, and one or two peaks for heal-toe contacts simulation.

We have conducted different experiments to evaluate the preferences of users navigating with or without the various KKE. Taken together, our results identify the best choices in term of sensation of walking for future uses of visual and tactile KKE, and they suggest a preference for multisensory combinations. Our King-Kong effects could be used in a variety of VR applications targeting the immersion of a user walking in a 3D virtual scene.

6.2.5. Haptic Interaction

Participants: Fernando Argelaguet, Fabien Danieau, Anatole Lécuyer [contact], Maud Marchal, Anthony Talvas.

6.2.5.1. Pseudo-Haptic Feedback

Pseudo-haptic feedback is a technique meant to simulate haptic sensations in virtual environments using visual feedback and properties of human visuo-haptic perception. Pseudo-haptic feedback uses vision to distort haptic perception and verges on haptic illusions. Pseudo-haptic feedback has been used to simulate various haptic properties such as the stiffness of a virtual spring, the texture of an image, or the mass of a virtual object.

In [13], we focused on the improvement of pseudo-haptic textures. Pseudo-haptic textures allow to optically-induce relief in tex- tures without a haptic device by adjusting the speed of the mouse pointer according to the depth information encoded in the texture. In this work, we have presented a novel approach for using curvature information instead of relying on depth information. The curvature of the texture is encoded in a normal map which allows the computation of the curvature and local changes of orientation, according to the mouse position and direction.

A user evaluation was conducted to compare the optically-induced haptic feedback of the curvature-based approach versus the original depth-based approach based on depth maps. Results showed that users, in addition to being able to efficiently recognize simulated bumps and holes with the curvature-based approach, were also able to discriminate shapes with lower frequency and amplitude.

6.2.5.2. Bi-Manual Haptic Feedback

In the field of haptics and virtual reality, two-handed interaction with virtual environments (VEs) is a domain that is slowly emerging while bearing very promising applications.

In [32] we could present a set of novel interactive techniques adapted to two-handed manipulation of objects with dual 3DoF single- point haptic devices (see Figure 5). We first proposed the double bubble for bimanual haptic exploration of virtual environments through hybrid position/rate controls, and a bimanual viewport adaptation method that keeps both proxies on screen in large environments. We also presented two bimanual haptic manipulation techniques that facilitate pick-and-place tasks: the joint control, which forces common control modes and control/display ratios for two interfaces grabbing an object, and the magnetic pinch, which simulates a magnet-like attraction between both hands to prevent unwanted drops of that object.

An experiment was conducted to assess the efficiency of these techniques for pick-and-place tasks, by comparing the double bubble with viewport adaptation to the clutching technique for extending the workspaces, and by measuring the benefits of the joint control and magnetic pinch.

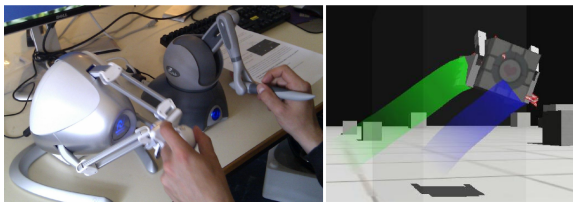


Figure 5. Bimanual pick-and-place task in a large virtual environment. A bimanual haptic setup made of two single-point devices (on left) allows to carry and displace a virtual cube using novel interactive techniques (on right).

6.2.5.3. Haptic Feedback and Haptic Seat for Enhancing AudioVisual Experience

This work aims at enhancing a classical video viewing experience by introducing realistic haptic feelings in a consumer environment.

First, in [16] a complete framework to both produce and render the motion embedded in an audiovisual content was proposed to enhance a natural movie viewing session. We especially considered the case of a first-person point of view audiovisual content and we propose a general workflow to address this problem. This latter includes a novel approach to both capture the motion and video of the scene of interest, together with a haptic rendering system for generating a sensation of motion. A complete methodology to evaluate the relevance of our framework was finally proposed and could demonstrate the interest of our approach.

Second, leveraging on the techniques and framework introduced previously, in [17] we could introduce a novel way of simulating motion sensations without calling for expensive and cumbersome motion plat- forms. The main idea consists in applying multiple force- feedbacks on the user's body to generate a sensation of motion while seated and experiencing passive navigation. A set of force-feedback devices are therefore arranged around a seat, as if various components of the seat could apply forces on the user, like mobile armrests or headrest. This new approach is called HapSeat (see Figure 6). A proof-of-concept has been designed within a structure which relies on 3 low-cost force-feedback devices, and two models were implemented to control them.

Results of a first user study suggests that subjective sensations of motion can be generated by both approaches. Taken together, our results pave the way to novel setups and motion effects for consumer living-places based on the HapSeat.

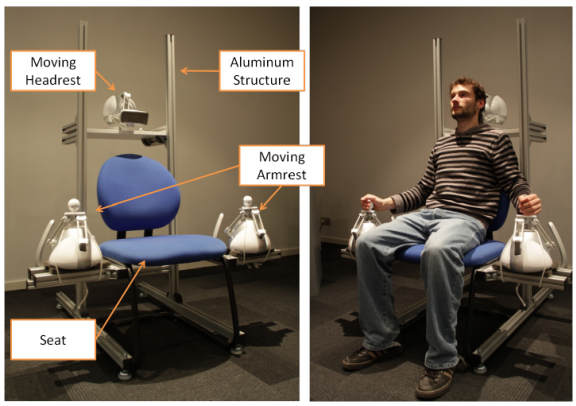


Figure 6. Prototype of the HapSeat. Left: seat structure with 3 force-feedback devices. Right: the system in use.

6.2.6. Interactions within 3D virtual universes

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

We have proposed some new metaphors allowing a guiding user to be fully aware of what the main user was seeing in the virtual universe and of what were the physical constraints of this user. We made a first prototype that made it possible to participate to the 3DUI 2012 contest [26], then we made further experiments showing the interest of the approach, these results will be presented in [25].

Our work focuses upon new formalisms for 3D interactions in virtual environments, to define what an interactive object is, what an interaction tool is, and how these two kinds of objects can communicate together. We also propose virtual reality patterns to combine navigation with interaction in immersive virtual environments. We are currently working about new multi-point interaction techniques to allow users to precisely manipulate virtuel objects.

6.3. Collaborative work in CVE's

6.3.1. The immersive interactive virtual cabin (IIVC)

Participants: Thierry Duval [contact], Valérie Gouranton [contact], Alain Chauffaut, Bruno Arnaldi, Cédric Fleury, Thi Thuong Huyen Nguyen, Georges Dumont.

We are still improving the architecture of our Immersive Interactive Virtual Cabin to improve the user's immersion with all his real tools and so to make the design and the use of 3D interaction techniques easier, and to make possible to use them in various contexts, either for different kinds of applications, or with different kinds of physical input devices.

The IIVC is now fully implemented in our two VR platforms: OpenMASK 5.1 and Collaviz 7.1.2 .

We have used the IIVC in order to provide efficient collaboration between users in a guiding task, allowing a guiding user to be fully aware of what the main user was seeing in the virtual universe and of what were the physical constraints of this user. We made a first prototype that made it possible to participate to the 3DUI 2012 contest [26], then we made further experiments showing the interest of the approach, these results will be presented in [25]. We also proposed to use the IIVC to enhance the communication between users sharing a virtual universe by helping them to build a cognitive model of the other users' environment [19]

6.3.2. *Generic architecture for 3D interoperability*

Participants: Thierry Duval [contact], Valérie Gouranton, Cédric Fleury, Rozenn Bouville Berthelot, Bruno Arnaldi.

Our goal is to allow software developers to build 3D interactive and collaborative environments without bothering with the 3D graphics API they are using. This work is the achievement of the IIVC software architecture. We have proposed PAC-C3D (Figure 7), a new software architectural model for collaborative 3D applications, in order to provide a higher abstraction for designing 3D virtual objects, and in order to provide interoperability, making it possible to share a virtual universe between heterogeneous 3D viewers.

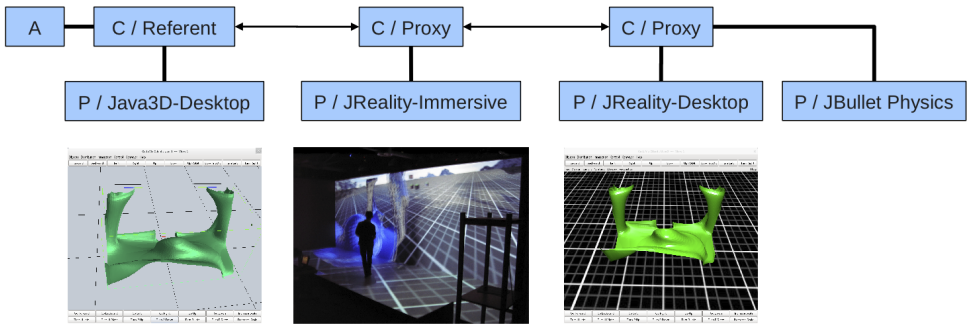


Figure 7. The PAC-C3D software architectural model makes interoperability possible between heterogeneous 3D viewers

We also study how to offer interoperability between virtual objects that are loaded in the same virtual environment but that are described using different formats. This is why we have proposed a generic architecture for enabling interoperability between 3D formats (Figure 8), the Scene Graph Adapter. Our SGA is now able to allow events coming from a 3D format to act upon data provided in another format, such as X3D events operating on Collada data, and makes also it possible to compose different format files [15].

6.3.3. *Collaborative interaction model*

Participants: Bruno Arnaldi, Valérie Gouranton [contact], Andrés Saraos Luna.

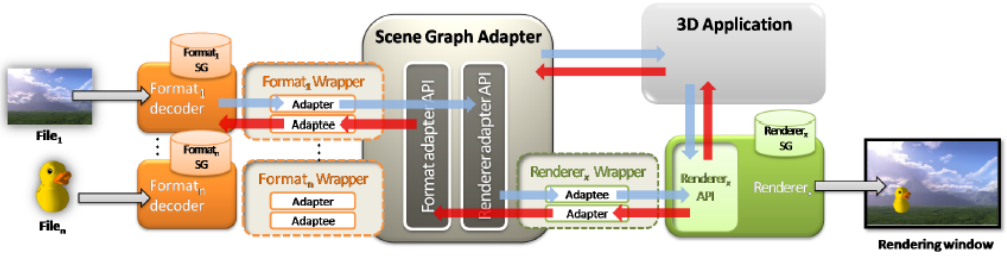


Figure 8. Our architecture allows the loading of any 3D graphics format simultaneously in any available rendering engine. The scene graph adapter is an interface that adapts a scene graph (SG) of a given format into a renderer scene graph and which also allows the rendering part to request this scene graph.

Our work ponders on collaborative interactions in Collaborative Virtual Environments for Training, with an emphasis on collaborative interactions between Real Humans and Virtual Humans working as a team. We propose [30] a new collaborative interaction model and from it construct a set of tools to describe and define such collaborative interactions [34].

6.4. Immersia Virtual Reality room

Participants: Georges Dumont [contact], Alain Chauffaut, Ronan Gaugne [contact], Marwan Badawi.

The team was the first in France to host a large-scale immersive virtual reality equipment known as Immersia (see figure 9). This platform, with full visual and sound immersion, is dedicated to real-time, multimodal (vision, sound, haptic, BCI) and immersive interaction. It will accommodate experiments using interactive and collaborative virtual-reality applications that have multiple local or remote users. Our new wall has four faces: a front, two sides and a ground. Dimensions are 9.6 m wide, 2.9 m deep and 3.1 m hight. The visual reproduction system combines ten Barco Galaxy NW12 projectors and three Barco Galaxy 7+ projectors. Visual images from Barco projectors are rendered on glass screens. They are adjusted for the users position, and this, together with their high resolution and homogeneous colouring, make them very realistic. The ART localization system, constituted of 16 ARTtrack 2 cameras, enables real objects to be located within the U-shape. Sound rendering is provided by a Yamaha processor, linked either to Genelec speakers with 10.2 format sound or Beyer Dynamic headsets with 5.1 virtual format sound, controlled by the users position.

The Immersia Virtual Reality room has been inaugurated on 2012, june, the 20th. We have hosted the project VR-GO, a Trans National Acces VISIONAIR project in june 2012. The goal was to evaluate an assembly by comparing different types of interaction : real, virtual and virtual + force feedback [28].

Figure 9. Immersia Virtual Reality Room

AXIS Project-Team

5. New Results

5.1. Introduction

As planned, our new results are splitted into our three sub-objectives as introduced below:

5.1.1. Mining for Knowledge Discovery in Information Systems

This year we get six main results: one related to how to integrate domain knowledge in a multi-view KDD process (cf. section 5.2.4), two on new KDD methods involving clustering (cf. sections 5.2.3 and 5.2.2), one on the construction of hierarchical structures of concepts in the field of e-tourism (cf. section 5.2.6), one on partitioning objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices (cf. section 5.2.1), and finally improvment of our work on critical edition of Sanskrit texts (cf. section 5.2.5).

- Zhang based on his thesis (2010) has published this year his work on modeling and clustering users with evolving profiles in usage streams [32]. This paper will propose three models to summarize bi-streaming data, which are the batch model, the Evolving Objects (EO) model and the Dynamic Data Stream (DDS) model. Through creating, updating and deleting user profiles, the models summarize the behaviours of each user as an object. Based on these models, clustering algorithms are employed to identify the user groups. The proposed models are tested on a real-world data set showing that the DDS model can summarize the bi-streaming data efficiently and effectively, providing better basis for clustering user profiles than the other two models.
- The work described in 2011(see our AxIS annual report) on critical edition of Sanskrit texts and submitted as a paper at the Cicling 2012 conference has been accepted [21].
- A past work accepted in an international journal with A. Ciampi and colleagues [16].
- One article in an international conference on functional data analysis issued from a collaboration with F. Rossi [40].
- Two articles have been deposit in the Computing Research Repository (CoRR): one on clustering Dynamic Web Usage data [65] from A. Da Silva's thesis and one on functional data analysis [66].

5.1.2. Information and Social Networks Mining for Supporting Information Retrieval

This year, we pursued two main works on clustering methods:

- the detection of communities in a social network (graph extracted from relationnal data) (cf. section 5.3.1),
- the improvment of our dynamic hard clustering method for relational data (cf. section 5.3.2).

5.1.3. Multidisciplinary Research For Supporting User Oriented Innovation

With the expansion of the innovation community beyond the firm's boundaries (the so-called "open innovation") a lot of changes have been introduced in design and evaluation processes : the users can become co-designers, HCI design and evaluation focus is no longer placed on usability only but also on the whole user experience, experimentations take place out of lab with large number of heterogeneous people instead of carefully controlled panels of users ... All these deep changes require improvements of existing practices, methods and tools for the design / evaluation of information systems as well as for usage analysis. This evolution calls also for a structured user centered methodology (methods and ICT tools) to deal with open innovation. Various different disciplines and trends are dedicated in understanding user behavior on Internet and with Digital Technologies, notably Human Computer Interaction community (HCI), CSCW, Workplace Studies, Distributed Cognition and Data Mining. Our contribution to open innovation research keeps its focus on usage analysis for design, evaluation and maintenance of information systems and our activities this year, as indicated in our roadmap presented at the Inria theme evaluation (2011) have been conducted both breadth wise and in depth with two main objectives :

- Improving, designing and evaluation support tools for innovation,
- Development of the FocusLab platform.

The research was conducted along three focus:

- Extension of usability methods and models (cf. section 5.4),
- Designing and evaluating user experience in the context of a living lab process (cf. section 5.5),
- FocusLab Platform (cf. section 5.6).

Let us note one research work related to Living labs done in 2011 and published in 2012 [26].

5.2. Mining for Knowledge Discovery in Information Systems

5.2.1. Clustering on Multiple Dissimilarity Matrices

Participants: Yves Lechevallier, F.A.T. de Carvalho, Guillaume Pilot, Brigitte Trousse.

In [17], we introduce hard clustering algorithms that are able to partitioning objects taking into account simultaneously their relational descriptions (relations + values) given by multiple dissimilarity matrices. The aim is to obtain a collaborative role of the different dissimilarity matrices in order to obtain a final consensus partition. These matrices could have been generated using different sets of variables and a fixed dissimilarity function or using a fixed set of variables and different dissimilarity functions, or using different sets of variables and dissimilarity functions.

During 2012 we show interest and disadvantages of these approaches to classifying curves with a Urso and Vichi distance based on the mathematical properties of curves (first derivative and second). The curves are issued from temperature sensors placed in 40 offices during one year (See section 6.1.3). This period was divided into the periods before and after challenge and the challenge period. During the challenge period the occupants had information by bonus / malus messages on energy consumption [34].

5.2.2. Web Page Clustering based on a Community Detection Algorithm

Participants: Yves Lechevallier, Yacine Slimani.

Extracting knowledge from Web user's access data in Web Usage Mining (WUM) process is a challenging task that is continuing to gain importance as the size of the Web and its user-base increase. That is why meaningful methods have been proposed in the literature in order to understand the behaviour of the user in the Web and improve the access modes to information. We pursued our previous work [102] and defined a new approach of knowledge extraction using graph theory. which is described in [29].

This work is done in collaboration with the laboratory LRIA At the Ferhat Abbas University, Sétif, Algérie.

5.2.3. Multi-criteria Clustering with Weighted Tchebycheff Distances for Relational Data

Participants: F.A.T. de Carvalho, Yves Lechevallier.

The method described in [27] uses a nonlinear aggregation criterion, weighted Tchebycheff distances, more appropriate than linear combinations (such as weighted averages) for the construction of compromise solutions. We obtain a partition of the set of objects, the prototype of each cluster and a weight vector that indicates the relevance of each criterion in each cluster. Since this is a clustering algorithm for relational data, it is compatible with any distance function used to measure the dissimilarity between objects.

5.2.4. Knowledge management in Multi-View KDD Process

Participant: Brigitte Trousse.

E.L. Moukhtar Zemmouri, in the context of his PhD thesis supervised by Hicham Behja, A. Marzark and B. Trousse pursued his work based on a Viewpoint Model in the context of a KDD process [30], [19].

Knowledge Discovery in Databases (KDD) is a highly complex, iterative and interactive process aimed at the extraction of previously unknown, potentially useful, and ultimately understandable patterns from data. In practice, a KDD process (data mining project according to CRISP-DM vocabulary) involves several actors (domain experts, data analysts, KDD experts, etc.) each with a particular viewpoint. We define a multi-view analysis as a KDD process held by several experts who analyze the same data with different viewpoints.
We propose to support users of multi-view analysis through the development of a set of semantic models to manage knowledge involved during such an analysis. Our objective is to enhance both the reusability of the process and coordination between users.
To do so, we propose first a formalization of viewpoint in KDD and a Knowledge Model that is "a specification of the information and knowledge structures and functions involved during a multi-view analysis". Our formalization, using OWL ontologies, of viewpoint notion is based on CRISP-DM standard through the identification of a set of generic criteria that characterize a viewpoint in KDD. Once instantiated, these criteria define an analyst viewpoint. This viewpoint will guide the execution of the KDD process, and then keep trace of reasoning and major decisions made by the analyst.
Then, to formalize interaction and interdependence between various analyses according to different viewpoints, we propose a set of semantic relations between viewpoints based on goal-driven analysis. We have defined equivalence, inclusion, conflict, and requirement relations. These relations allow us to enhance coordination, knowledge sharing and mutual understanding between different actors of a multi-view analysis, and reusability in terms of viewpoint of successful data mining experiences within an organization.

5.2.5. Critical Edition of Sanskrit Texts

Participants: Yves Lechevallier [correspondant], Marc Csernel, Ehab Assan.

With the help of Ehab Assan we improved the prototype made last year by Nicolas Bèchet (cf. 2011 AxIS activity report,[21]). It is now included in the construction process of critical editions of Sanskrit texts. Ehab also added LaTeX output to the process, we now have paper as well as Web output. It was possible to present these new features [33], [36] at the 13th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing) in Delhi.

5.2.6. Construction and Settlement of hierarchical Structures of Concepts in E-tourism

Participant: Yves Lechevallier.

The work of Nicolas Bechet (AxIS member in 2011) and Yves Lechevallier in collaboration with Marie-Aude Aufaure (Ecole Centrale de Paris), was published in 2012 [20] related to a method for the construction and the automatic settlement of hierarchical structures of concepts. We were particularly interested in the construction of a hierarchical structure of services offered in Hotels from a data set of an application in the field of e-tourism motivated by our contacts with the SME Addictrip. The goal is to associate to each service a concept that provides a common representation of all services. Our experiments are carried out using resources from partners specialized in online hotel booking, in particular from Addictrip. The establishment of a structure of concepts is essential to these partners that use their own terminologies description of hotel services. Indeed it provides a common representation space allowing the comparison of service coming from different resources. Our approach is based on proximity of literal terms in the service having a nearby measure based on n-grams of characters. The results during our experiments show the quality of this approach and its limitations.

5.3. Information and Social Networks Mining for Supporting Information Retrieval

5.3.1. Clustering of Relational Data and Social Network Data

Participants: Yves Lechevallier, Amine Louati, Bruno Almeida Pimentel.

The automatic detection of communities in a social network can provide a kind of graph aggregation. The objective of graph aggregations is to produce small and understandable summaries and it can highlight communities in the network, which greatly facilitates the interpretation.

Social networks allow having a global view of the different actors and different interactions between them, thus facilitating the analysis and information retrieval.

In the enterprise context, a considerable amount of information is stored in relational databases. Therefore, relational database can be a rich source to extract social network.

During this year many updates of the program developed by Louati Amine in 2011 were performed by Bruno Almeida Pimentel. A book chapter, included the new aggregation criteria proposed ans evaluted by Bruno Almeida Pimentel, was written and will be published in 2013.

This work is done in collaboration with Marie-Aude Aufaure, head of the Business Intelligence Team, Ecole Centrale Paris, MAS Laboratory.

5.3.2. Multi-View Clustering on Relational Data

Participants: Thierry Despeyroux, Yves Lechevallier.

In the work reported in [23] in collaboration with Francisco de A.T. de Carvalho, we introduce an improvement of a clustering algorithm described in [17] that is able to partition objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices. In this version of the prototype clusters depend on the variables of the representation space. These matrices could have been generated using different sets of variables and dissimilarity functions. This method, which is based on the dynamic clustering algorithm for relational data, is designed to provided a partition and a vector of prototypes for each cluster as well as to learn a relevance weight for each dissimilarity matrix by optimizing an adequacy criterion that measures the fit between clusters and their representatives. These relevance weights change at each algorithm iteration and are different from one cluster to another. Moreover, various tools for the partition and cluster interpretation furnished by this new algorithm are also presented.

Two experiments demonstrate the usefulness of this clustering method and the merit of the partition and cluster interpretation tools. The first one uses a data set from UCI machine learning repository concerning handwritten numbers (digitalized pictures). The second uses a set of reports for which we have an expert classification given a priori. which we have an expert classification given a priori.

5.4. Extension of Usability Methods and Tools

5.4.1. User Evaluation and Tailoring of Personal Information

Participants: Claudia Detraux, Dominique Scapin.

In the context of the ANR project PIMI (Personal Information Management through Internet see section 6.2.1), ergonomic inspections have been carried out to evaluate the usability of the PIMI V0.1. prototype, in its PC and mobile versions [49], [48]. Also, an experiment [24], [35] was conducted on a mockup of a Personal Information Space. Users were asked to perform scenario-based data entry and retrieval tasks, then to modify the mockup according to their wishes and needs. The results allowed to validate the item content and structure for the future personal space, as well as to assess the role of user modifications as evaluation cues, and for the development of further ergonomic recommendations. Detailed information was obtained on how users enter and retrieve data, by modifying the interface settings and shows that the adaptable nature of a Personal Information Space can indeed influence its acceptance, and provides useful cues for ergonomic evaluation

5.4.2. Usability Methods for Information Visualization

Participant: Dominique Scapin.

A collaboration between UFRGS (Federal University of Rio Grande do Sul, Institute of Informatics), Brazil and Inria-AxIS led to a book chapter [37] dealing with potential methodologies for including a user-centered approach into information visualization techniques. It starts by presenting the evolution of visualization techniques evaluation, briefly summarizing the main contributions in this area since its humble beginning as a collateral activity until the recent growth of interest. Then, the focus is on current issues related to such evaluations, particularly concerning the way they are designed and conducted, taking into account a background of well-known usability evaluation methods from HCI to help understanding why there are still open problems. A set of guidelines for a (more) user-centered usability evaluation of information visualization techniques is proposed and discussed.

5.4.3. Usability Recommendations for MIS (Mixed Interactive Systems)

Participant: Dominique Scapin.

A collaboration between University of Toulouse – IUT Tarbes, IRIT and Inria-AxIS led to a book chapter [38] dealing with Mixed Interactive Systems (MIS) which denote an advanced form of interaction that aims at combining physical and digital worlds, such as mixed and augmented reality, tangible user interfaces, ubiquitous computing, etc. Their main interest relates to the use of physical artifacts from the user's activity customary context. The book chapter first reports on a systematic review of the literature on MIS evaluation. From that review, usability recommendations were selected and deciphered before reformulating them under a common format. Finally, three different classification schemes of the usability recommendations obtained are proposed to facilitate search and retrieval, but also to better integrate them into the MIS development process.

5.5. Designing and Evaluating User Experience and Methods for Open Innovation

5.5.1. From Usability to User Experience: an HCI Review

Participants: Dominique Scapin, Bernard Senach, Brigitte Trousse, Marc Pallot.

Through an extensive review of the literature, a paper [28] attempted to characterize a rather novel and popular view on human-computer interaction: User Experience (UX). After introducing its polysemous nature, this paper describes the origins of UX, its scope, underlying concepts and components, as well as its various definitions. Then, UX methods are surveyed and classified, distinguishing processes, frameworks, and specific methods. The paper identifies a set of issues about the needs for increased UX maturity. Even though UX can still be viewed as an extension of usability, its future may correspond to a paradigm evolution rather than simply a buzz word. The evolution is not drastic, but it adds complexity (including new measurements) by considering a few more user areas than traditional usability.

5.5.2. Evaluation of our Methods for Idea Generation Process

Participants: Anne-Laure Negri, Caroline Tiffon, Brigitte Trousse, Bernard Senach.

In 2011 we proposed a methodology coupling two methods [25] (GenIoT a generative method based on probes (fake sensors and/or actuators) and ALoHa! a bodystorming method for designing service concepts in the specific paradigm of the Internet of Things (IoT). In the frame of the European project ELLIOT - Experiential Living Lab for Internet Of Things -, ICT Usage Lab (cf. section 6.1.8) aims at co-creating "green" services, i.e. services based on air quality and noise measurement.

Both IoT ideation methods Aloha! and GenIoT were used for the co-creation of health related services (cf. section 6.3.1.1). The participants of the methods were Environment and Health professional. Results were very different than the workshops run with citizen in the frame of the mobility scenario in ELLIOT (see 2011 AxIS activity report). Comparison of these workshops shows that hybrid approaches –i.e. co-creation approaches mixing both real and virtual meetings are not working as well as pure face to face or pure online approaches. Moreover, GenIoT method seems to be more effective with citizen than with professional. Aloha! is effective in both groups but more efficient with professionals. However the participant experience of Aloha! is higher in the case of citizen (mainly because professionals are not used to practice creative thinking methods and do not appreciate to go out of their comfort zone).

5.5.3. Leading People Behavior Changes: Mining Evolutive Data

Participants: Brigitte Trousse, Yves Lechevallier, Guillaume Pilot, Carole Goffart, Bernard Senach.

The ECOFFICES project (cf. [62], [22] and section 6.1.3) was for AxIS project team our first step towards eco-behavior study. It provided us a very rich context to study how to analyse the evolution of the energy consuming of employees during an energy challenge. A qualitative analysis from questionnaires (before and after the challenge) has been done as well quantitative analysis. The data set for quantitative data is composed of heterogeneous data issued from around 400 sensors (temperature, presence, behavior in terms of opening doors, windows, bonus, malus, etc.). We made different studies related to data preprocessing and data analysis. In our first study [64], we cleaned the data set and selected reliable data for data analysis (only temperature of various equipments, user presence and bonus/malus points). We decided not to work with aggregated variables such as the initial ponderation (defined by partners) for the various bonus-malus rules and the energy consuming at the office level. We decided to use (office, day) as statistical unit (i.e. 9995 units) with a vectoral representation. Finally we realized that the three initial periods (before the challenge, during and after the challenge) on 379 days (2011-2012) should be in fact decomposed in five periods, due to the fact the first and the last periods were split into two subperiods (with and without heat). For the analysis, we apply for each (office,day) a first analysis on a vectorial representation of temperature with the MND method (cf. section focuslab) in order to identify the best partition of these. The MND method uses euclidean distance between each value of the vectorial representation and the prototypes are defined by the means. Second we did a clustering of these units based on bonus and malus and finally we made the correspondence between these two partitions. Three classes for (office, day) are obtained. The interpretation in terms of team relied difficult but we proposed various conclusions for a winner for managing a specific bonus, or in managing ambient temperature or in behavioral change.
In our second study [34] in collaboration with Francisco de A.T. De Carvalho, our goal was to improve the interpretation task at the office and team level by applying AxIS advanced methods. To do this, we applied our hard clustering method presented in [34] on this dataset where each office was characterized by two different representations:

- Interval representation: each office is characterized by a vector of intervals corresponding to the average, minimum and maximum of daily temperatures on the three temperature sensors during these five periods. Then the office is represented by a vector of 15 intervals and the distance used is Hausdorff distance. This classification is consistent with the partition into three classes obtained during the ECOFFICES project. The class obtained with nine eco-responsible ecoffices is the same. However, other offices are divided into two classes according to the type of heating used during the winter period. The classical method divided these offices into two clusters, one of which contains the offices using the radiators during the winter period.
- Sequential representation: Each office is characterized by a vector of 9 measures, the min, max and average of daily temperatures of the three sensors in these five periods. The values are ordinate versus time and the distance used is Urso and Vichi distance (adequate for curves). The results of this approach are quite different from the classical approach results. These results required more effort for their interpretation in collaboration by specialists.

5.5.4. Leading People Behavior Changes: a Literature Review

Participants: Bernard Senach, Anne-Laure Negri.

Our research towards eco-behavior study started with the ECOFFICES project (cf. sections 5.5.3 and 6.1.3 for more details) was recently complemented with a litterature review aiming at a deeper understanding of breaks and levers to eco behavior adoption. A first work was focused on the so-called "modal change problem", compiling methods and tools aimed at supporting people to use public transportation system rather than their personal car. A second work was initiated to get a better understanding of the role that users interface could play in encouraging people to adopt a specific behavior. This work is still in progress.
Eco mobility : prompting people to adopt public transportation mode rather than their personal car.

The first review of work conducted in the fields of Persuasive Communication, Commitment, Nudges and Persuasive Technology showed that behavioral change is a process with many steps requiring to support each step with specific means. For instance, if mass communication can support the public awareness of a problem, information is not sufficient to convince people to really change their behavior. It is necessary to push them to act and numerous well-known influence techniques are nowadays available. All recent technological development (geo localization, mobile devices, social networks) can provide very effective support for behavioral changes as far as they rely on design principles identified by research in Persuasive Technology. A presentation was done on this topic for GreenCode Forum [67] (see the video on youtube).

5.5.5. *Future of Internet and User-Open Innovation for Smart Cities*

Participants: Marc Pallot, Brigitte Trousse, Bernard Senach.

We pursued our work on this topic and contributed to a white paper [59] which is one of the main outcomes of the FIREBALL project [cf. section 6.3.1.2), a Coordination Action within the 7th Framework Programme for ICT, running in the period 2010-2012. The aim of this project was to bring together communities and stakeholders who are active in three areas, namely: research and experimentation on the Future Internet (FIRE); open and user-driven innovation in Living Labs; and urban development for smarter cities. The goal was to develop a common vision on how the different methodologies and concepts in these areas can be aligned for cities as playgrounds of open and user driven innovation related to the Future Internet.

The white paper addresses several aspects that are critical for understanding the 'smart city' concept and the current progress in this area. Based on cases studies and foresight reports we aim to shed light on how the concept of smart city is currently adopted by European Cities and what the ambitions and expectations are in using this concept. It investigates the drivers and bottlenecks that influence the transformation towards a "smart city". Underlying approaches to smart cities are discussed, both in terms of the strategies and planning approaches. From this point of view, this paper explores the conditions that must be established to stimulate the transformation towards smart cities, and the resources that are available or should be made available such as investments in broadband networks and in smart applications, as well as in the capabilities to innovate. This also points to the changing structures and processes of innovation and city development. Interestingly, we see a tendency towards more decentralized and bottom-up approaches to planning and innovation. Innovation ecosystems are characterized by a combination of top down and bottom up initiatives, leading to networking and collaboration among stakeholders, which eventually extend to real innovation communities. Increasingly, citizens, advanced businesses and local governments act as proactive catalysers of innovation, shaping cities as "agents of change".

5.6. FocusLab Platform

5.6.1. *FocusLab platform: software part*

Participants: Brigitte Trousse, Yves Lechevallier, Semi Gaieb, Xavier Augros, Guillaume Pilot, Florian Bonacina.

FocusLab v1.3 (software component) done inside the ELLIOT project (cf. section 6.3.1.1) and for the purposes of the CPER Telius (cf. section 6.1.5) corresponds to the design and the implementation of a set of web-services providing basic and advanced functionalities for data analysis and some other tools supporting the living lab process.

In this version, five data analysis web services are proposed including three generic web services: a classical linear regression and two AxIs methods:

- SMDS/SCDS [91]: SCDS (Sequence Clustering in Data Stream) is a clustering algorithm for mining sequential patterns (Java) in data streams developed by A. Marascu during her thesis. This software takes batches of data in the format "Client-Date-Item" and provides clusters of sequences and their centroids in the form of an approximate sequential pattern calculated with an alignment technique. We propose in this version to return the apparition frequency (min, max, average, slope) of a sequential pattern from data streams (SCDS algorithm) (see references

- GEAR for data streams compression [93], [91], [92], [94]: GEAR (REGLO in french) is an implementation of the history management strategy proposed in Marascu's thesis [1]. It takes a set of time series and provides a memory representation of these series based on a new principle, where salient events are important (in contrast to the recent events of decaying models) .

Other data analysis services and tools have been added for Living Labs needs. We propose also two clustering methods which must be downloaded as standalone software and used for mining data from living labs:

- ATWUEDA (Axis Tool for Web Usage Evolving Data Analysis) for Analysing Evolving Web Usage Data (Da Silva 'thesis 2009 [79], [83], [81], [82]) was developed in Java and uses the JRI library (http://www.r-project.org/). The ATWUEDA tool is able to read data from a cross table in a MySQL database, split the data according to the user specifications (in logical or temporal windows) and then apply the proposed approach in order to detect changes in dynamic environment. Such an approach characterizes the changes undergone by the usage groups (e.g. appearance, disappearance, fusion and split) at each timestamp. Graphics are generated for each analysed window, exhibiting statistics that characterizes changing points over time. This application for the next experiment of Green services use case is under study.
- MND method (Dynamic Clustering Method for Multi-Nominal Data) [90]: The proposed MND method (developed in C++ language) determines iteratively a series of partitions which improves at each step the underlying clustering criterion. The algorithm is based on: a) Prototypes for representing the classes; b) Representation space; c) Proximities (distances or similarities) between two individuals; d) Context-dependent proximity functions for assigning the individuals to the classes at each step. The clustering criterion to be optimized is based on the sum of proximities between individuals and the prototype of the assigning clusters.

  This method has been also successfully applied on Web logs in 2003. This year we improved our code and tested it on IoT data (temperature) issued from the ECOFFICES project (cf. sections 5.5.3 and 6.1.3).

The application of the services provided by FocusLab 1.3 and other AxIS data mining methods for the purposes of ELLIOT use cases and other experimental projects are under study.

DAHU Project-Team

5. New Results

5.1. Distributed data management

Participants: Serge Abiteboul, Émilien Antoine, Cristina Sirangelo, Nadime Francis, Luc Segoufin.

Distributed knowledge base. We are developing the system Webdamlog [16], [13], [14] to address the challenges faced by everyday Web users, who interact with inherently heterogeneous and distributed information. Managing such data is currently beyond the skills of casual users. In Webdamlog, we see the Web as a knowledge base consisting of distributed logical facts and rules. The objective is to enable automated reasoning over this knowledge base, ultimately improving the quality of service and of data. The system supports the Webdamlog language, a Datalog style language with rule delegation.

Deduction in uncertain worlds. Motivated by reasoning in distributed environments in which disagreements arise between different actors, we study in [17] deduction (captured by datalog programs) in the presence of inconsistencies (induced by functional dependency (FD) violations). We adopt an operational semantics for datalog with FDs based on inferring facts one at a time, while never violating the FDs. This yields a set of possible worlds that we capture by c-tables of possibly exponential size. We propose to use probabilities to measure this nondeterminism and define a probabilistic semantics that can be captured by probabilistic conditional tables. Not surprisingly, we show that computing the probability of a query answer in our setting is expensive, which leads us to introduce a sampling algorithm to estimate answer probabilities. We then turn our attention to the problem of explaining why a particular answer holds. This leads us to consider two novel notions: the most influential extensional facts, and the most likely proofs for an answer. We study algorithms for ranking facts and proofs based on their contribution to the derivation of an answer. Finally, we consider how our framework can be adapted to a distributed setting, and in particular, how sampling can be performed in a distributed manner.

Access rights in a distributed setting. We started considering access right issues in Webdamlog. This is related to specifying access right on views in standard databases. There is also the issues of controlling rules that are run locally but were specified by other peers.

Incomplete information in Web data. Incomplete information often arises from the integration of different Web data sources, as well as from the exchange of data between communicating Web applications. The semantics of incompleteness (i.e. which possible complete databases are represented by an incomplete one) depends on the context and the particular scenario where incompleteness raises from. We have studied how to deal with the presence of incomplete information under different possible semantics. We have in particular studied in which condition it is possible to query incomplete data “naively", i.e. as if it were complete. We have exhibited “natural" fragments of first order logic for which naive evaluation is possible, under different semantics.

Graph data management. Graph structured data can be found in new emerging applications such as RDF and linked data, or social networks. The peculiarity of queries over graphs is that they are interested in both data carried by the graph and in the graph topology; they are often based on reachability patterns. In a distributed setting it is very common to be able to query only a partial description or a "view" of the graph. We studied the problem of answering queries using only the information provided by the views. The presence of a form of recursion in views and queries presents new challenges. We found restricted classes of graph views and queries that allow efficient query answering over views.

5.2. Tree automata theory

Participants: Luc Segoufin, Serge Abiteboul, M Praveen.

Tree automata We studied the expressive power of a subclass of regular tree languages. We gave a decidable characterization of those languages that are "piecewise testable", i.e. definable using boolean combination of existential first-order formulas [12].

Automata with counters. We studied extending techniques used in standard Petri nets to other models. We extended the Rackoff technique to decide coverability and boundedness problems for Strongly Increasing Affine nets, a subclass of Affine nets [20].

Languages on trees. We studied in [18] highly expressive query languages for unordered data trees, using as formal vehicles Active XML and extensions of languages in the while family. All languages may be seen as adding some form of control on top of a set of basic pattern queries. The results highlight the impact and interplay of different factors: the expressive power of basic queries, the embedding of computation into data (as in Active XML), and the use of deterministic vs. nondeterministic control. All languages are Turing complete, but not necessarily query complete in the sense of Chandra and Harel. Indeed, we show that some combinations of features yield serious limitations, analogous to FOk definability in the relational context. On the other hand, the limitations come with benefits such as the existence of powerful normal forms. Other languages are "almost" complete, but fall short because of subtle limitations reminiscent of the copy elimination problem in object databases.

Probabilistic XML. In [15], we study the problem of, given a corpus of XML documents and its schema, finding an optimal (generative) probabilistic model, where optimality here means maximizing the like- lihood of the particular corpus to be generated. Focusing first on the structure of documents, we present an efficient algorithm for finding the best generative probabilistic model, in the absence of constraints. We further study the problem in the presence of integrity constraints, namely key, inclusion, and domain constraints. We study in this case two different kinds of generators. First, we consider a continuation-test generator that performs, while generating documents, tests of schema satisfiability ; these tests prevent from generating a document violating the constraints but, as we will see, they are computationally expensive. We also study a restart generator that may generate an invalid document and, when this is the case, restarts and tries again. Finally, we consider the injection of data values into the structure, to obtain a full XML document. We study different approaches for generating these values.

Infinite alphabet. We studied the complexity of satisfiability of linear temporal logics extended to reason about repetitions of values from an infinite data domain. We refined an existing result that reduced this problem to Petri net reachability, and showed that it can be reduced to the coverability problem. Using this refinement, we gave the precise complexity of the satisfiability problem. We also characterized the complexity of satisfiability for many fragments and extensions of the logic.

DREAM Project-Team

6. New Results

6.1. Diagnosis of large scale discrete event systems

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

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

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

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

This year, we have improved the prototype and worked on a journal paper that will be submitted in 2013.

6.1.2. Scenario patterns for exploring qualitative ecosystems

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

The approach has been experimented on a marine ecosystem under fishing pressure. The model describes the trophodynamic interactions between fish trophic groups as well as interactions with the fishery activities and with an environmental context. A paper has been accepted for publication in the Environmental Modelling Software Journal [52].

6.1.3. Controler synthesis for dealing with “How to” queries

We extended the approach to deal with “How to” queries. As before, we rely on a qualitative model in the form of timed automata and on model-checking tools to answer queries. We proposed and compared two approaches to answer questions such as “How to avoid a given situation ?”(safety query). The first one exploits controller synthesis and the second one is a “generate and test” approach. We evaluated these two approaches in the context of an application that motivates this work, i.e the management of a marine ecosystem and the evaluation of fishery management policies. The results have been accepted for publication in [17].

More recently, we use similar methodological tools to model herd management on a catchment and analyse the best/optimal farming practices in order to reduce nitrate pollution due to livestock effluents. An hybrid model has been built using hierarchical timed automata. Scenarios can already be simulated and evaluated. We currently work on adapting controler synthesis tools in order to get the best strategies. This work is made in collaboration with our colleagues of INRA.

6.2. Machine learning for model acquisition

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

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

6.2.1. Mining temporal patterns with numerical information

We are interested in mining interval-based temporal patterns from event sequences where each event is associated with a type and time interval. Temporal patterns are sets of constrained interval-based events. This year we have begun to work on multiscale temporal abstraction to represent time series by codewords at different temporal and amplitude scales. We have improved the method of Wang et al. [70] by introducing Dynamic Time Warping to compute better codewords for time series abstraction. The codeword-based time series representation is then used by QTIPrefixSpan [3] to extract temporal patterns. A paper is in preparation. We are also working on a multivariate version of the method for mining multivariate temporal patterns at different resolution levels.

6.2.2. Incremental sequential mining

Sequential pattern mining algorithms operating on data streams generally compile a summary of the data seen so far from which they compute the set of actual sequential patterns. We propose another solution where the set of actual sequential patterns are incrementally updated as soon as new data arrive on the input stream. Our work stands in the framework of mining an infinite unique sequence. Though being of great importance, this problem has not received a similar attention as mining from a transaction database. Our method [13] provides an algorithm that maintains a tree representation (inspired by the PSP algorithm [56]) of frequent sequential patterns and their minimal occurrences [54] in a window that slides along the input data stream. It makes use of two operations: deletion of the itemset at the beginning of the window (obsolete data) and addition of an itemset at the end of the window (new data). The experiments were conducted on simulated data and on real data of instantaneous power consumption. The results show that our incremental algorithm significantly improves the computation time compared to a non-incremental approach [14].

6.2.3. Incremental learning of preventive rules

The problem is to learn preventive rules in order to avoid malfunctioning on smartphones. A monitoring module is embedded on the phones and sends reports to a server. Reports are labeled with a normal or abnormal label. From this set of reports new rules are learned. As a lot of smartphones are supervised, it is impossible to store all the reports. Therefore incremental learning has to be used.

Last year, we achieved two main tasks: a report database has been built in order to test the future algorithms, and a new algorithm [20] has been developed for implementing an incremental version of the algorithm AQ21 [72].

6.2.4. Multiscale segmentation of satellite image time series

Satellite images allow the acquisition of large-scale ground vegetation. Images are available along several years with a high acquisition frequency (1 image every two weeks). Such data are called satellite image time series (SITS). In [12] , we present a method to segment an image through the characterization of the evolution of a vegetation index (NDVI) on two scales: annual and multi-year. We test this method to segment Senegal SITS and compare our method to a direct classification of time series. The results show that our method using two time scales better differentiates regions in the median zone of Senegal and locates fine interesting areas (cities, forests, agricultural areas).

6.2.5. Mining a big unique graph for spatial pattern extraction

Researchers in agro-environment needs a great variety of landscapes to test the agro-ecological models of their scientific hypotheses. As the representation of real landcapes necessitates lots of on-land measures, good big representations are difficult to acquire. Working with landscape simulations is then an alternative to get a sufficient variety of experimental landscapes. We propose to extract spatial patterns from a well described geographic area and to use these patterns to generate realistic landscapes. We have begun the exploration of graph mining techniques to discover the relevant spatial patterns present in a graph expressing the spatial relationships between the agricultural plots as well as the roads, the rivers, the buildings, etc., of a specific geographic area.

This year, we have been working on extending algorithm gSPAN [73] with an adaptive support threshold and with a taxonomy to be able to extract interesting patterns involving agricultural plots with rare features. We plan to submit a paper in 2013.

6.3. Decision aiding with models and simulation data

Participants: Louis Bonneau de Beaufort, Tassadit Bouadi, Marie-Odile Cordier, Véronique Masson, Florimond Ployette, René Quiniou, Karima Sedki.

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

6.3.1. Exploring models thanks to scenarios: a generic framework

In the framework of the Appeau project (see 7.2.1) a paper describing a generic framework for scenario exercises using models applied to water-resource management, has been written in cooperation with all the partners and published in Environmental Modelling and Software [5].

6.3.2. A datawarehouse for simulation data

The ACASSYA project 7.2.2 aims at providing experts or stakeholders or farmers with a tool to evaluate the impact of agricultural practices on water quality. As the simulations of the deep model TNT2 are time-consuming and generate huge data, we have proposed to store these simulation results in a datawarehouse and to extract relevant information, such as prediction rules, from the stored data. We have devised a general architecture for agro-environmental data on top of the framework Pentaho.

This year we have been working on the efficient computation of OLAP queries related to realistic scenarios proposed by experts in the domain. Precisely, we have devised indexing schemes to access the data in the OLAP cube. We have also worked on the visualization by a GIS (Geographical Information System) of the query results on maps of the geographical area under interest. A paper will be submitted to the COMPAG Journal in beginning 2013.

6.3.3. Efficient computation of skyline queries in an interactive context

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

This year we have worked at improving the formal setting of the partial materialization of skyline queries when dynamic preferences are refined online by the user. We have explicited which parts of the skyline evolve (which point are added or removed) when a new dimension is introduced in the computation. This led to an efficient incremental method for the online computation of the skyline corresponding to new user preferences [9]. An extended version of this paper is under submission to the Journal "Transactions on Large Scale Data and Knowledge Centered Systems" (TLDKS).

We are working now on a hierachical extension of our method that could be introduced in a datawarehouse context.

6.3.4. Influence Diagrams for Multi-Criteria Decision

For multi-criteria decision-making problems, we propose in [7] a model based on influence diagrams able to handle uncertainty, represent interdependencies among the different decision variables and facilitate communication between the decision-maker and the analyst. The model makes it possible to take into account the alternatives described by an attribute set, the decision-maker's characteristics and preferences, and other information (e.g., internal or external factors) that influence the decision. Modeling the decision problem in terms of influence diagrams requires a lot of work to gather expert knowledge. However, once the model is built, it can be easily and efficiently used for different instances of the decision problem. In fact, using our model simply requires entering some basic information, such as the values of internal or external factors and the decision-maker's characteristics.

6.3.5. Modeling influence propagation by Bayesian causal maps

The goal of this project is modeling shellfish fishing to assess the impact of management pollution scenarios on the *Rade de Brest*. Cognitive maps were built from interviews with fishermen. To represent and reason about these cognitive maps, we propose to use Bayesian Causal Maps making use of fishermen knowledge, particularly to perform influence propagation [11].

However, this model does not take into account the variety of influences asserted by the fishermen, but only the "mean" causal map. This year we have been working on an approach that could combine individual knowledge with belief functions in the way of Philippe Smets's Transferable Belief Model [67].

This work is done in the framework of the RADE2BREST project, involving Agrocampus Ouest and CNRS (GEOMER/LETG), funded by "Ministère de l'Ecologie" (This project is not mentioned in section 7.2 because DREAM is not an official partner of this project.).

6.3.6. Mining simulation data by rule induction

In the framework of the SACADEAU project (see 7.2.1), a paper dedicated to mining simulation data by rule induction has been published in the COMPAG Journal [8]. Both qualitative and quantitative predictions from a model of an agro-environmental system are analysed. Two approaches in rule learning from spatial data (ILP and attribute-value approaches) are compared and show that results help identify factors with strong influence on herbicide stream-pollution.

We have also participated in a collaboration for modeling the effects crop rotations the results of which were published in the Science of the Total Environment Journal [6].

6.4. Diagnostic and causal reasoning

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

Stemming on [29], [30], [31], [32], [33], we have designed an inference system based on causal statements. This is related to diagnosis (observed symptoms explained by faults). The aim is to produce possible explanations for some observed facts. Previously existing proposals were ad-hoc or, as in [36], [47], they were too close to standard logic in order to make a satisfactory diagnosis. A key issue for this kind of work is to distinguish logical implication from causal links and from ontological links. This is done by introducing a simple causal operator, and an *is-A* hierarchy. These two operators are added to a restricted first order logic of the Datalog kind (no function symbols). Then, our system produces elementary *explanations* for some set of observed facts. Each explanation links some facts to the considered observation, together with a set of atoms called the *justifications*: The observation is explained from these facts, provided the justifications are possible (not contradicted by the available data). This formalism has also been translated into answer set programming [57], [58]).

This year, we have extended our formalism in order to deal with more complex problems such as finding explanations for the hurricane Xynthia (2010, February 28). In such situations, there are many data and many possible elementary explanations can be examined. This involves an extension of our formalism, in order to deal with more complex chains of causations and *is-A* links. We are on the way to end this task. Our formalism makes precise what all these possible explanations are. Then, in order to deal with so many possible complex explanations, we integrate this causal formalism into an argumentation framework. Logic-based formalizations of argumentation [34] take pros and cons for some conclusion into account. These formalizations assume a set of formulae and then exhaustively lay out arguments and counterarguments. This involves providing an initiating argument for the inference and then providing undercuts to this argument, and then undercuts to undercuts. So here our causal formalism provides a (rather large) set of explanations, and the argumentation part allows to select the best ones, under various criteria.

Then, since answer set programming can easily deal with logical formalisms, the argumentation part will be incorporated into our already existing answers set programming translation of the causal formalism. Regarding answer set programming, we have also examined some more difficult examples [16] and participated to a chapter in the to be published "Panorama de l'intelligence artificielle. Ses bases méthodologiques, ses développements" [19].

EXMO Project-Team

6. New Results

6.1. Ontology matching and alignments

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

6.1.1. Evaluation

Evaluation of ontology matching algorithms requires to confront them with test ontologies and to compare the results. Since 2004, we run the Ontology Alignment Evaluation Initiative (OAEI) which organises evaluation campaigns for assessing the degree of achievement of actual ontology matching algorithms [2].

This year, we ran two evaluation campaigns named 2011.5 and 2012. This was justified by the will to complete full evaluations using the support of the SEALS platform. Hence, the main activities carried out in 2012 were related to the automation of the evaluation. This involved providing participants with a better way to bundle their tools so that they can be evaluated both offline and within the SEALS platform. It also required to support more organisers to provide test case within the plaform.

This work has been used in the OAEI 2012 evaluation campaign. OAEI 2012 offered 9 different test sets (7 of which under the SEALS platform). This issue brought the following results:

- More participants than ever (21);
- All ontology matchers running on the SEALS platform (18);
- Increased performances in terms of precision and recall;
- Matchers are now very scalable and can deal with the largest available ontologies (9 systems able to deal with the very large medical ontology SnoMed);

We have also introduced as a data set, the benchmark for multilingual ontology matching developed last year [6]. It has pushed matcher developers to address multilingual issues.

The participating systems and evaluation results were presented in the 7th Ontology Matching workshop, that was held in Boston, MA, US [22], [7]. More information on OAEI can be found at http://oaei.ontologymatching.org/.

6.1.2. Semantics for weighted correspondences

Alignment correspondences are often assigned a weight or confidence factor by matchers. Nonetheless, few semantic accounts have been given so far for such weights. We have proposed a formal semantics for weighted correspondences between different ontologies. It is based on a classificational interpretation of correspondences: if o and o' are two ontologies used to classify a common set X, then alignments between o and o' are interpreted as encoding how elements of X classified in the concepts of o are re-classified in the concepts of o', and weights are interpreted as measures of how precise and complete re-classifications are. This semantics is justifiable for extensional matchers. We have proven that it is a conservative extension of a semantics of absolute correspondences, and we have provided properties that relate correspondence entailment with description logic constructors [8].

This work has been made in cooperation with Alexander Borgida (Rutgers University) and Chiara Ghidini and Luciano Serafini (Fondazione Bruno Kessler).

6.2. Data interlinking

The web of data uses semantic web technologies to publish data on the web in such a way that they can be interpreted and connected together. It is thus critical to be able to establish links between these data, both for the web of data and for the semantic web that it contributes to feed.

6.2.1. Keys and pseudo-keys detection for web datasets cleansing and interlinking

We have proposed a method for analysing web datasets based on key dependencies. For this purpose, we have adapted the classical notion of a key in relational databases to the case of RDF datasets [9], [16]. In order to better deal with web data of variable quality, we have introduced the definition of a pseudo-key. We have also provided an RDF vocabulary for representing keys and pseudo-keys and designed and implemented an algorithm for discovering them. Experimental results show that, even for a large dataset such as DBpedia, the runtime of the algorithm is still reasonable. We are currently working on two applications: data cleansing, i.e., detection of errors in RDF datasets and recovery, and datasets interlinking.

The algorithm is publicly available at https://gforge.inria.fr/projects/melinda/.

6.2.2. Data interlinking from expressive alignments

In the context of the DATALIFT project (see §7.1.1), we are developing a data interlinking module. Based on our analysis of the relationships between ontology matching and data interlinking [15], our goal is to generate data interlinking scripts from ontology alignments. For that purpose, we have integrated existing technologies within the DATALIFT platform: the Alignment API, for taking advantage of the EDOAL language and SILK, developed by Frei Universtität Berlin, for processing linking scripts. So far, we have generated SILK script from ontology alignments in order to produce links.

This work is part of the PhD of Zhengjie Fan, co-supervised with François Scharffe (LIRMM).

6.3. Ontology networks

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

6.3.1. Path queries and μ-calculus

Querying the semantic web is mainly done through the SPARQL language [18]. We designed one of its extensions, PSPARQL (Path SPARQL) which provides queries with paths of arbitrary length. We continue this work by connecting it to the work of the WAM team on static analysis of XPATH expressions. More specifically, we consdider query contrainment, i.e., determining whether, for any graph, the answers to a query are contained in those of another query. This is achieved by reducing this problem to satisfiability in the μ-calculus. In this work, RDF graphs are considered as transition systems and important fragments of RDFS and SPARQL as propositional μ-calculus formulas. It is then possible to use solvers of this logic to test query containment of SPARQL queries under RDFS and OWL schema constraints [11], with paths or under particular entailment regimes [10]. We have also implemented the proposed techniques and provided a first benchmark for query containment available under http://sparql-qc-bench.inrialpes.fr.

This work is part of the PhD of Melisachew Wudage Chekol [4], co-supervised with Nabil Layaïda (WAM).

GRAPHIK Project-Team

6. New Results

6.1. Ontology-Based Query Answering with Existential Rules

Participants: Jean-François Baget, Mélanie König, Michel Leclère, Marie-Laure Mugnier, Michaël Thomazo.

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

In collaboration with: Sebastian Rudolph (Karlsruhe Institute of Technology)

We have pursued the work on the existential rule framework in the context of Ontology-Based Query Answering. See the 2011 activity report for details on this framework also known as Datalog+/-. The ontology-based query answering problem consists of querying data while taking into account inferences enabled by an ontology (described by existential rules in our case).

From 2009 to 2011, we mainly investigated decidability and complexity issues. In 2012, while still interested in deepening decidability and complexity results, we tackled the next step: algorithms. Our aim is to develop algorithms with good theoretical properties (at least they should run in "the good worst-case complexity class") and with good performance in practice. There are two main ways of processing rules, namely forward chaining and backward chaining. In forward chaining, rules are applied to enrich the initial facts and query answering is solved by evaluating the query against the "saturated" facts (as in a classical database system). When it is finite, the backward chaining process can be divided into two steps: first, the query is rewritten into a first-order query (typically a union of conjunctive queries) using the rules; then the rewritten query is evaluated against the initial facts (again, as in a classical database system).

6.1.1. Forward Chaining Algorithms

Considering the expressive class of greedy bounded-treewidth set of rules (in short *gbts*), which we defined in 2011, we have designed a query answering algorithm which has several advantages over 2011 algorithm, while staying optimal with respect to worst-case combined and data complexities.

1. It is much more implementable (previous algorithm was using an oracle).
2. It is generic in the sense that it works for any class of rules that fulfills the gbts property, but it can also be easily specialized for specific gbts subclasses whith lower complexities, such as frontier-guarded or guarded rules, in such a way that it runs in the good complexity class.
3. It allows for separation between offline and online processing steps: the knowledge base can be compiled independently from queries, which are evaluated against the compiled form.

One of the lightweight description logics used for ontology-based query answering is $\mathcal{EL}$. We designed a subclass of existential rules that covers $\mathcal{EL}$ with the same complexity of reasoning, while allowing for any predicate arity and some cycles on variables. We also added complex role inclusions like transitivity and right/left identity rules to enhance expressivity, while staying polynomial in data complexity and generalizing existing results.

- *Results published in [36], [37] and [32] (invited conference). See also our research report [49] for a longer version.*
- *A journal version extending the papers at IJCAI 2011 and KR 2012 is in preparation, to be submitted to a major artificial intelligence journal.*

6.1.2. Backward Chaining Algorithms

We consider query rewriting techniques that output a union of conjunctive queries, which we see as a set of conjunctive queries. More specifically, only the most general elements of this set need to be kept in the output. We first proved that all sound and complete query rewriting algorithms necessarily produce the same result (up to redundancy) when restricted to their most general elements. It follows that comparing existing algorithms with respect to the size of the produced query is pointless.

Existing query rewriting algorithms accept only specific classes of existential rules (mainly corresponding to the translation of some lightweight description logics). We designed an algorithm that accept as input any set of existential rules and stops if this set of rules fufills so-called *fus* property (meaning that the set of most general rewritings of any initial conjunctive query is finite). This algorithm has been implemented and first experimentations have been led on rule bases obtained by translating description logic bases.

- *Results published in [31] (best paper price)*

6.1.3. Querying Optimization (Work in Progress)

Our current work aims at improving previous algorithms, in particular: the online querying step in the gbts algorithm; the query rewriting algorithm, by avoiding generating several times equivalent rewritings; for specific subclasses, query rewriting into a set of so-called semi-conjunctive queries instead of conjunctive queries, which reduces the size of the output query.

6.2. Reasoning with Imperfect Information and Priorities

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

In collaboration with: Joël Abecassis (IATE/INRA), Patrice Buche (IATE/INRA), Nir Oren (Univ. of Aberdeen, Scotland), Leon van der Torre (University of Luxembourg) and Nouredine Tamani (post-doc IATE).

This year, we mainly investigated decision support based on argumentations systems and preferences, either in relation with application needs in agronomy or on more fundamental aspects.

6.2.1. Argumentation for Decision Making in Agronomy

Historically, scientific investigations in this axis are guided by applications of our partners in agronomy (IATE laboratory). Part of the work has consisted of analyzing the proposed applications and the techniques they require in order to select appropriate applications with respect to our team project.

In the context of the EcoBioCap project (see Sect. 8.2), the different stakeholders have expressed conflicting preferences for the packaging quality. However, when discussing with domain experts they have raised the need for a tool which allows them to highlight a conflict and see the reasons behind it. In order to achieve this goal two steps were taken. First we have instantiated a popular logical argumentation framework (ASPIC+) with a simple preference logic. This allowed the different experts to express arguments about their preferences. We can then extract maximal consistent subsets of preferences by the means of extensions.

- *This work was performed in collaboration with the University of Aberdeen (Dr. Nir Oren) and the results were published and presented at the COMMA conference [24].*

Second, a negotiation phase was introduced to the previously described system in order for the domain experts to refine and extend their preferences. This tool was the aim of the master thesis of Patricio Mosse.

- *This work was published and presented at the Effost conference [23], based upon Patricio Mosse's Master Thesis [48]. A detailed journal article reporting on the two steps is under preparation and will be submitted beginning 2013.*

Let us mention additional results related to the applications in agronomy on semi-automatic data extraction from web data (tables) [39], [40], [41], data reliability, and the representation and flexible querying of imprecise data with fuzzy sets [42], [15]. These investigations are complementary to the above mentioned results on argumentation and generally relate to other aspects in the same applicative projects.

6.2.2. Conditional Acceptance Functions

Dung-style abstract argumentation theory centers on argumentation frameworks and acceptance functions. The latter take as input a framework and return sets of labelings. A labeling assigns "in", "out" or "undecided" to each arguments. Arguments having "in" assignment are acceptable arguments. This methodology however assumes full awareness of the arguments relevant to the evaluation. There are two reasons why this is not satisfactory. Firstly, full awareness is, in general, not a realistic assumption. Second, frameworks have explanatory power, which allows us to reason abductively or counterfactually, but this is lost under the usual

semantics. To recover this aspect, we generalized conventional acceptance, and we present the concept of a conditional acceptance function which copes with the dynamics of argumentation frameworks.

- *Results published in [28].*

6.2.3. Foundational Aspects of Preferences

Preferences are the backbone of various fields as they naturally arise and play an important role in many real-life decisions. Preferences are fundamental in scientific research frameworks as well as applications. One of the main problems an individual faces when expressing her preferences lies in the number of variables (or attributes or criteria) that she takes into account to evaluate the different outcomes. Indeed, the number of outcomes increases exponentially with the number of variables. Moreover, due to their cognitive limitation, individuals are generally not willing to compare all possible pairs of outcomes or evaluate them individually. These facts have an unfortunate consequence that any preference representation language that is based on the direct assessment of individual preferences over the complete set of outcomes is simply infeasible.
Fortunately, individuals can abstract their preferences. More specifically, instead of providing preferences over outcomes (by pairwise comparison or individual evaluation), they generally express preferences over partial descriptions of outcomes. Often such statements take the form of qualitative comparative preference statements e.g., "I like London more than Paris" and "prefer tea to coffee". Conditional logics aim at representing such partial descriptions of individual preferences which we refer to as comparative preference statements. They use different completion principles in order to compute a preference relation induced by a set of preference statements. In particular they use various more or less strong semantics to interpret comparative preference statements. So far the main objective in artificial intelligence has been to rank-order the set of outcomes given a set of comparative preference statements and one or several semantics. We addressed this problem from a different angle. We considered a set of postulates studied in preference logics and non-monotonic reasoning which formalize intuition one may have regarding the behavior of preference statements. We analyzed the behavior of the different semantics w.r.t. these postulates. Our analysis gives a complete picture of the behavior of our (five) semantics.

In the last decade, AI researchers have pointed out the existence of two types of information: positive information and negative information. This distinction has also been asserted in cognitive psychology. Distinguishing between these two types of information may be useful in both knowledge and preference representation. In the first case, one distinguishes between situations which are not impossible because they are not ruled out by the available knowledge, and what is possible for sure. In the second case, one distinguishes between what is not rejected and what is really desired. Besides it has been shown that possibility theory is a convenient tool to model and distinguish between these two types of information. Knowledge/Preference representation languages have also been extended to cope with this particular kind of information. Nevertheless despite solid theoretical advances in this topic, the crucial question of "which reading (negative or positive) one should have" remains a real bottleneck. We focused on comparative statements and presented a set of postulates describing different situations one may encounter. We provided a representation theorem describing which sets of postulates are satisfied by which kind of information (negative or positive) and conversely. One can then decide which reading to apply depending on which postulates she privileges.

- *Results published in [29] and [30].*

6.2.4. Argumentation for Inconsistency-Tolerant Query Answering (Work in Progress)

Argumentation allows to encode by the means of extensions maximal subsets of the knowledge base which are consistent (given the logic chosen). We are currently investigating the link between different argumentation extensions and the notion of a maximal repair as introduced by [51], [50] in the context of the positive existential subset of first order logic we are mainly working with. We are then interested in comparing the semantics proposed in the literature for query answering with inconsistent knowledge bases and argumentation reasoning paradigms. This study has been performed jointly with the University of Luxembourg during a research visit during end of November. We plan to submit our results at a conference beginning January.

6.3. Semantic Data Integration

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

It often happens that different references (i.e. data descriptions), possibly coming from heterogeneous data sources, concern the same real world entity. In such cases, it is necessary: (i) to detect whether different data descriptions really refer to the same real world entity and (ii) to fuse them into a unique representation. Since the seminal paper [52], this issue has been been studied under various names:"record linking", "entity resolution","reference resolution", "de-duplication", "object identification", "data reconciliation", etc., mostly in databases (cf. the bibliography by William E. Winckler [1]). It has become one of the major challenges in the Web of Data, where the objective is to link data published on the web and to process them as a single distributed database. Most entity resolution methods are based on classification techniques; Fatiha Saïs, Nathalie Pernelle and Marie-Christine Rousset proposed the first logical approach [53]. Many experiments on public data are underway, in France (cf. DataLift[2] and ISIDORE[3] projects) or internationally (e.g., VIAF project[4] led by OCLC[5], whose aim is to interconnect authority files coming from 18 national organizations).

Three years ago, we began a collaboration with ABES (National Bibliographic Agency for Universities, which takes part in the VIAF project). The aim of this collaboration is to enable the publication of ABES metadata based on the Web of Data and to provide an identification service dedicated to bibliographic notices. ABES bibliographic bases, and more generally document metadata bases, appear to be a privileged application domain for the representation and reasoning formalisms developed by the team. This work has an interdisciplinary dimension, as it also requires experts in the Library and Information Science domain. We think that a logical approach is able to provide a generic solution for entity resolution in document metedata bases, even though it is generally admitted in Library and Information Science that *"there is no single paradigmatic author name disambiguation task—each bibliographic database, each digital library, and each collection of publications, has its own unique set of problems and issues"* [54].

6.3.1. Implementation of an Entity Identification Service

Last year, we have developed a method and a prototype to perform entity resolution between on one hand the authors of a new bibliographic notice, and, on the other the domain experts of an authority catalog (and namely the Sudoc catalogue from the ABES agency). The prototype providing this service has been implemented on top of Cogui and experiments have been led in the context of the SudocAd project (jointly conducted by ABES and GraphIK). This work has been continued this year on the following issues as part of the Qualinca project:

- *generalizing the developed method with the aim to define a generic combined (numerical/logical) framework for entity resolution. This work is reported in the research report [44] that we plan to submit to a conference in January.*
- *Defining evaluation measures of the quality of an entity resolution tool. This work is still on-going.*

6.3.2. Quality of Document Catalogs

The SudocAd project showed the feasability and pertinence of a mixed approach for data interlinking problems. It also showed the immediate necessity of taking into account the existence of human errors already present in document catalogues. This led us to propose Qualinca, an ANR Contint project, accepted beginning 2012 and started in April 2012. The partners include two major actors in the document catalogues field: ABES and INA, as well as three academic research groups.

In this context we currently investigate a formal approach to the notion of a "key" in the web of data. Our immediate objective is to define the notion of a discovered key used then in order to evaluate the quality of data inter linking of a meta data catalogue.

[1] http://www.hcp.med.harvard.edu/statistics/survey-soft/docs/WinklerReclinkRef.pdf
[2] DataLift, http://datalift.org/
[3] ISIDORE, http://www.rechercheisidore.fr/
[4] The Virtual International Authority File, http://www.oclc.org/research/activities/viaf/
[5] Online Computer Library Center, http://www.oclc.org

We also study the methodology of linking error detection and fixing based on a partitioning (clustering) method on authors of bibliographic records. This study is part of the PhD thesis of LÃ©a Guizol (jointly funded by GraphIK and ABES). The above mentioned methodology is based on a set of criteria which will allow us to cluster "similar" authors together. Each criterion represents a point of view on the author: name, publication time span, publication domain etc. The first challenge consists of defining for each of such view points the respective criteria. The second challenge is to propose an aggregation semantics of such criteria which is well adapted for the problem at hand.

- *The methodology of using such clustering techniques for this problem has been published in [25]. A certain number of criteria have already been implemented and different partitioning semantics proposed. We are currently evaluating these on the ABES data.*

6.3.3. Multi Agent Knowledge Allocation

The assumption behind semantic data integration and querying is that different agents accessing the integrated data repository will have equal interest in the querying results. This is not always true in a data sensitive scenario where the knowledge provider might want to allocate the query answers to the agents based on their valuations. Furthermore, the agents might want some information exclusively (and thus offer a valuation that allows it) while others might want it shared. To this end we have proposed a new mechanism of allocation of query answers inspired from combinatorial auctions. We have defined the newly introduced scenario of Multi Agent Knowledge Allocation and proposed a graph based method, inspired on network flows, for solving it.

- *These results were published in [26] and [35]. We are currently investigating the mechanism design aspects of such valuations in collaboration with the University of Athens (Dr. Iannis Vetsikas).*

MAIA Project-Team

6. New Results

6.1. Decision Making

6.1.1. Accounting for Uncertainty in Penetration Testing

Participants: Olivier Buffet, Jörg Hoffmann.

Carlos Sarraute (Core Security Technologies) is an external collaborator.

Core Security Technologies is an U.S.-American/Argentinian company providing, amongst other things, tools for (semi-)automated security checking of computer networks against outside hacking attacks. For automation of such checks, a module is needed that automatically generates potential attack paths. Since the application domain is highly dynamic, a module allowing to declaratively specify the environment (the network and its configuration) is highly advantageous. For that reason, Core Security Technologies have been looking into using AI Planning techniques for this purpose. After consulting by Jörg Hoffmann, they are now using a variant of Jörg Hoffmann's FF planner in their product. While that solution is satisfactory in many respects, it also has weaknesses. The main weakness is that it does not handle the incomplete knowledge in this domain – figuratively speaking, the attacker is assumed to have perfect information about the network. This results in high costs in terms of runtime and network traffic, for extensive scanning activities prior to planning.

We are currently working with Core Security's research department to overcome this issue, by modeling and solving the attack planning problem as a POMDP instead. A workshop paper detailing the POMDP model has been published at SecArt'11. While such a model yields much higher quality attacks, solving an entire network as a POMDP is not feasible. We have designed a decomposition method making use of network structure and approximations to overcome this problem, by using the POMDP model only to find good-quality attacks on single machines, and propagating the results through the network in an appropriate manner. This work has been published in ICAPS'12 [34].

6.1.2. Searching for Information with MDPs

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

In the context of Mauricio Araya's PhD, we are working on how MDPs —or related models— can search for information. This has led to various research directions, such as extending POMDPs so as to optimize information-based rewards, or actively learning MDP models. This year, we have focused on a novel optimistic Bayesian Reinforcement Learning algorithm –as described below– and on Mauricio's dissertation.

Exact or approximate solutions to Model-based Bayesian RL are impractical, so that a number of heuristic approaches have been considered, most of them relying on the principle of "optimism in the face of uncertainty". Some of these algorithms have properties that guarantee the quality of their outcome, inspired by the PAC-learning (Probably Approximately Correct) framework. For example, some algorithms provably make in most cases the same decision as would be made if the true model were known (PAC-MDP property).

We have proposed a novel optimistic algorithm, BOLT, that is

- appealing in that it is (i) optimistic *about* the uncertainty in the model and (ii) deterministic (thus easier to study); and
- provably PAC-BAMDP, i.e., makes in most cases the same decision as a perfect BRL algorithm would.

This work has been published in ICML'12 [9] and (in French) in JFPDA'12 [30], additional details appearing in [40].

6.1.3. Scheduling for Probabilistic Realtime Systems

Participant: Olivier Buffet.

Maxim Dorin, Luca Santinelli, Liliana Cucu-Grosjean (Inria, TRIO team), and Rob Davies (U. of York) are external collaborators.

In this collaborative research work (mainly with the TRIO team), we look at the problem of scheduling periodic tasks on a single processor, in the case where each task's period is a (known) random variable. In this setting, some job will necessarily be missed, so that one will try to satisfy some criteria depending on the number of deadline misses.

We have proposed three criteria: (1) satisfying pre-defined deadline miss ratios, (2) minimizing the worst deadline miss ratio, and (3) minimizing the average deadline miss ratio. For each criterion we propose an algorithm that computes a provably optimal fixed priority assignment, i.e., a solution obtained by assigning priorities to tasks and executing jobs by order of priority.

This work has been presented in RTNS'11, and an extended version is currently in preparation.

6.1.4. Adaptive Management with POMDPs

Participant: Olivier Buffet.

Iadine Chadès, Josie Carwardine, Tara G. Martin (CSIRO), Samuel Nicol (U. of Alaska Fairbanks) and Régis Sabbadin (INRA) are external collaborators.

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

Here, we consider that a number of experts provide us with one possible model each, assuming that one of them is the true model. This allows making decisions by solving a *hidden model MDP* (hmMDP). An hmMDP is essentially a simplified mixed observability MDP (MOMDP), where the hidden part of the state corresponds to the model (in cases where all other variables are fully observable).

From a theoretical point of view, we have proved that deciding whether a finite-horizon hmMDP problem admits a solution policy of value greater than a pre-defined threshold is a PSPACE-complete problem. We have also conducted preliminary studies of this approach, using the scenario of the protection of the Gouldian finch, and focusing on the particular characteristics that could be exploited to more efficiently solve this problem. These results have been presented in AAAI'12 [14].

6.1.5. Multi-Camera Tracking in Partially Observable Environment

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

Fabien Flacher (Thales THERESIS) is an external collaborator.

In collaboration with Thales ThereSIS - SE&SIM Team (Synthetic Environment & Simulation), we focus on the problem of following the trajectories of several persons with the help of several actionable cameras. This problem is difficult since the set of cameras cannot cover simultaneously the whole environment, since some persons can be hidden by obstacles or by other persons, and since the behavior of each person is governed by internal variables which can only be inferred (such as his motivation or his hunger).

The approach we are working on is based on (1) POMDP formalisms to represent the state of the system (person and their internal states) and possible actions for the cameras, (2) a simulator provided and developed by Thales ThereSIS and (3) particle filtering approaches based on this simulator.

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

6.1.6. Scaling Up Decentralized MDPs Through Heuristic Search

Participant: Jilles Dibangoye.

External collaborators: Christopher Amato, Arnaud Doniec.

Decentralized partially observable Markov decision processes (Dec-POMDPs) are rich models for cooperative decision-making under uncertainty, but are often intractable to solve optimally (NEXP-complete). The transition and observation independent Dec-MDP is a general subclass that has been shown to have complexity in NP, but optimal algorithms for this subclass are still inefficient in practice. We first provide an updated proof that an optimal policy does not depend on the histories of the agents, but only the local observations. We then present a new algorithm based on heuristic search that is able to expand search nodes by using constraint optimization. We show experimental results comparing our approach with the state-of-the-art Dec-MDP and Dec-POMDP solvers. These results show a reduction in computation time and an increase in scalability by multiple orders of magnitude in a number of benchmarks.

This work was presented in UAI'2012 [16].

6.1.7. Approximate Modified Policy Iteration

Participant: Bruno Scherrer.

External collaborators: Victor Gabillon, Mohammad Ghavamzadeh and Matthieu Geist.

Modified policy iteration (MPI) is a dynamic programming (DP) algorithm that contains the two celebrated policy and value iteration methods. Despite its generality, MPI has not been thoroughly studied, especially its approximation form which is used when the state and/or action spaces are large or infinite. We have proposed three implementations of approximate MPI (AMPI) that are extensions of well-known approximate DP algorithms: fitted-value iteration, fitted-Q iteration, and classification-based policy iteration. We have provided an error propagation analysis that unifies those for approximate policy and value iteration. For the classification-based implementation, we have developed a finite-sample analysis that shows that MPI's main parameter allows to control the balance between the estimation error of the classifier and the overall value function approximation.

This work was presented in JFPDA'2012 [35] and ICML'2012 [45].

6.1.8. A Dantzig Selector Approach to Temporal Difference Learning

Participant: Bruno Scherrer.

External collaborators: Matthieu Geist, Mohammad Ghavamzadeh and Alessandro Lazaric.

LSTD is one of the most popular reinforcement learning algorithms for value function approximation. Whenever the number of samples is larger than the number of features, LSTD must be paired with some form of regularization. In particular, L_1-regularization methods tend to perform feature selection by promoting sparsity and thus they are particularly suited in high-dimensional problems. Nonetheless, since LSTD is not a simple regression algorithm but it solves a fixed-point problem, the integration with L_1-regularization is not straightforward and it might come with some drawbacks (see e.g., the P-matrix assumption for LASSO-TD). We have introduced a novel algorithm obtained by integrating LSTD with the Dantzig Selector. In particular, we have investigated the performance of the algorithm and its relationship with existing regularized approaches, showing how it overcomes some of the drawbacks of existing solutions.

This work was presented at JFPDA'2012 [33] and ICML'2012 [20].

6.1.9. On the Use of Non-Stationary Policies for Stationary Infinite-Horizon Markov Decision Processes

Participants: Bruno Scherrer, Boris Lesner.

In infinite-horizon stationary γ-discounted Markov Decision Processes, it is known that there exists a stationary optimal policy. Using Value and Policy Iteration with some error ϵ at each iteration, it is well-known that one can compute stationary policies that are $\frac{2\gamma}{(1-\gamma)^2}\epsilon$-optimal. After having shown that this guarantee is tight, we have developed variations of Value and Policy Iteration for computing non-stationary policies that can be up to $\frac{2\gamma}{1-\gamma}\epsilon$-optimal, which constitutes a significant improvement in the usual situation when γ is close to 1. Surprisingly, this shows that the problem of "computing near-optimal non-stationary policies" is much simpler than that of "computing near-optimal stationary policies".

This work was presented and selected for a full oral presentation at NIPS'2012 [28].

6.1.10. Developmental Reinforcement Learning

Participant: Alain Dutech.

External collaborators: Matthieu Geist (IMS Supelec), Olivier Pietquin (IMS Supelec)

Reinforcement Learning in rich, complex and large sensorimotor spaces is a difficult problem mainly because the exploration of such a huge space cannot be done in an extensive way. The idea is thus to adopt a developmental approach where the perception and motor skills of the robot can grow in richness and complexity during learning, as a consequence the size of the state and action spaces grows progressively when the performances of the learning agent increases. The learning framework relies on function approximators with specific properties (continuous input space, life-long adaptation, knowledge transfer). Architectures based on "reservoir learning" and "dynamical self-organizing maps" kind of artificial neural networks have been investigated [32], [18].

6.1.11. Dialog and POMDPs

Participant: Lucie Daubigney.

Reinforcement learning (RL) is now part of the state of the art in the domain of spoken dialog systems (SDS) optimization. The best performing RL methods, such as those based on Gaussian Processes, require to test small changes in the policy to assess them as improvements or degradations. This process is called on policy learning. Nevertheless, it can result in system behaviors that are not acceptable by users. Learning algorithms should ideally infer an optimal strategy by observing interactions generated by a non-optimal but acceptable strategy, that is learning off-policy. Such methods usually fail to scale up and are thus not suited for real-world systems. In this work, a sample-efficient, on-line and off-policy RL algorithm is proposed to learn an optimal policy [15]. This algorithm is combined to a compact non-linear value function representation (namely a multilayer perceptron) enabling to handle large scale systems. One of the application domain is the teaching of a second language [31].

6.1.12. SAP Speaks PDDL: Exploiting a Software-Engineering Model for Planning in Business Process Management

Participant: Jörg Hoffmann.

Ingo Weber (NICTA) and Frank Michael Kraft (bpmnforum.net) are external collaborators.

Planning is concerned with the automated solution of action sequencing problems described in declarative languages giving the action preconditions and effects. One important application area for such technology is the creation of new processes in Business Process Management (BPM), which is essential in an ever more dynamic business environment. A major obstacle for the application of Planning in this area lies in the modeling. Obtaining a suitable model to plan with – ideally a description in PDDL, the most commonly used planning language – is often prohibitively complicated and/or costly. Our core observation in this work is that this problem can be ameliorated by leveraging synergies with model-based software development. Our application at SAP, one of the leading vendors of enterprise software, demonstrates that even one-to-one model re-use is possible.

The model in question is called Status and Action Management (SAM). It describes the behavior of Business Objects (BO), i.e., large-scale data structures, at a level of abstraction corresponding to the language of business experts. SAM covers more than 400 kinds of BOs, each of which is described in terms of a set of status variables and how their values are required for, and affected by, processing steps (actions) that are atomic from a business perspective. SAM was developed by SAP as part of a major model-based software engineering effort. We show herein that one can use this same model for planning, thus obtaining a BPM planning application that incurs no modeling overhead at all.

We compile SAM into a variant of PDDL, and adapt an off-the-shelf planner to solve this kind of problem. Thanks to the resulting technology, business experts may create new processes simply by specifying the desired behavior in terms of status variable value changes: effectively, by describing the process in their own language.

This work has been published in JAIR [6].

6.1.13. Resource-Constrained Planning: A Monte Carlo Random Walk Approach

Participant: Jörg Hoffmann.

Hootan Nakhost and Martin Müller (University of Alberta) are external collaborators.

The need to economize limited resources, such as fuel or money, is a ubiquitous feature of planning problems. If the resources cannot be replenished, the planner must make do with the initial supply. It is then of paramount importance how constrained the problem is, i.e., whether and to which extent the initial resource supply exceeds the minimum need. While there is a large body of literature on numeric planning and planning with resources, such resource constrainedness has only been scantily investigated. We herein start to address this in more detail. We generalize the previous notion of resource constrainedness, characterized through a numeric problem feature $C \leq 1$, to the case of multiple resources. We implement an extended benchmark suite controlling C. We conduct a large-scale study of the current state of the art as a function of C, highlighting which techniques contribute to success. We introduce two new techniques on top of a recent Monte Carlo Random Walk method, resulting in a planner that, in these benchmarks, outperforms previous planners when resources are scarce (C close to 1). We investigate the parameters influencing the performance of that planner, and we show that one of the two new techniques works well also on the regular IPC benchmarks.

This work has been published in ICAPS-12 [26].

6.1.14. How to Relax a Bisimulation?

Participants: Michael Katz, Jörg Hoffmann.

Malte Helmert (Basel University) is an external collaborator.

Merge-and-shrink abstraction (M&S) is an approach for constructing admissible heuristic functions for cost-optimal planning. It enables the targeted design of abstractions, by allowing to choose individual pairs of (abstract) states to aggregate into one. A key question is how to actually make these choices, so as to obtain an informed heuristic at reasonable computational cost. Recent work has addressed this via the well-known notion of bisimulation. When aggregating only bisimilar states – essentially, states whose behavior is identical under every planning operator – M&S yields a perfect heuristic. However, bisimulations are typically exponentially large. Thus we must relax the bisimulation criterion, so that it applies to more state pairs, and yields smaller abstractions. We herein devise a fine-grained method for doing so. We restrict the bisimulation criterion to consider only a subset K of the planning operators. We show that, if K is chosen appropriately, then M&S still yields a perfect heuristic, while abstraction size may decrease exponentially. Designing practical approximations for K, we obtain M&S heuristics that are competitive with the state of the art.

This work has been published in ICAPS-12 [22], and as Inria research report RR-7901 [42].

6.1.15. Semi-Relaxed Plan Heuristics

Participants: Emil Keider, Jörg Hoffmann.

Patrik Haslum (ANU) is an external collaborator.

Heuristics based on the delete relaxation are at the forefront of modern domain-independent planning techniques. Here we introduce a principled and flexible technique for augmenting delete-relaxed tasks with a limited amount of delete information, by introducing special fluents that explicitly represent conjunctions of fluents in the original planning task. Differently from previous work in this direction, conditional effects are used to limit the growth of the task to be linear, rather than exponential, in the number of conjunctions that are introduced, making its use for obtaining heuristic functions feasible. We discuss how to obtain an informative set of conjunctions to be represented explicitly, and analyze and extend existing methods for relaxed planning in the presence of conditional effects. The resulting heuristics are empirically evaluated, and shown to be sometimes much more informative than standard delete-relaxation heuristics.

This work has been published in ICAPS-12 [24].

6.1.16. Structural Patterns Beyond Forks: Extending the Complexity Boundaries of Classical Planning

Participants: Michael Katz, Emil Keider.

Tractability analysis in terms of the causal graphs of planning problems has emerged as an important area of research in recent years, leading to new methods for the derivation of domain-independent heuristics (Katz and Domshlak 2010). Here we continue this work, extending our knowledge of the frontier between tractable and NP-complete fragments. We close some gaps left in previous work, and introduce novel causal graph fragments that we call the hourglass and semi-fork, for which under certain additional assumptions optimal planning is in P. We show that relaxing any one of the restrictions required for this tractability leads to NP-complete problems. Our results are of both theoretical and practical interest, as these fragments can be used in existing frameworks to derive new abstraction heuristics. Before they can be used, however, a number of practical issues must be addressed. We discuss these issues and propose some solutions.

This work has been published in AAAI-12 [23].

6.2. Understanding and mastering complex systems

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

Participants: Vincent Chevrier, Tomas Navarrete.

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

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

We propose a control architecture[1] based on an "equation-free" approach. We use a multi-agent model to evaluate the global impact of local control actions before applying the most pertinent set of actions.

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

6.2.2. Multi Modeling and multi-simulation

Participants: Vincent Chevrier, Christine Bourjot, Benjamin Camus.

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

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

We proposed the AA4MM meta-model [56] is to build a society of models, simulators and simulation softwares that solves the core challenges of multimodelling and simulation coupling in an homogeneous perspective.

This year we focused on systems that naturally involve entities at different levels of description: micro and macro levels with their dynamics and and their articulations : emergence (upward causation, from micro to macro levels) and immergence (downward causation, from macro to micro levels). We relied on Bourgine's generic view of the relationship between complex phenomenon's levels and their temporal evolution [50]. We proposed an extension of the AA4MM concepts[13] in order to adapt them to emergence and immergence specifications. A simple example of multi-level modeling of a flocking phenomenon has been implemented to illustrate our proposal.

6.2.3. Robustness of Cellular Automata and Reactive Multi-Agent Systems

Participants: Olivier Bouré, Vincent Chevrier, Nazim Fatès.

Our research on emergent collective behaviours focuses on robustness analysis, that is the behavioural resistance to perturbations in collective systems. We progressed in the knowledge of how to tackle this issue in the case of cellular automata (CA) and multi-agent systems (MAS).

The density classification problem was taken as a simple example for studying how decentralised computations can be carried out with simple cells. Although it is known that this problem can not be solved perfectly, we derived analytic calculations to understand how stochastic cellular automata provide good solutions [3]. A collaboration with mathematicians lead us to study how to extend this result to the infinite-space case [25] and to the 2D finite case [19].

Two papers resulting from the Amybia projects were published : experimental results on phase transitions obtained with FPGAs [7] and the description on a robotics experiment that demonstrates the robustness of a bio-inspired aggregation method [5].

The results on asynchronous information transmission in cellular automata were consolidated [2]. Original definitions of asynchronism were also developed in lattice-gas cellular automata [11], which allows us to complete our spectrum of models for which robustness can be studied analytically and with numerical simulations.

6.2.4. Robotics Systems and Ambiant Intelligence

6.2.4.1. Robotics systems : autonomy, cooperation, robustness

6.2.4.1.1. Local control based platooning

Participants: Alexis Scheuer, Olivier Simonin, François Charpillet, Jano Yazbeck.

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

In order to reduce the tracking error (*i.e.* the distance between each follower's path and the path of its predecessor), we developed both an innovative approach [58] and a new lateral control law. This lateral control law reduces the tracking error faster than other existing control laws. This control law, and the experimental results obtained with it, has been submitted to 2013 IEEE International Conference on Robotics and Automation. Its integration with a previously defined secure longitudinal control law [55] has also been studied, and will be submitted soon to 2013 IFAC Intelligent Autonomous Vehicles Symposium.

6.2.4.1.2. Adaptation of autonomous vehicle traffic to perturbations

Participants: Mohamed Tlig, Olivier Simonin, Olivier Buffet.

In the context of the European project InTraDE, the problem studied in the context of Mohamed Tlig's PhD thesis is to handle the displacements of numerous IAVs (Intelligent Autonomous Vehicles) in a seaport. Here we assume a supervisor planning the routes of the vehicles in the port. However, in such a large and complex system, different unexpected events can arise and degrade the traffic : failure of a vehicle, human mistake while driving, obstacle on roads, local re-planning, and so on.

We started focusing on a first important sub-problem of space resource sharing among multiple agents: how to ensure the crossing of two opposed flows of vehicles on a road when one of the two paths is blocked by an obstacle. To overcome this problem, blocked vehicles have to coordinate with vehicles of the other side to share the road and manage delays. The objective is to improve traffic flow and reduce the emergence of traffic jam. After formalizing this problem, we have defined and studied in simulation two decision rules that produce two different strategies: the first one alternates between two vehicles from each side of the road, and the second one gives priority to the vehicle with the highest delay. This work has been presented in ICTAI'12 [29].

We are now considering more complex situations, e.g., when multiple flows of vehicles share more than one crossroad.

6.2.4.1.3. Multi-robot exploration and mapping : The Carotte Challenge

Participants: Olivier Simonin, François Charpillet, Antoine Bautin, Nicolas Beaufort.

In the context of the ANR/DGA Carotte Challenge, we study since 2009 new strategies and algorithms for multi-robot exploration and mapping. The proposed models are experimented with real autonomous mobile robots at LORIA and every year at the Carotte challenge. Our consortium, called "Cart-o-matic", is composed of members from Université d'Angers (LISA) and from Maia team-project (our industrial partner has left the consortium in 2011).

The year 2012 produced several results :

- In June, **we won the final edition of the Carotte challenge !** This result was obtained in particular by the efficiency and the robustness of the multi-robot strategy we proposed. Our system also provided one of the best map of the contest.
- We developed a software platform, including SLAM, Planning and multi-robot explorations algorithms. These softwares have been protected by copyrights (APP), see 5.4 .
- We presented the results in different publications : RIA revue [8], ICIRA'2012 International Conference [10] (Finalist for the Best student paper).
- Antoine Bautin wrote his PhD thesis, that he will defend in the beginning of year 2013. This work proposes new frontier assignation algorithms for multi-robot exploration. We defined a new heuristics, based on counting the robots towards a frontier rather than considering only the distance between robots and frontiers. For these purpose we developed algorithms based on wavefronts computations (artificial potential fields).We measured on benchmarks that our algorithm outperforms the two classical approaches *closest frontier* and *Greedy assignation*.
- In Oct. 2012, Nassim Kaldé started a PhD thesis (MENRT scholarship), advised by F. Charpillet and O. Simonin. We aim at continuing the work of the Cartomatic project, under new hypothesis and constrains on communications and complexity of the environment to explore.

6.2.4.2. Intelligent environments and health assistance

6.2.4.2.1. Spatial computing: iTiles network

Participants: Olivier Simonin, François Charpillet, Lionel Havet, Mihai Andries.

Olivier Rochel (Inria research engineer, SED Nancy) is an external collaborator.

In the context of ambient intelligence and robotic assistance, we explore the definition of an active floor composed of connected nodes, forming a network of cells. We consider different way of computation, as spatial calculus, to define robust and self-adaptive functions in the environment. We aim at dealing with walk analysis, surveillance of people activity (actimetry) and assistance (control of assistant robots, etc.).

This work can be summarized in several points :

- We asked Hikob company to design the iTile model we defined at the end of year 2011. In 2012, a network of 90 iTiles has been installed on the floor of the smart apartment of the center. This apartment is an experimental platform developed in the context of the "Situated Computer Science" Action of the CPER MISN (Lorraine region, Inria and government fundings). See InfoSitu.
- Each iTile is composed of one node connected to embedded sensors and to its neighboring tiles. A tile holds 4 weight sensors, an accelerometer and 16 LEDs. A simulator of the iTile network has been developed by Olivier Rochel. This tools makes easier the development on the real tiles.
- Several functions have been developed and are currently under experiments: (i) detection of a person walking on the floor (ii) tracking of feet position (iii) propagation and display of information in the network.
- We are involved since 2010 in the PAL Inria large scale initiative (Personally Assisted Living). In this context, Mihai Andries started a PhD thesis in oct. 2012 (funded by Inria-PAL). This PhD. aims at studying the iTiles model and its possibility for assistance functions. We also study models allowing robots to interact and to use the iTile network.

6.2.4.2.2. Center of pressure and Step Detection of a person walking on our intelligent floor

Participants: Amandine Dubois, François Charpillet.

It is quite easy to estimate in realtime the center of pressure of a person walking on the intelligent floor described above. From a sequence of center of pressure, we conceived a system categorizing the set of measures into two sets :

- foot: the measure belongs to the pressure trace left by a foot on the floor,
- transition: the center of pressure corresponds to what happens when the person passes his right leg or left from backwards to forwards.

This has been done in a first time using an heuristic algorithm and then using an HMM. From this categorization it's then easy to estimate classical gait parameters such as length of the steps or speed of the walk.

6.2.4.2.3. Pose estimation of several kinects

Participants: Nicolas Beaufort, François Charpillet.

Tracking one or several persons using several Kinects required to solved the calibration, i.e estimation of the pose of each kinect in the scene, knowing that the area covered by each Depth camera don't overlap with other (because of interference). We have addressed this issue using a SLAM approach implemented within a GPU.

6.2.4.2.4. Fall prevention and Fall detection

Participants: Amandine Dubois, François Charpillet.

A major problem of public health is the loss of autonomy of elderly people usually caused by the falls. Since 2003 one of the goal of MAIA team is to develop a system allowing to detect falls and also to analyze the gait deterioration to prevent falls. A first approach consisted in developing a markerless human motion capture system estimating the 3D positions of the body joints over time. This system used a dynamic Bayesian network and a factored particle filtering algorithm. Since 2011, we used a new approach using Microsoft Kinect camera which allows to acquire at the same time a RGB and a depth image to deal of the problem of the gait. After the extraction of the human from the background, we calculate the gait parameters from the center of mass of a person. Some parameters, as the length and the time of steps, the speed of the gait, allow to predict a deterioration of the gait of a person and an increase of the risk of falls [17].

Another use of the extraction of center of mass of a person from the Kinect camera is to determine the activity of a person. The method uses a Hidden Markov Model to distinguish eight activities of the daily life (sitting, walking, lying (on a couch, on a bed), lying down, falling, going up on the obstacles, squatting and bending). We set up an experiment in a smart room to validate our results. Concerning the gait parameters we compare them to the real values obtained making the young subjects wake with pads soaked with ink under the shoes on the paper. The results show that there is a difference of 3-4cm between length provided by our Kinect algorithm and the real length provided by the paper. Concerning the detection of the activity, we ask to 28 subjects to perform eight situations (corresponding to the eight states of the HMM). The results showed that each situation is recognized exept “bending”, falls are detected correctly and there are no false positives except “sitting” and “qqsquatting” which are detected instead of “bending”.

MOSTRARE Project-Team

6. New Results

6.1. Modeling XML document transformations

Participants: Joachim Niehren, Angela Bonifati, Sophie Tison, Sławek Staworko, Aurélien Lemay, Anne-Cécile Caron, Yves Roos, Benoît Groz, Antoine Ndione, Tom Sebastian.

XML Schema Validation Groz, Staworko et. al. [26] present a new algorithm that tests determinism of regular expressions in linear time. All regular expressions used in DTDs and XML Schemas are required to be deterministic by the recommendation of the W3C. Whether this is the case can indeed been tested in linear time, as shown in this paper. The best known previous algorithm, which was based on the Glushkov automaton, required O(σ|e|) time, where σ is the number of distinct symbols in e. They also show that matching a word w against a deterministic regular expression e can be achieved in combined linear time O(|e| + |w|) for a wide range of cases.

Staworko et. al. studied bounded repairability for regular tree languages modulo the tree edit distance [28].

Ndione, Niehren, and Lemay [33] present a new probabilistic algorithm for approximate membership of words to regular languages modulo the edit distance on words. In the context of XML, this algorithm is relevant for sublinear DTD validity testing. The time complexity of the algorithm is independent of the size of the input word and polynomial in the size of the input automaton and the inverse error precision. All previous property testing algorithms for regular languages run in exponential time.

XML Query Answering Debarbieux, Niehren, Sebastian et. al. [32] present new algorithms for early XPath node selection on XML Streams. Early selection and rejection is crucial for efficiency, while earliest selection and rejection has high computational complexity in the general case. In contrast to all previous approaches, there algorithm does not rely on any expensive static analysis method. Instead, it is based on a compiler from XPath to nested word automata with selection and rejection states that they introduce. They cover a large fragment of downward XPath, with the main restriction that negation is forbidden above descendant axis and disjunctions. Non-determinism is used to deal with descendant axis and disjuctions. High run-time efficiency in practice is obtained by on-the-fly determinization for nested word automata, even in cases where static determinization produces automata of more than exponential size. Our experimental results confirm a very high efficiency in space and time. An implementation of our FXP/QuiXPath system is freely available and used for industrial transfer in the QuiXProc system.

Staworko et. al. tackled prioritized repairing and consistent query answering in relational databases in [20].

External Cooperations with other teams in Lille lead to the following publications [19], [31], [30].

6.2. Machine learning for XML document transformations

Participants: Adrien Boiret, Jean Decoster, Pascal Denis, Jean-Baptiste Faddoul, Antonino Freno, Gemma Garriga, Rémi Gilleron, Mikaela Keller, Grégoire Laurence, Aurélien Lemay, Joachim Niehren, Sławek Staworko, Marc Tommasi, Fabien Torre.

Learning XML Queries. Staworko et. al. [29] proposed learning twig and path queries.

Niehren, Champavère, Gilleron, and Lemay [34] propose new algorithm and learnability result for XML query induction based on schema-guided pruning strategies. Pruning strategies impose additional assumptions on node selection queries that are needed to compensate for small numbers of annotated examples. The class of regular queries that are stable under a given schema-guided pruning strategy was distinguished and shown to be learnable with polynomial time and data. The learning algorithm is obtained by adding pruning heuristics to the traditional learning algorithm for tree automata from positive and negative examples. While justified by a formal learning model, their learning algorithm for stable queries also performs very well in practice of XML information extraction.

Learning XML Transformations. Boiret, Lemay, and Niehren [21] solved the long open question of how to learn rational functions with polynomial time and data. Rational functions are transformations from words to words that can be defined by deterministic string transducers with lookahead. No previous learning results for classes of transducers with look-ahead existed, so this results is relevant for learning XML transformations defined by transducers with look-ahead, as with XSLT.

Multi-task Learning. We address the problem of multi-task learning with no label correspondence among tasks. In [22], Faddoul, Chidlovskii, Gilleron and Torre propose the multi-task Adaboost algorithm with Multi-Task Decision Trees as weak classifiers. They conduct experiments on multi-task datasets, including the Enron email set and Spam Filtering collection. Faddoul successfully defended his PhD thesis [16] in June 2012.

Probabilistic models for large graphs. We propose new approaches for the statistical analysis of large-scale undirected graphs. The guiding idea is to exploit the spectral decomposition of subgraph samples, and in particular their Fiedler eigenvalues, as basic features for density estimation and probabilistic inference. In [24], Freno, Keller, Garriga, and Tommasi develop a conditional random graph model for learning to predict links in information networks (such as scientific coauthorship and email communication). In [25], Freno, Keller, and Tommasi propose instead to estimate joint probability distributions through (non-linear) random fields, applying the resulting model to graph generation and link prediction.

Learning in Multiple graphs Ricatte, Garriga, Gilleron and Tommasi focus on learning from several sources of heterogeneous data. They represent each source as a graph of data and they propose to combine the multiple graphs with the help of small number of labeled nodes. They obtain a kernel that can be used as input to different graph-learning tasks such as node classification and clustering. The paper is under submission. Along a collaboration with physicians, Keller and Tommasi consider graphs that represents the structural connectivity of the brain (connectome). They develop a spatially constrained clustering method, combining heterogeous descriptions of the same objects through the graph of neighborhood on the cortex and the graph of connectivity. The paper is under submission.

Starting PhDs Boneva, Bonifati and Staworko started to supervise the PhD of R. Ciucanu on learning cross-model database mappings. Denis and Tommasi has begun to supervise the PhD of David Chatel on guided clustering for graphs (of texts).

OAK Team

6. New Results

6.1. Efficient XML and RDF data management

6.1.1. Efficient and safe management of XML and JSON data

We addressed the problem of detecting independence between XML queries and updates. Since the problem is undecidable for XQuery queries and updates, and is intractable even for restricted fragments, we adopted an approximating technique based on a schema-based static analysis. Our analysis turned to be precise and, at the same time, fast to run. Main result about this research line have been published in [6], while the complete study is reported in Federico Ulliana's PhD Thesis (defended in December 12) [5].

To address the problem of manipulating large XML documents via main-memory XQuery engines, largely used for their efficiency and easiness of integration in a programming environment, we developed partitioning techniques for both XQuery queries and updates. Our technique is based on a static analysis over queries and updates (no schema is used) able to infer information that is used to partition the input document, in a streaming fashion. Besides allowing existing main-memory system to scale up in terms of query/update input size, our technique also admits a MapReduce implementation. Main results have been published in [11], while the complete study is reported in Noor Malla's PhD Thesis (defended on September 21) [3].

We also tackled the problem of safe manipulation of JSON data. Some typed and MapReduce-based programming languages for manipulating JSON data have been recently proposed. However, the problem of inferring a schema for untyped JSON data was still open, and having a schema for manipulated data is fundamental for the afore mentioned programming languages. We started investigating technique able to deal with massive JSON data sets. To ensure efficiency, our technique is based on Map-Reduce, while to ensure precision and conciseness it adopts type rewriting rules able to: i) compact as much as possible intermediate inferred types, and ii) to avoid gross approximation when compacting types. Some preliminary results are quite encouraging, and appeared in [21].

6.1.2. Hybrid models for XML and RDF

Considerable energy is spent towards enriching XML data on the web with semantics through annotations. These annotations can range from simple metadata to complex semantic relationships between data items. Although the vision of supporting such annotations is spreading, it still lacks the infrastructure that will enable it. To this end we have proposed a framework enabling the storage and querying of annotated documents. We have introduced (i) the XR data model, in which annotated documents are XML documents described by RDF triples and (ii) the query language XRQ to interrogate annotated documents through their structure and their semantics. A prototype platform XRP for the management of annotated documents has also been developed, to show the relevance of our approach through experiments [9].

6.1.3. RDF query answering

A promising method for efficiently querying RDF data consists of translating SPARQL queries into efficient RDBMS-style operations. However, answering SPARQL queries requires handling *RDF reasoning*, which must be implemented outside the relational engines that do not support it. We have introduced the *database (DB) fragment of RDF*, going beyond the expressive power of previously studied RDF fragments. Within this fragment, we have devised novel *sound and complete techniques* for answering *Basic Graph Pattern (BGP)* queries, exploring the two established approaches for handling RDF semantics, namely reformulation and saturation. In particular, we have focused on handling database *updates* within each approach and proposed a method for incrementally maintaining the saturation; updates raise specific difficulties due to the rich RDF semantics. Our techniques have been designed to be deployed on top of any RDBMS(-style) engine, and we have experimentally studied their performance trade-offs [20], [14], [25].

6.1.4. Efficient and scalable Web Data Entity Resolution

We addressed the problem of detecting multiple heterogeneous representations of a real-world object (often referred to as record linkage, duplicate detection, or entity resolution) in two contexts, i.e., for hierarchical data and for data where relationships between entities form a graph.

Concerning XML entity resolution, we contributed to a novel algorithm that uses a Bayesian network to determine the probability of two XML elements being duplicates. The probability is based both on content and on structure information given by the hierarchical XML model. To efficiently evaluate the Bayesian network to find duplicates, we devised two pruning techniques. Whereas the first is lossless in terms of not loosing any true duplicates, the second pruning heuristic trades off runtime for a somewhat lower accuracy of the duplicate detection result. An experimental evaluation shows that the proposed solutions are capable of outperforming other state-of-the art XML duplicate detection methods [8].

As for duplicate detection in entity graphs, we defined a general framework for algorithms tackling this problem. The general process consists of three steps, namely retrieval, classification, and update. We further proposed an algorithm complying to the framework that leverages an off-the-shelf relational database to store and to efficiently query information (both data and relationships) relevant for duplicate classification. We further extended our framework and algorithm to allow for parallel and batched processing. Our experimental validation on data of up to two orders of magnitude larger than data considered by other state-of-the-art algorithms showed that the proposed methods allow to scale duplicate detection in entity graphs to large volumes of data [7].

6.1.5. Warehousing RDF data

Data warehousing (DW) research has lead to a set of tools and techniques for efficiently analyzing large amounts of multi-dimensional data. As more data gets produced and shared in RDF, analytic concepts and tools for analyzing such irregular, graph-shaped, semantic-rich data are needed. We have introduced *the first all-RDF model for warehousing RDF graphs*. Notably, we have defined *RDF analytical schemas*, themselves full RDF graphs, and *RDF analytical queries*, corresponding to the relational DW star/snowflake schemas and cubes. We have shown how *RDF OLAP operations* can be performed on our RDF cubes. We have also performed experiments validating the practical interest of our approach.

6.2. Cloud-based Data Management

We investigate architectures for storing Web data (in particular, XML documents and RDF graphs) based on commercial cloud platforms. In particular, we have developed the AMADA platform, which operates in a Software as a Service (SaaS) approach, allowing users to upload, index, store, and query large volumes of Web data. Since cloud users support monetary costs directly connected to their consumption of cloud resources, we focus on indexing content in the cloud. We study the applicability of several indexing strategies, and show that they lead not only to reducing query evaluation time, but also, importantly, to reducing the monetary costs associated with the exploitation of the cloud-based warehouse [10], [12], [13].

6.3. Data Transformation Management

When developing data transformations – a task omnipresent in applications like data integration, data migration, data cleaning, or scientific data processing – developers quickly face the need to verify the semantic correctness of the transformation. Declarative specifications of data transformations, e.g. SQL or ETL tools, increase developer productivity but usually provide limited or no means for inspection or debugging. In this situation, developers today have no choice but to manually analyze the transformation and, in case of an error, to (repeatedly) fix and test the transformation.

The above observations call for a more systematic management of a data transformation. Within Oak, we have so far focused on the first phase of the process described above, namely the analysis phase. Leveraging results obtained in previous years (by us and others), we solidified the theory of why-not provenance. Analogously to a distinction between different types of why-provenance, we defined three types of why-not provenance. For each of the three types, we surveyed the semantics employed by different approaches, e.g., set vs. bag semantics or existential vs. universal quantification. We also identified cases of implication and equivalence between why-not provenance of different types. We have leveraged this theoretical work during the design of a novel algorithm that has the potential to overcome usability and efficiency limitations of previous algorithms after further optimization, implementation, and validation in the future. Furthermore, we implemented different approaches for why-provenance and why-not provenance and included them in the Nautilus Analyzer, a system prototype for declarative query debugging. We demonstrated this prototype at CIKM 2012 [15].

ORPAILLEUR Project-Team

6. New Results

6.1. The Mining of Complex Data

Participants: Mehwish Alam, Thomas Bourquard, Aleksey Buzmakov, Victor Codocedo, Adrien Coulet, Elias Egho, Nicolas Jay, Florence Le Ber, Ioanna Lykourentzou, Luis Felipe Melo, Amedeo Napoli, Chedy Raïssi, My Thao Tang, Yannick Toussaint.

formal concept analysis, relational concept analysis, pattern structures, search for frequent itemsets, association rule extraction, mining of complex data, graph mining, skylines, sequence mining, FCA in spatial and temporal reasoning

Formal concept analysis, together with itemset search and association rule extraction, are suitable symbolic methods for KDDK, that may be used for real-sized applications. Global improvements may be carried on the scope of applicability, the ease of use, the efficiency of the methods, and on the ability to fit evolving situations. Accordingly, the team is working on extensions of such symbolic data mining methods to be applied on complex data such as biological or chemical data or textual documents, involving objects with multi-valued attributes (e.g. domains or intervals), n-ary relations, sequences, trees and graphs.

6.1.1. FCA, RCA, and Pattern Structures

Recent advances in data and knowledge engineering have emphasized the need for Formal Concept Analysis (FCA) tools taking into account structured data. There are a few extensions of FCA for handling contexts involving complex data formats, e.g. graphs or relational data. Among them, Relational Concept Analysis (RCA) is a process for analyzing objects described both by binary and relational attributes [116]. The RCA process takes as input a collection of contexts and of inter-context relations, and yields a set of lattices, one per context, whose concepts are linked by relations. RCA has an important role in KDDK, especially in text mining [86], [85].

Another extension of FCA is based on Pattern Structures (PS) [94], which allows to build a concept lattice from complex data, e.g. nominal, numerical, and interval data. In [101]), pattern structures are used for building a concept lattice from intervals, in full compliance with FCA, thus benefiting of the efficiency of FCA algorithms. Actually, the notion of similarity between objects is closely related to these extensions of FCA: two objects are similar as soon as they share the same attributes (binary case) or attributes with similar values or the same description (at least in part). Various results were obtained in the study of the relations existing between FCA with an embedded explicit similarity measure and FCA with pattern structures [100]. Moreover, similarity is not a transitive relation and this lead us to the study of tolerance relations. In addition, a new research perspective is aimed at using frequent itemset search methods for mining interval-based data being guided by pattern structures and biclustering as well.

6.1.2. Advances in FCA and Pattern Mining

In the context of environmental sciences, research work is in concern with the mining of complex hydroecological data with concept lattices. In particular, Florence Le Ber –as a member of UMR 7517 Lhyges, Strasbourg– is the scientific head of an ANR project named "FRESQUEAU" (2011–2014) dealing with FCA and data mining and hydroecological data (see http://engees.unistra.fr/site/recherche/projets/anr-fresqueau/).

In this framework, concept lattices based on multi-valued contexts have been used for characterizing macroinvertebrate communities in wetland and their seasonal evolution [19]. Within the ANR Fresqueau project we are studying tools for sequential pattern extraction taking into account spatial relations [56], [43].

From another point of view, miscanthus is a perennial crop used for biomass production. Its implantation is rather new, and there is few farms cultivating miscanthus in France. Understanding the farmers' choices for allocating miscanthus in their farmland is a main challenge. The CBR model is investigated for modeling these choices from farm surveys, including spatial reasoning aspects [20], [47] [41].

For completing the work on FCA and itemset search, there is still on-going work on frequent and rare itemset search, for being able to build lattices from very large data and completing the algorithm collection of the Coron platform. Work is still in progress on the design of an integrated and modular algorithm for searching for closed and generators itemsets, and equivalence classes of itemsets, thus enabling the construction of the associated lattice [121]. This research aspect is also linked to the research carried on within a the PICS CaDoE research project (see Section 8.1.1.3). In addition, there is also research work carried on different aspects involving the management of big data in the context of the BioIntelligence Project and the Quaero Project.

6.1.3. Skylines, sequential data, privacy and E-sports analytics

Pattern discovery is at the core of numerous data mining tasks. Although many methods focus on efficiency in pattern mining, they still suffer from the problem of choosing a threshold that influences the final extraction result. One goal is to make the results of pattern mining useful from a user-preference point of view. That is, take into account some domain knowledge to guide the pattern mining process. To this end, we integrate into the pattern discovery process the idea of skyline queries in order to mine *skyline patterns* in a threshold-free manner. This forms the basis for a novel approach to mining skyline patterns. The efficiency of our approach was illustrated over a use case from *chemoinformatics* and we showed that small sets of dominant patterns are produced under various measures that are interesting for chemical engineers and researchers.

Sequence data is widely used in many applications. Consequently, mining sequential patterns and other types of knowledge from sequence data has become an important data mining task. The main emphasis has been on developing efficient mining algorithms and effective pattern representation.

However, important fundamental problems still remained open: (i) given a sequence database, can we have an upper bound on the number of sequential patterns in the database? (ii) Is the efficiency of the sequence classifier only based on accuracy? (iii) Do the classifiers need the entire set of extracted patterns or a smaller set with the same expressiveness power?

In the field of the management of sequential date in medicine, analysis of health care trajectories led to the development of a new sequential pattern mining method [42]. The MMISP algorithm is able to efficiently extract sequential patterns composed of itemsets and multidimensional items. The multidimensional items can be described with additional taxonomic knowledge, allowing mining with appropriate levels of granularity. In parallel, a new measure has been created to compute the similarity between sequences of itemsets [78].

Orpailleur is one of the few project-teams working on privacy challenges which are becoming a core issue with different scientific problems in computer science. With technology infiltrating more and more every aspect of our lives, each human activity leaves a digital trace in some repository. Vast amounts of personal data are implicitly or explicitly created each day, and rarely one is aware of the extent of information that is kept, processed and analyzed without his knowledge or consent. These personal data give rise to significant concerns about user privacy, since important and sensitive details about private life are collected and exploited by third parties. The goal of privacy preservation technologies is to provide tools that allow greater control over the dissemination of user data. A promising trend in the field is Privacy Preserving Data Publishing (PPDP), which allows sharing of anonymized data. Anonymizing a dataset is not limited to the removal of direct identifiers that might exist in a dataset, e.g. the full name or the Social Security Number of a person. It also includes removing secondary information, e.g. like age, zip code that might lead indirectly to the true identity of an individual.

Existing research on this problem either perturbs the data, publishes them in disjoint groups disassociated from their sensitive labels, or generalizes their values by assuming the availability of a generalization hierarchy. In a recent work, we proposed a novel alternative [54]. Our publication method also puts data in a generalized

form, but does not require that published records form disjoint groups and does not assume a hierarchy either. Instead, it employs generalized bitmaps and recasts data values in a nonreciprocal manner.

One of the most fascinating challenges of our time is understanding the complexity of the global interconnected society we inhabit. Today we have the opportunity to observe and measure how our society intimately works, by analyzing the big data. i.e, the digital breadcrumbs of human activities sensed as a by-product of the ICT systems that we use. These data describe the daily human activities: for instance, automated payment systems record the tracks of our purchases, search engines record the logs of our queries for finding information on the web, social networking services record our connections to friends, colleagues and collaborators, wireless networks and mobile devices record the traces of our movements and our communications. These social data are at the heart of the idea of a knowledge society, where decisions can be taken on the basis of knowledge in these data.

Social network data analysis raises concerns about the privacy of related entities or individuals. We theoretically establish that any kind of structural identification attack can effectively be prevented using random edge perturbation and show that, surprisingly, important properties of the whole network, as well as of subgraphs thereof, can be accurately calculated and hence data analysis tasks performed on the perturbed data, given that the legitimate data recipient knows the perturbation probability as well [53].

"Electronic-sport" (E-Sport) is now established as a new entertainment genre. More and more players enjoy streaming their games, which attract even more viewers. In fact, in a recent social study, casual players were found to prefer watching professional gamers rather than playing the game themselves. Within this context, advertising provides a significant source of revenue to the professional players, the casters (displaying other people's games) and the game streaming platforms. In a recent work with Mehdi Kaytoue, we started focusing on the huge amount of data generated by electronic games. We crawled, during more than 100 days, the most popular among such specialized platforms: Twitch.tv.

Thanks to these gigabytes of data, we proposed a first characterization of a new Web community, and we showed, among other results, that the number of viewers of a streaming session evolves in a predictable way, that audience peaks of a game are explainable and that a Condorcet method can be used to sensibly rank the streamers by popularity [45]. This work should bring to light the study of E-Sport and its growing community for computer scientists and sociologists. They indeed deserve the attention of industrial partners (for the large amount of money involved) and researchers (for interesting problems in social network dynamics, personalized recommendation, sentiment analysis, etc.).

6.1.4. KDDK in Text Mining

Ontologies help software and human agents to communicate by providing shared and common domain knowledge, and by supporting various tasks, e.g. problem-solving and information retrieval. In practice, building an ontology depends on a number of "ontological resources" having different types: thesaurus, dictionaries, texts, databases, and ontologies themselves. We are currently working on the design of a methodology and the implementation of a system for ontology engineering from heterogeneous ontological resources. This methodology is based on both FCA and RCA, and was previously successfully applied in contexts such as astronomy and biology. At present, an engineer is implementing a robust system being guided by the previous research results and preparing the way for some new research directions involving trees and graphs (see also the work on the ANR Hybride project).

6.2. KDDK in Life Sciences

Participants: Yasmine Assess, Emmanuel Bresso, Thomas Bourquard, Adrien Coulet, Marie-Dominique Devignes, Anisah Ghoorah, Renaud Grisoni, Jean-François Kneib, Florence Le Ber, Bernard Maigret, Jean-François Mari, Amedeo Napoli, Violeta Pérez-Nueno, Dave Ritchie, Malika Smaïl-Tabbone.

The Life Sciences constitute a challenging domain in which to implement knowledge-guided approaches for knowledge discovery. Biological data are complex from many points of views: voluminous, high-dimensional, deeply inter-connected, etc. Analyzing such data and extracting hidden knowledge has become a crucial issue in important domains such as health, environment and agronomy. More and more bio-ontologies are available and can be used to enhance the knowledge discovery process [88], [117]. In the next few years, the experience of the Orpailleur team in KDDK applied to the Life Sciences will be further developed in two directions: the use of bio-ontologies to improve approaches for data integration and mining when applied to real-world data, and the study of the synergy between numeric and symbolic data-mining methods in life-science applications.

6.2.1. Relational data mining applied to complex biological object characterization and prediction

Inductive Logic Programming (ILP) is a learning method which allows expressive representation of the data and produces explicit first-order logic rules. However, any ILP system returns a single theory based on heuristic user-choices of various parameters and learning biases, thus ignoring potentially relevant rules. Accordingly, we propose an approach based on Formal Concept Analysis for effective interpretation of reached theories with the possibility of adding domain knowledge. Our approach was applied to the characterization of three-dimensional (3D) protein-binding sites, namely phosphorylation sites, which are the protein portions on which interactions with other proteins take place [33]. In this context, we defined a logical representation of 3D patches and formalized the problem as a concept learning problem using ILP. Another application of this KDDK methodology concerns the characterization and prediction of drug side-effect profiles (Journal manuscript in preparation). In this case, maximal frequent itemsets are extracted and allow us to propose relevant side-effect profiles of drugs which are further characterized by ILP.

6.2.2. Functional classification of genes using semantic similarity matrix and various clustering approaches

In the last report, we proposed a measure called IntelliGO which computes semantic similarity between genes for discovering biological functions shared by a set of genes (e.g., showing the same expression profile). This measure takes into account domain knowledge represented in Gene Ontology (GO) [83].

Functional classification aims at grouping genes according to their molecular function or the biological process they participate in. Evaluating the validity of such unsupervised gene classification remains a challenge given the variety of distance measures and classification algorithms that can be used. We evaluated functional classification of genes with the help of reference sets. Overlaps between clusters and reference sets are estimated by the F-score metric. We test the IntelliGO measure with hierarchical and fuzzy C-means clustering algorithms and we compare results with the state-of-the-art DAVID functional classification method (Database for Annotation Visualization and Integrated Discovery). Finally, study of best matching clusters to reference sets leads us to propose a method based on set-differences for discovering missing information.

The IntelliGO-based functional clustering method was tested on four benchmarking datasets consisting of biological pathways (KEGG database) and functional domains (Pfam database) [13]. The IntelliGO measure is usable on line (see http://bioinfo.loria.fr/Members/benabdsi/intelligo_project/).

We are currently investigating the clustering problem when objects are not represented as feature vectors in a vector space but as a pairwise similarity matrix. In biology such similarity measures are often computationally expensive or incompatible with *bona fide* distance definition. Embedding techniques of pairwise data into Euclidean space aim at facilitating subsequent clustering of the objects [115]. Spectral clustering methods are also relevant in this case [127]. We are conducting comparative and large-scale gene clustering evaluation using the Intelligo measure and reference sets.

6.2.3. Analysis of biomedical data annotated with ontologies

Annotating data with concepts of an ontology is a common practice in the biomedical domain. Resulting annotations define links between data and ontologies that are key for data exchange, data integration and data analysis tasks. In 2011 we collaborated with the National Center for Biomedical Ontologies (NCBO) to

develop of large repository of annotations named the NCBO Resource Index [99]. The resulting repository contains annotations of 34 biomedical databases annotated with concepts of 280 ontologies of the BioPortal [2]. We proposed a comparison of the annotations of a database of biomedical publications (Medline) with two databases of scientific funding (Crisp and ResearchCrossroads) to profile disease research [18]. The annotation of these three databases with a unique ontology about diseases enable to consider their content conjointly and consequently to analyze and compare, for distinct disease (or family of diseases), trends in term of number of publications and funding amounts.

We started a new project that aims at exploring biomedical annotations with FCA techniques. One main challenge here is to develop a knowledge discovery approach that consider the knowledge represented in the ontologies employed for the annotations.

6.2.4. Connecting textual biomedical knowledge with the Semantic Web

A large amount of biomedical knowledge is in the form of text embedded in published articles, clinical files or biomedical public databases. It is consequently of high interest to extract and structure this knowledge to facilitate its consideration when processing biomedical data. We benefited from advances in Natural Language Processing (NLP) techniques to extract fine-grained relationships mentioned in biomedical text and subsequently published such relationships on line in the form of RDF triples [91], [90]. In a collaborative work with the Health Care and Life Science (HCLS) interest group of the W3C, we demonstrated how biomedical knowledge extracted from text, along with Semantic Web technologies has high potential for recommendation systems and knowledge discovery in biomedicine [118].

6.3. Structural Systems Biology

Participants: Thomas Bourquard, Marie-Dominique Devignes, Anisah Ghoorah, Van-Thai Hoang, Bernard Maigret, Violeta Pérez-Nueno, Dave Ritchie, Malika Smaïl-Tabbone.

knowledge discovery in life sciences, bioinformatics, biology, chemistry, gene

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

6.3.1. Accelerating protein docking calculations using graphics processors

We have recently adapted the *Hex* protein docking software [113] to use modern graphics processors (GPUs) to carry out the expensive FFT part of a docking calculation [114]. Compared to using a single conventional central processor (CPU), a high-end GPU gives a speed-up of 45 or more. This software is publicly available at http://hex.loria.fr. A public GPU-powered server has also been created (http://hexserver.loria.fr) [105]. The docking server has performed some 12,000 docking runs during 2012. A book chapter describing the Hex docking algorithm has been published [75]. Our docking work has facilitated further developments on modeling the assembly of multi-component molecular structures using a particle swarm optimization technique [25], and on modeling protein flexibility during docking [24].

6.3.2. KBDOCK: Protein docking using Knowledge-Based approaches

In order to explore the possibilities of using structural knowledge of protein-protein interactions, Anisah Ghoorah recently developed the KBDOCK system as part of her doctoral thesis project. KBDOCK combines residue contact information from the 3DID database [119] with the Pfam protein domain family classification [92] together with coordinate data from the Protein Data Bank [87] in order to describe and analyze all known protein-protein interactions for which the 3D structures are available. We have demonstrated the utility of KBDOCK [96] for template-based docking using 73 complexes from the Protein Docking Benchmark [98]. KBDOCK is available at http://kbdock.loria.fr. Anisah Ghoorah successfully defended her thesis in November 2012 [10].

[2] http://bioportal.bioontology.org/

6.3.3. Kpax: A new algorithm for protein structure alignment

We have developed a new protein structure alignment approach called Kpax [6]. The approach exploits the fact that each amino acid residue has a carbon atom with a highly predictable tetrahedral geometry. This allows the local environment of each residue to be transformed into a canonical orientation, thus allowing easy comparison between the canonical orientations of residues within pairs of proteins using a novel scoring function based on Gaussian overlaps. The overall approach is two or three orders of magnitude faster than most contemporary protein structure alignment algorithms, while still being almost as accurate as the state-of-the-art TM-Align approach [126]. The Kpax program is available at http://kpax.loria.fr/.

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

Solving the structures of large protein assemblies is a difficult and computationally intensive task. Multiple two-dimensional (2D) images must be processed and classified to identify protein particles in different orientations. These images may then be averaged and stacked to deduce the three-dimensional (3D) structure of a protein. In order to help accelerate the first of these tasks we have recently developed a novel and highly parallel algorithm called "gEMpicker" which uses multiple graphics processors to detecting 2D particles in cryo-electron microscopy images. We have also developed a 3D shape matching algorithm called "gEMfitter" which also exploits graphics processors, and which will provide a useful tool for the final 3D assembly step. Both programs will soon be made publicly available, and two manuscripts describing our approach are in preparation.

6.3.5. DOVSA: Developing new algorithms for virtual screening

In 2010, Violeta Pérez-Nueno joined the Orpailleur team thanks to a Marie Curie Intra-European Fellowship (IEF) award to develop new virtual screening algorithms (DOVSA). The aim of this project is to advance the state of the art in computational virtual drug screening by developing a novel consensus shape clustering approach based on spherical harmonic (SH) shape representations [111]. The main disease target in this project is the acquired immune deficiency syndrome (AIDS), caused by the human immuno-deficiency virus (HIV) [109]. However, the approach will be quite generic and will be broadly applicable to many other diseases. Good progress has been made on calculating and clustering spherical harmonic "consensus shapes" which represent rather well the essential features of groups of active molecules [110]. The approach has since been extended to provide a rapid way to cluster drug families according to the Gaussian distributions of their surface shapes, and to predict possible cross-interactions of drug families [21]. We have also published a review on the state of the art in 3D virtual drug screening [15].

6.4. Around the Taaable research project

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

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

The Taaable project (http://taaable.fr) has been originally created as a challenger of the Computer Cooking Contest (ICCBR Conference). A candidate to this contest is a system whose goal is to solve cooking problems on the basis of a recipe book (common to all candidates), where each recipe is a shallow XML document with an important plain text part. The size of the recipe book (about 1500 recipes) prevents from a manual indexing of recipes: this indexing is performed using semi-automatic techniques.

Beyond its participation to the CCCs, the Taaable project aims at federating various research themes: case-based reasoning, information retrieval, knowledge acquisition and extraction, knowledge representation, minimal change theory, ontology engineering, semantic wikis, text-mining, etc. Case-based reasoning is used to perform adaptation of recipe to user constraints. The reasoning process uses a cooking domain ontology (especially hierarchies of classes) and adaptation rules. The knowledge base used by the inference engine is encoded within a semantic wiki, which contains the recipes, the domain ontology, and adaptation rules.

The most important original features of this version are:

Modules for computing adaptation knowledge. Using adaptation knowledge, and especially adaptation rules, is a way to better adapt cooking recipes to user constraints. A previous work for extracting adaptation rules has been performed in 2011 [93]. In this work, variation of ingredients between couple of recipes are mined using closed itemsets extraction. The adaptation rules come from the interpretation of closed itemsets whose items correspond to the ingredients that have to be removed, kept, or added. This approach has been integrated as a wiki extension, providing a collaborative environment in which humans and machines may now collaborate to better acquire adaptation rules [38]. Humans (expert in cooking) may trigger automatic processes (knowledge discovery processes) and and may validate, using a specific user interface, proposition of adaptation rules as adaptation knowledge, which is then added to the knowledge base. In the same way, this environment integrates also the results of a new work on knowledge extraction where specific cooking adaptation rules (i.e. that can be applied to a single recipe) are generalized using close itemsets into generic adaptation rules, to make them usable on other recipes [60].

A module for acquiring a process semantic representation. While a process for acquiring cases from recipe preparation texts exists, the results are not perfect. In order for valid case representations to be available in the semantic wiki, a semi-automatic case acquisition tool was created [59]. This tools presents the user with a graphical interface through which it is able to interact with the case acquisition process. In order to limit the effort required, each correction entered by the user is propagated by the tool to the rest of the case representation.

Some other theoretical studies have been carried out that should be applied to some future versions of Taaable:

- The combination of workflows and interval algebras to represent procedural knowledge [55].
- The revision-based adaptation of cases represented in a qualitative algebra [41].
- The study of taxonomy merging [39]: several versions of the taxonomies used in Taaable (such as the food hierarchy) can be incoherent one with the others and a merging process is defined in order to obtain a consistent merged taxonomy.
- A continuous knowledge extraction process to ensure the non regression of the reasoning system according to the ontology evolution [50].

SMIS Project-Team

6. New Results

6.1. Embedded Data Management

Participants: Nicolas Anciaux, Luc Bouganim, Lionel Le Folgoc, Yanli Guo, Saliha Lallali, Philippe Pucheral, Iulian Sandu Popa, Shaoyi Yin.

Inspired by low cost economic models, this work draws the idea of a one-dollar database machine, with the objective to disseminate databases everywhere, up to the lightest smart objects. In contrast to traditional database machines relying on massively parallel architectures, the one-dollar database machine considers the cheapest form of computer available today: a microcontroller equipped with GBs size (external) Flash storage. Designing such a database machine is very challenging due to a combination of conflicting RAM and NAND Flash constraints. To tackle this challenge, this work proposes a new paradigm based on database serialization (managing all database structures in a pure sequential way) and stratification (restructuring them into strata when a scalability limit is reached).We show that a complete DBMS engine can be designed according to this paradigm and demonstrate the effectiveness of the approach through a performance evaluation. This work capitalizes on previous results related to the indexing of Flash resident data [16] and has also obvious connections with the more general study we are conducting on Flash-based data management (see Section 6.2). Partial elements of this solution have been demonstrated at [13]. In 2012, we have extended our previous results on indexation of flash resident data [IS] and we have proposed the design of a complete DBMS engine [DAPD] complying by nature with the conflicting RAM and NAND Flash constraints we are facing. Currently, we work at the extension of the embedded DBMS engine to support document data (e.g., text documents or any type of documents that are tagged)) and spatio-temporal data (e.g., vehicle trajectory data or any type of time-stamped and/or geo-located data).

6.2. Flash-based Data Management

Participants: Matias Bjørling, Philippe Bonnet, Luc Bouganim, Niv Dayan.

Solid State Drives (SSDs) are replacing magnetic disks as secondary storage for database management, as they offer orders of magnitude improvement in terms of bandwidth and latency. In terms of system design, the advent of SSDs raises considerable challenges. First, the storage chips, which are the basic component of a SSD, have widely different characteristics – e.g., copy-on-write, erase-before-write and page-addressability for flash chips vs. in-place update and byte-addressability for PCM chips. Second, SSDs are no longer a bottleneck in terms of I/O latency forcing streamlined execution throughout the I/O stack. Finally, SSDs provide a high degree of parallelism that must be leveraged to reach nominal bandwidth. This evolution puts database system researchers at a crossroad. The first option is to hang on to the current architecture where secondary storage is encapsulated behind a block device interface. This is the mainstream option both in industry and academia. This leaves the storage and OS communities with the responsibility to deal with the complexity introduced by SSDs in the hope that they will provide us with a robust, yet simple, performance model. We showed that this option amounts to building on quicksand. We illustrated our point by debunking some popular myths about flash devices and by pointing out mistakes in the papers we have published throughout the years. The second option is to abandon the simple abstraction of the block device interface and reconsider how database storage managers, operating system drivers and SSD controllers interact. We gave our vision of how modern database systems should interact with secondary storage. This approach requires a deep re-design of the database system architecture, which is the only viable option for database system researchers to avoid becoming irrelevant. This work started at the end of 2011 and was published at CIDR'13 [20], in cooperation with the IT University of Copenhagen.

6.3. Minimal Exposure

Participants: Nicolas Anciaux, Walid Bezza, Danae Boutara, Benjamin Nguyen, Michalis Vazirgiannis.

When users request a service, the service provider usually asks for personal documents to tailor its service to the specific situation of the applicant. For example, the rate and duration of consumer's loans are usually adapted depending on the risk based on the income, assets or past lines of credits of the borrower. In practice, an excessive amount of personal data is collected and stored. Indeed, a paradox is at the root of this problem: service providers require users to expose data in order to determine whether that data is needed or not to achieve the purpose of the service. We explore a reverse approach, where service providers would publicly describe the data they require to complete their task, and where software (placed, depending on the context, on the client, on the server, or in a trusted hardware component) would use those descriptions to determine a minimum subset of information to expose. In 2012, we have presented our general framework called Minimum Exposure [14], we have modelled the underlying problem (for simple tasks) and proposed resolution algorithms [19], [24], and we have addressed the case of multi-label classifiers [18]. In the short term, we plan to adapt the minimum exposure architecture to support hidden decision rules using smart cards. Then, we will investigate new privacy metrics to capture the degree of exposure of sets of personal data items better.

6.4. Secure Global Computing on Asymmetric Architecture

Participants: Tristan Allard, Benjamin Nguyen, Philippe Pucheral, Quoc-Cuong To.

This research direction is based on the asymmetric architecture, composed of a powerful, available and untrusted computing infrastructure (server or cloud), and a large set of low powered, highly disconnected trusted devices. Trust is assumed ad hoc and can be justified by the use of secure tokens, open source software, friend relationships etc. In our work, we use tamper resistant secure tokens running trusted software, which provide a high degree of trust, due to the overwhelming cost of hardware tampering. The main difficulty on such an architecture is global processing i.e. constructing aggregate data from the individual records, because the entity in charge of executing the global computation is untrusted. Given our large scale data centric applications (e.g. nationwide surveys), we also discard solutions based on secure multi-party computation, which do not scale. We have studied the execution of Privacy Preserving Data Publishing (PPDP) algorithms on such an architecture, and provided generic protocols to deal with all kinds of PPDP algorithms, which are robust against honest-but-curious and malicious adversaries. This work is an extension of [31] We are now studying more generally the execution of SQL "Group by" queries on this architecture, which is the topic of Quoc-Cuong To's Ph.D. thesis started in sept. 2012. We have published preliminary results on this novel problem in [23], which adapts the techniques proposed in [31].

6.5. Trusted Cell Data Management

Participants: Nicolas Anciaux, Philippe Bonnet, Luc Bouganim, Benjamin Nguyen, Philippe Pucheral, Iulian Sandu Popa.

With the convergence of mobile communications, sensors and online social networks technologies, we are witnessing an exponential increase in the creation and consumption of personal data. Such data is volunteered by users, automatically captured by sensors or inferred from existing data. Today, there is a wide consensus that individuals should have increased control on how their personal data is collected, managed and shared. Yet there is no appropriate technical solution to implement such personal data services: centralized solutions sacrifice security for innovative applications, while decentralized solutions sacrifice innovative applications for security. In this work, we argue that the advent of secure hardware in all personal IT devices, at the edges of the Internet, could trigger a sea change. We propose the vision of trusted cells: personal data servers running on secure smart phones, set-top boxes, secure portable tokens or smart cards to form a global, decentralized data platform that provides security yet enables innovative applications. We motivate our approach, describe the trusted cells architecture and define a range of challenges for future research in a paper published at CIDR'13 (Int. Conf on Innovative Data Systems Research) [17].

6.6. Experiment in the medical field

Participants: Nicolas Anciaux, Luc Bouganim, Philippe Pucheral, Alexei Troussov.

The PlugDB engine is being experimented in the field since September 2011 to implement a secure and portable medical-social folder. The objective is to improve the coordination of medical care and social services provided at home for dependent people. Details related to this experiment are given in Section 7.2 . While this action did not generate new academic results (though it helped us validating some previous results), it imposed us a strong investment in terms of test and optimization for our prototype and in terms of communication to promote this experiment at the regional level.

WAM Project-Team

6. New Results

6.1. Multimedia Models and Formats

In the context of the CLAIRE project (see section 7.1.1), a new model for educational documents has been defined. The objectives of this model are:

- to seamlessly handle conventional and richmedia content in the context of a unique pedagogical web platform.
- to be able to store and recover any multimedia document including its spatial and time structure, consistent with HTML5 and Timesheets specifications.
- to have a data model which is format agnostic to cope with existing and future rendering systems.
- to cope with the authoring needs of all users.

We have more specifically worked on the multimedia modelling part for defining spatial and temporal fragment types. These types are used to express the synchronization beween different elements within the document.

We are now using this model in the definition and implementation of a web-based authoring user interface.

6.2. XML Processing

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

- We have introduced the first system capable of statically verifying properties of a given cascading style sheet (CSS) over the whole set of documents to which this style sheet applies [5]. Properties include coverage of styling information and absence of erroneous rendering.
- In a joint work with the EXMO team, we have introduced a novel approach for deciding the SPARQL query containment problem in the presence of schemas, that paves the way for future extensions [4] [3] [8] [1].
- We have revisited the problem of XML Query-Update Independence Analysis, and showed the relevance of an approach that has been neglected in the literature so far [6]. In particular, we have compared an SMT-modulo with a tree logic approach to Independence Analysis.
- We have made progress on the characterization of the impacts of schema changes on XQuery programs [7].
- We have formally proved a result about the factorization power of the Lean: a construction that we use to speed up the XML Reasoning Solver. We have characterized which kind of duplicate subformulas this construction eliminates, and how [10].
- We have proposed a novel technique and a tool for the static type-checking of XQuery programs, using backward type inference [11].
- We have defined a type system for integrating session types for objects in object-oriented languages such as Java, with full structural subtyping, without altering the language semantics [9]. Session types are protocol specifications which describe which sequences of method calls are allowed or disallowed on a given object.

We briefly review these results below.

6.2.1. Automated Analysis of Cascading Style Sheets (CSS)

Developing and maintaining cascading style sheets (CSS) is an important issue to web developers as they suffer from the lack of rigorous methods. Most existing means rely on validators that check syntactic rules, and on runtime debuggers that check the behavior of a CSS style sheet on a particular document instance. However, the aim of most style sheets is to be applied to an entire set of documents, usually defined by some schema. To this end, a CSS style sheet is usually written w.r.t. a given schema. While usual debugging tools help reducing the number of bugs, they do not ultimately allow to prove properties over the whole set of documents to which the style sheet is intended to be applied. We have developed a novel approach to fill this lack [5]. The main ideas are borrowed from the fields of logic and compile-time verification and applied to the analysis of CSS style sheets. We have implemented an original tool (see section 5.1.1) based on recent advances in tree logics. The tool is capable of statically detecting a wide range of errors (such as empty CSS selectors and semantically equivalent selectors), as well as proving properties related to sets of documents (such as coverage of styling information), in the presence or absence of schema information. This new tool can be used in addition to existing runtime debuggers to ensure a higher level of quality of CSS style sheets.

6.2.2. Deciding Satisfiability and Containment for Semantic Web Queries

The problem of SPARQL query containment is defined as determining if the result of one query is included in the result of another for any RDF graph. Query containment is important in many areas, including information integration, query optimization, and reasoning about Entity-Relationship diagrams [1].

We encode this problem into an expressive logic called μ-calculus: where RDF graphs become transition systems, queries and schema axioms become formulas [4] [3]. Thus, the containment problem is reduced to formula satisfiability test. Beyond the logic's expressive power, satisfiability solvers are available for it. Hence, this study allows to exploit these advantages.

In addition, in order to experimentally assess implementation limitations, we have designed a benchmark suite offering different experimental settings depending on the type of queries, projection and reasoning (RDFS) [8]. We have applied this benchmark to three available systems using different techniques highlighting the strengths and weaknesses of such systems.

6.2.3. XML Query-Update Independence Analysis Revisited

XML transformations can be resource-costly in particular when applied to very large XML documents and document sets. Those transformations usually involve lots of XPath queries and may not need to be entirely re-executed following an update of the input document. In this context, a given query is said to be independent of a given update if, for any XML document, the results of the query are not affected by the update. We have revisited Benedikt and Cheney's framework for query-update independence analysis and we have shown that performance can be drastically enhanced, contradicting their initial claims [6]. The essence of our approach and results resides in the use of an appropriate logic, to which queries and updates are both succinctly translated. Compared to previous approaches, ours is more expressive from a theoretical point of view, equally accurate, and more efficient in practice. We have illustrated this through practical experiments and comparative figures.

6.2.4. Toward Automated Schema-directed Code Revision

Updating XQuery programs in accordance with a change of the input XML schema is known to be a time-consuming and error-prone task. We have designed an automatic method aimed at helping developers realign the XQuery program with the new schema [7]. First, we have devised a taxonomy of possible problems induced by a schema change. This allows to differentiate problems according to their severity levels, e.g. errors that require code revision, and semantic changes that should be brought to the developer's attention. Second, we have provided the necessary algorithms to detect such problems using our solver (see section 5.1) to check satisfiability of XPath expressions.

6.2.5. Logical Combinators for Rich Type Systems

We have developed a functional approach to design rich type systems based on an elegant logical representation of types [10]. The representation is not only clean but it also avoids exponential increases in combined complexity due to subformula duplication. This opens the way to solving a wide range of problems such as subtyping in exponential-time even though their direct translation into the underlying logic results in an exponential blowup of the formula size, yielding an incorrectly presumed two-exponential time complexity.

6.2.6. Backward type inference for XQuery

We have designed a novel technique and a tool for static type-checking of XQuery programs [11]. The tool looks for errors in the program by jointly analyzing the source code of the program, input and output schemas that respectively describe the sets of documents admissible as input and as output of the program. The crux and the novelty of our results reside in the joint use of backward type inference and a two-way logic to represent inferred tree type portions. This allowed us to design and implement a type-checker for XQuery which is more precise and supports a larger fragment of XQuery compared to the approaches previously proposed in the literature; in particular compared to the only few actually implemented static type-checkers such as the one in Galax. The whole system uses compilers and a satisfiability solver for deciding containment for two-way regular tree expressions. Our tool takes an XQuery program and two schemas S_{in} and S_{out} as input. If the program is found incorrect, then it automatically generates a counter-example valid w.r.t. S_{in} and such that the program produces an invalid output w.r.t S_{out}. This counter-example can be used by the programmer to fix the program.

6.2.7. Session types

Session types allow communication protocols to be specified type-theoretically so that protocol implementations can be verified by static type checking. In [9], we extend previous work on session types for distributed object-oriented languages in three ways. (1) We attach a session type to a class definition, to specify the possible sequences of method calls. (2) We allow a session type (protocol) implementation to be *modularized*, i.e. partitioned into separately-callable methods. (3) We treat session-typed communication channels as objects, integrating their session types with the session types of classes. The result is an elegant unification of communication channels and their session types, distributed object-oriented programming, and a form of typestate supporting non-uniform objects, i.e. objects that dynamically change the set of available methods. We define syntax, operational semantics, a sound type system, and a sound and complete type checking algorithm for a small distributed class-based object-oriented language with structural subtyping. Static typing guarantees that both sequences of messages on channels, and sequences of method calls on objects, conform to type-theoretic specifications, thus ensuring type-safety. The language includes expected features of session types, such as delegation, and expected features of object-oriented programming, such as encapsulation of local state. The main ideas have been implemented as a prototype, extending Java 1.4.

6.3. Multimedia Authoring

In cooperation with EPFL (Lausanne) we pursue our research on template-driven editing for XML multimedia contents (see section 3.3.2). Experiments with very different types of contents have been done with the AXEL library. AXEL is developed by EPFL, based on our joint work on template languages. It is an innovative multi-purpose client-side authoring framework intended for web users with limited skills.

We have addressed the issue of authoring XML multimedia content on the web, focusing on methods that apply to such contents as structured documents, factual data, and multimedia objects [2]. We have shown that a template-based approach enhances the ability for multiple applications to use the produced content.

6.4. Augmented Environments

Most results in the area of augmented environments were presented through various software products and prototypes, including:

- IXE, Interactive eXtensible Engine (see section 5.5 for details). In particular, IXE allowed us to show that a precision of one step is attainable, guidance being done through a mix of spatialized vocal instructions and 3D audio.
- GIF Demonstrator: This application was used to showcase our technologies at the Grenoble Innovation Fair (GIF). Augmented reality was used to find the various booths and products, while our indoor navigation system was guiding visitors to any booth.
- Interactive Audio Panorama: A fun interactive experience with virtual audio. It immerses the user in a complete 360° audio panorama and allows her/him to discover a futuristic house. It demonstrates the authoring possibilities offered by the MAUDL interactive audio language.
- PDRTrack: An indoor localization utility demonstrating the various correction parameters of our IMU-based localization system. The user can record data sets and simulate using various parameters to find out the effect of different map matching settings and their result on localization accuracy. The user can also simply walk in real-time with tracking enabled on a given OpenStreetMap network.
- Sugimotocho Stn: A model of this railway station has been built with the help of the GISLab (Osaka City University). An electronic kick-scooter was used to measure distances and a navigation network was designed to help people to move around in the station.

These products and prototypes were presented in various fora in 2012, in particular at:

- Grenoble Innovation Fair
- 4I Forum
- 6th European eAccessibility Forum
- State Of The Map 2012

WIMMICS Team

6. New Results

6.1. Linked Data Access

Participants: Serena Villata, Luca Costabello, Fabien Gandon.

We designed and developed a context-aware access control framework for the Web of Data called Shi3ld [7]. The framework protects access to SPARQL endpoints, and it adopts Semantic Web languages only, as in the philosophy of the Web of Data. The innovative feature of the proposed framework consists in evaluating the accessibility to the data considering the attributes of the users. These attributes are defined following three main dimensions: user, device, and environment. The evaluation of the model shows that access control comes with a cost but it guarantees the protection of the data published on the Web of Data. The results of this research activity have been published in international conferences in the area of Artificial Intelligence (ECAI, [35]) and the Web (WWW [61], HT [34]).

On the same line, we have proposed a framework for attaching the licenses to the data resulting from a query on the Web of Data. The rationale is that the licenses associated to the data returned by the query are selected, and using a number of rules their compatibility is assessed. If the licenses are evaluated compatible, then they are composed into a composite license which is released to the user together with the data. The results of this research have been published in the COLD international workshop [74] and in the ISWC international conference [75] (best poster award). These two research lines have been performed in the context of the DataLift ANR project.

The PhD thesis of Luca Costabello, directed by F. Gandon and I. Herman (CWI and Semantic Web Activity Lead at W3C) investigates Web of Data interaction from mobile environments. Two main research activities have been carried out in 2012: i) PRISSMA [8], an adaptive rendering engine for RDF and ii) Shi3ld[9], a context-aware access control framework for Linked Data.

The goal of PRISSMA is delivering an *adaptive* rendering engine for Linked Data resources. PRISSMA tweaks RDF visualization to the context in which the resource consumption is performed. Work in 2012 has been mainly focused on determining the algorithm that selects the best RDF visualization according to the real, sensed context. The uncertain and incomplete nature of context data, led to investigate strategies that model the task as an inexact RDF instance matching operation.

The second main research line carried out in 2012 led to the creation of Shi3ld, an access control framework for Linked Data SPARQL endpoints. Shi3ld authorization procedure and Access Policies, defined using Semantic Web languages only, have been enriched with the notion of mobile context, thus enabling context-based access control (e.g. geo-temporal authorization policies).

A collateral research line has been carried out as the follow-up of the 9th edition of the Summer School on Ontology Engineering and the Semantic Web [10]. We contributed to an exploration work on the problem of spamming in Linked Data, providing a classification of potential spamming techniques and populating and publishing a dataset containing spammed triples[11]. The dataset is useful to train anti-spamming mechanisms.

6.2. ISICIL Platform

Participants: Nicolas Delaforge, Michel Buffa, Fabien Gandon, Alain Giboin.

[7] http://wimmics.inria.fr/projects/shi3ld/
[8] http://wimmics.inria.fr/projects/prissma
[9] http://wimmics.inria.fr/projects/shi3ld
[10] http://sssw.org/2012/
[11] http://www-sop.inria.fr/members/Luca.Costabello/spam/

ISICIL is an ANR project studying social networks and Semantic Web communities to support corporate intelligence.

First, ISICIL proposes a multidisciplinary design of a new form of corporate intelligence. The challenge of this project is to reconcile the new viral Web applications with formal representations of business processes and to integrate them into practical intelligence communities of actors working in the company. We designed, studied and experimented with new tools to support collaborative tasks in business intelligence by leveraging Web 2.0 interfaces (blog, wiki, social bookmarking) for interactions and Semantic Web technologies for interoperability and information processing. ISICIL also allowed to explore new scientific developments of the notion of epistemic cooperation (human interaction oriented toward the development and transmission of knowledge) and to identify usable technological solutions. An ergonomic approach, combining impregnation of ground truth data and freer technological inspirations from bibliographic and webographic sources, was proposed.

Secondly, ISICIL uses typed graphs as models underlying epistemic communities. The entire model relies on a unifying model based on RDF graphs to represent resources and community stakeholders. These models are integrated with bookmarking tools or "Web scraping" the results of which are tagged. The tags used are collected to form folksonomies and a semi-automatic thesaurus structure in these folksonomies. User feedback on this structure is captured when they use the search engine which offers tags related to their keywords and the user can accept, reject or adjust these suggestions. User profiles and links between them, considered as a network, are processed by a series of operators to propose a semantic analysis of social network, for example the computation of indicators of centrality parameterized by an ontology. Merged graphs of structured folksonomies and of social networks finally allow the detection and labeling of epistemic communities. Meanwhile we study how the analysis of user interactions can determine the trust and how to represent and control access to data and their semantics in a social network.

As part of the paradigm of Social Epistemology, ISICIL combined in an ergonomic approach, impregnation of ground truth data, a bibliometric study and technological inspirations to offer patterns and inferences exploiting the Semantic Web social networks to assist corporate intelligence. An open-source platform is available under CeCILL-C licence and was tested at ADEME and Orange. ISICIL is a proof of concept of the compatibility of the Semantic Web formalisms, practices and models of Web 2.0 and the philosophical framework of social epistemology.

The project has resulted in three PhD theses (Florie Bugeaud 2011, Guillaume Erétéo 2011 [98], Freddy Limpens 2010 [99]) and publications at the following conferences: ISWC 2009 IEEE / WIC / ACM Web Intelligence 2011, Web Science 2010, WWW 2011, ASWC 2009, COOP 2010, PRO-VE 2009, VSST 2010, EGC 2010, IC 2009 & 2011, Psycho Ergo Days 2010.

The ISICIL project participated to Colloque ANR STIC, January 4-6 2012.

6.3. Natural Language Processing

Participants: Elena Cabrio, Julien Cojan, Fabien Gandon, Maxime Lefrançois, Serena Villata.

We have proposed a combined framework of natural language processing and argumentation theory to support the users in their interactions within online debate platforms. The framework combines a natural language processing module which exploits the textual entailment approach and detects the arguments in natural language debates and the relationships among them, and an argumentation module which represents the debates as graphs and detects the accepted arguments. The results of this research have been published in one of the major conferences in the field of Artificial Intelligence (ECAI [30]), and in the areas of natural language processing (ACL [28]) and argumentation theory (COMMA [29]).

To enhance users interactions with the Web of Data, query interfaces providing an extensible mapping between natural language expressions, and concepts and relations in structured knowledge bases are becoming particularly relevant. As a result of the first year of her postdoc, Elena Cabrio (together with Julien Cojan) designed QAKiS (Question Answering wiKiframework-based System), that allows end users to submit a query in English to an RDF triple store and obtain the answer in the same language, hiding the complexity of the

non intuitive formal query languages involved in the resolution process. At the same time, the expressiveness of these standards is exploited to scale to the huge amounts of available semantic data.

In its current implementation, QAKiS addresses the task of Question Answering (QA) over structured knowledge bases (e.g. DBpedia) where relevant information is expressed also in unstructured form (e.g. Wikipedia pages). Its major novelty is to implement a relation-based match for question interpretation, to convert the user question into a query language (e.g. SPARQL). A demo of the system is available online [12]. The results of this research have been published as a demo paper in the main conference of Semantic Web, ISWC [57], and in the special issue of the journal Intelligenza Artificiale [14]. QAKiS has been evaluated with respect to state-of-the-art systems taking part into the QALD-2 (Question Answering over Linked Data) challenge at ESWC, obtaining satisfactory results [59].

In the PhD Thesis of Maxime Lefrançois, we are interested in bridging the world of natural language and the world of the Semantic Web in particular to support multilingual access to the Web of Data, and management of interlingual knowledge bases. In 2011 we introduced the ULiS project that aimed at designing a pivot-based NLP technique called Universal Linguistic System, 100% using the Semantic Web formalisms, and being compliant with the Meaning-Text theory [102].

We showed that neither Description Logics nor Conceptual Graphs suit our needs, so this Ph.D. now focuses on the formalization of the Unit Graphs mathematical framework that is conceived to fill the gap between the highly linguistically precise Explanatory Combinatorial Dictionaries of the Meaning-Text Theory and the Dependency Grammars, and the highly mathematically grounded model of the Conceptual Graphs.

Maxime finally joined the Multilingual-Web-LT W3C Working Group, and left it when the charter got revised.

6.4. Requirement Engineering

Participants: Zeina Azmeh, Isabelle Mirbel, Serena Villata.

Requirements engineering is an essential process of software engineering, during which the complete behavior of a software system can be defined. The success of this process plays a crucial role in the success of the whole software project. A key issue of requirements engineering is stakeholders participation, which is facilitated through the emergence of online collaborative working tools. These tools create new opportunities of practice regarding requirements elicitation. Nevertheless, they result in an information overload lacking structure and semantics. Consequently, requirements analysis and selection become more challenging.

Our current proposition is embodied in an approach based on Semantic Web languages as well as concept lattices to identify relevant communities of stakeholders depending on their past participation. These communities can be used to enable efficient decision-making and handling of requirements.

We exploited the idea of applying argumentation theory to deal with requirements engineering. In particular, the proposed framework detects consistent sets of goal-based requirements and maintains their consistency over time based on argumentation theory which allows to detect the conflicts among elements. More specifically, the framework relies on meta-argumentation, which instantiates abstract argumentation frameworks, where requirements are represented as arguments and the standard Dung-like argumentation framework is extended with additional relations between goal-based requirements. The results of this research have been published to the CLIMA international workshop [37].

6.5. Regulation Engineering

Participants: Khalil Bouzidi, Catherine Faron-Zucker, Olivier Corby.

Regulations in the Building Industry are becoming increasingly complex and involve more than one technical area, covering products, components and project implementations. They also play an important role in ensuring the quality of a building, and to minimize its environmental impact.

[12] http://dbpedia.inria.fr/qakis/

In a collaboration between CSTB and the I3S laboratory, we are carrying on research on the acquisition of knowledge from the technical and regulatory information contained in the REEF referential [13] and the automated processing of this knowledge with the final goal of assisting professionals in the use of these texts and the creation of new texts.

We are applying this work in CSTB to help industrials in the writing of Technical Assessments. The problem is how to specify these assessments and standardize their structure using models and adaptive semantic services.

A Technical Assessment (in French: Avis Technique ou ATec) is a document containing technical information on the usability of a product, material, component or element of construction, which has an innovative character. We chose this Technical Assessment as a case study because CSTB has the mastership and a wide experience in these kinds of technical documents.

In 2012, we were particularly interested in the modeling of the regulatory constraints derived from the Technical Guides used to validate the Technical Assessment. These Guides are regulatory complements offered by CSTB to the various industrials to enable easier reading of technical regulations. They collect execution details with a wide range of possible situations of implementations.

Our work aims to formalize the Technical Guides in a machine-processable model to assist the creation of Technical Assessments by automating their validation.

Our first contribution is the use of standard SBVR (Semantics of Business Vocabulary and Business Rules) and SPARQL to reformulate the regulatory requirements of guides on the one hand in a controlled language and on the other hand in a formal language

Second, our model incorporates expert knowledge on the verification process of Technical Documents. We have organized the SPARQL queries representing regulatory constraints into several processes. Each component involved in the Technical Document corresponds to an elementary process of compliance checking. An elementary process contains a set of SPARQL queries to check the compliance of an elementary component. A full complex process for checking a Technical Document is defined recursively and automatically built as a set of elementary processes relative to the components which have their semantic definition in the OntoDT ontology that we have designed.

Finally, we represent in RDF the association between the SBVR rules and SPARQL queries representing the same regulatory constraints. We use annotations to produce a compliance report in natural language to assist users in the writing of Technical Assessments.

As a result, we have designed a Semantic Web application to support and guide the process of writing Technical Assessment. The current version has allowed us to validate our approach. Also, we have developed a base of SBVR rules to describe business requirements of guides. This rule base is implemented in SPARQL.

6.6. Graph-based Knowledge Representation

Participants: Olivier Corby, Catherine Faron-Zucker, Fabien Gandon, Isabelle Mirbel, Adrien Basse, Oumy Seye.

We have designed a method to build pretty-printers for RDF Abstract Syntax Trees and Graphs, built on top of SPARQL Query Language. Pretty-print rules are written as SPARQL select-where queries. The *where* clause matches the target subtree to be printed and the *select* clause returns the pretty-printed statement using an external *kg:pprint* function. This function recursively calls the pretty printer, looking for appropriate pretty-print queries for the target subtrees.

We have designed a syntactic extension to SPARQL in order to ease writing pretty-printing rules. Below, an example of template for a SPIN like AST is shown:

```
template {
  "select " ?s
```

[13] http://reef.cstb.fr

```
  "\n"
  "where "  ?w
}
where {
  ?in ast:select ?s ;
      ast:where  ?w
}
```

We were able to write a pretty-printer for a SPIN like complete SPARQL 1.0 AST with 19 templates.

We have designed an extension to our KGRAM SPARQL interpreter that enables to consider an RDF Graph directly as a Query Graph. This enables to compute RDF subgraph matching.

We have completed SPARQL 1.1 implementation with the final version of Property Path (PP) and federated queries (service & bindings). In order to be able to query a SPARQL 1.0 endpoint with PP, we have written a compiler that translates PP into basic graph patterns.

The work on KGRAM is published in [33].

Alban Gaignard from the I3S Modalis team has designed a distributed version of KGRAM to query remote triple stores, in the context of Semantic Federation of Distributed Neurodata. This work is published in [63], [64].

We have implemented a prototype of C-SET Commutative Replicated Data Type for RDF in Corese with Pascal Molli and Luis Ibáñez from U. Nantes [69].

6.6.1. Extracting Graph Patterns to Characterize RDF Data Sources

This work takes place in the PhD Thesis of Adrien Basse.

Many Semantic Web applications address the issue of integrating data from distributed RDF triple stores. There are several solutions for distributed query processing such as SPARQL 1.1 Federation, which defines extensions to the SPARQL Query Language to support distributed query execution. Such extensions make it possible to formulate a query that delegates parts of the query to a series of services, but one issue remains: how to automate the selection of RDF triple stores containing relevant data to answer a query. This is especially true in the context of the Linking Open Data project where numerous and very heterogeneous datasets are interlinked, allowing for interesting queries across several sources. To decompose and send queries targeting only relevant stores, we need a means to describe each RDF triple store, i.e. an index structure which provides a complete and compact description of the content of the RDF triple store.

To know the content of a RDF triple store, we proposed to use graph patterns as basic structures for index items. In this thesis we present an approach to extract these graph patterns from RDF triple store. For this purpose, we extend Depth-First Search coding (DFS) [104] to RDF labeled and oriented multigraphs and we provide a join operator between two DFS codes so as to sequentially build the different levels of the index structure.

Insertion or deletion of annotations in the triple store may cause changes to the index structure. To handle updates in triple store, we proposed a procedure to identify exactly the changes in the first level of the index structure and propagate them to the following levels. The DFSR (Depth First Search for RDF) coding makes it possible for us to efficiently manipulate graph patterns, but is difficult to read (succession of integer numbers). To facilitate the reading of our index structure, we propose a visualization user-interface and algorithms to turn a DFS code into a more legible format like RDF. Our algorithm relies on Corese/KGRAM [95]. We have tested our algorithm on many datasets. During the building of index structures we keep a set of data in order to help us to better understand the progress of our algorithm and improve it.

6.6.2. Rules for the Web of Data

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

We have characterized the subset of SPARQL that can be expressed in RIF and, conversely, we have searched for the maximal RIF dialect that can be expressed as SPARQL construct-where queries. This work is published in [71], [72] were we present the implementation of a RIF dialect with a SPARQL Rule Engine in Corese/KGRAM.

We have designed online services for RIF-BLD parsers for presentation syntax and XML syntax [14]. We have also done an online service for RIF-BLD translation into SPARQL and RDF [15].

6.7. Business Intelligence

Participants: Corentin Follenfant, Olivier Corby, Fabien Gandon.

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

Industrial Business Intelligence proposes tools and methods to perform data analysis over heterogeneous enterprise sources. They allow one to harvest, federate, cleanse, annotate, query, organize and visualize data in order to support decision making with human-readable documents such as reports, dashboards, mobile visualizations. Authoring these dynamic documents requires proficiency in technical domains like relational modeling and SQL for one to produce relevant content: end users therefore praise example-driven and information retrieval (IR) systems that help them reusing existing content. Such systems need common structured metadata to enable comparison, search, matching and recommendation of (parts of) documents.

As target data sources are mainly tabular or relational, queries executed to feed the dynamic documents are SQL or derivatives. In [62] we proposed to model these queries as RDF named graphs, and use the graphs as documents annotations. Queries are represented through their abstract syntax trees (AST) represented with RDF graphs. The SQL-specific modeling contribution can therefore be applied to any generic query language. We identified two desirable features for IR systems that deal with queries repositories: search and rewriting, the latter allowing further annotation as well as reconciliation of source relational entities against LOD (Linked Open Data) repositories. On this basis we evaluated SPARQL 1.1 to perform SQL query analysis, i.e. pattern-matching search or rewriting, using in particular property paths. Resulting SPARQL queries are intuitive and concise.

Next steps include a quantitative evaluation by extracting RDF representations from a repository of SQL-fed documents, the production of a library of SPARQL queries that perform generic IR operations against RDF-modelled SQL queries, a formalization of the modeling and operations to compare them with generic tree manipulation methods. In further work we plan to investigate rewriting queries from different languages modelled with language-specific abstract syntax trees to generic abstract syntax trees and experiment cross-language query comparison with SPARQL.

6.8. Fuzzy Knowledge Representation

Participant: Andrea Tettamanzi.

Andrea Tettamanzi has joined the Wimmics research team in September 2012, after winning a position as a full professor at the University of Nice-Sophia Antipolis. He got a PhD in Computational Mathematics and Operations Research in 1995 from the University of Milan with a thesis on evolutionary algorithms; he became assistant professor at the University of Milan in 1998 and associate professor at the same university in 2002.

His research interests focus on combining different methods of computational intelligence, namely evolutionary algorithms, fuzzy logic, and neural networks, to solve real-world problems in the presence of imprecision, noisy data, and ill-defined optimization criteria, but also on the management of vagueness and uncertainty in knowledge representation, the automatic extraction of knowledge from data, possibility theory and its application to belief revision and goal generation in cognitive agents.

[14] http://wimmics-ws.inria.fr/rifparser
[15] http://wimmics-ws.inria.fr/riftosparql

After joining Wimmics, Andrea Tettamanzi has continued work on previous collaborations with other members of the team, namely Serena Villata [76], and has begun exploring with the rest of the team several research axes that could benefit from his contribution.

6.9. Emotion Modeling

Participants: Franck Berthelon, Imen Tayari, Nhan Le Thanh, Peter Sander.

In the PhD Thesis of Imen Tayari, an algebraic vector representation model of emotional states was designed. This multidimensional model provides a powerful mathematical tools for the analysis and the processing of emotions. It permits to integrate information from different modalities(speech, facial expressions, gestures) in order to allow more reliable estimation of emotional states. Indeed our proposal aims at efficient recognition of emotional states even when they appear to be superposed or masked. Experiments show the efficiency of the proposed method in detecting basic emotion by giving hight recognition rate. This work is published in [39], [41], [43], [40], [42].

In the PhD Thesis of Franck Berthelon, we are working in the domain of affective computing to create an emotion sensitive system. Interaction between human and computer can be improved via such a system because emotion is so important in everyday communication. Our research focuses on serious gaming, particularly on enabling a user and a virtual character to "share" an emotion. The two main problems that arise are:

- How to detect a user's emotions given that the stimulus comes from a virtual environment?
- How to give feedback based on the user's current emotion?

We propose to model emotions as a complex system where data are retrieved from physiological sensors such as heart rate, EMG or EEG. We need to map the multi-sensor data back into a dimensional model of emotion space. Finally, we aim to have an effect on the user's emotional state by varying the stimulus received from the virtual environment. This puts the user into different emotional situations determined by the task to accomplish, with an accompanying effect on their ability to carry out the task.

We developed an application for experimentation purposes; it aims to implement our model using an EEG headset (Emotiv EPOC). This application allows us to generate an emotional map from a slide show of emotion annotated picture. Based on the created map and the real time EEG data, this application can compute a user's instantaneous emotion.

In addition to the first development, we reuse experimental data from MIT to validate our model in a more controlled way. We take the same data, features, signal processing and feature reduction algorithm but instead of using the k-nearest neighbors (KNN) classification algorithm we use our model to identify and annotate discontinuities that represents emotion state changes in accordance with Klaus R. Scherer hypothesis [103].

We are continuing work on validating our model with quantitative results and applying those results on a more realistic system with the application we have developed.

6.10. DBpedia in French

Participants: Julien Cojan, Fabien Gandon.

The purpose of the project *DBpedia in French* is to extract data from Wikipedia in French and publish it under structured format. Wikipedia content is mainly meant to be read by human and is not suited for use in applications. DBpedia publishes the data extracted from Wikipedia articles in RDF W3C standard for the Semantic Web [16] thus readily available for many applications. For instance, DBpedia is used to generate indexes for cultural resources (e.g. HdA-lab project [17]), it can also be used for mobile applications thanks to the geographic data it contains, or to answer natural language questions, etc.

[16] http://www.w3c.org/RDF/
[17] http://hdalab.iri-research.org/hdalab/

The original version of DBpedia is focused on the English chapter of Wikipedia. Last versions also contain elements extracted from other chapters, but only when related to a page in English. Articles with no equivalent in English are skipped, leading to a significant number of pages being ignored and so a significant amount of data is lost. For instance, about 49 000 persons and 180 000 places described in the French chapter have no corresponding article in English and are then missing in the English DBpedia. Moreover, the description of the same topic can be different from one chapter to another, reflecting cultural diversity.

DBpedia in French publishes data extracted from the French Wikipedia in complement to the English DBpedia. Data are linked with the different chapters from the internationalization committee thus providing multilingual resources. In its release from October 2nd, DBpedia in French contains 130 million triples describing 1.3 million things, among them 260 000 places, 140 000 persons, 64 000 work pieces and 26 000 organizations.

This project is supported by the Semanticpedia collaboration platform [18] launched November 19th 2012 by Aurélie Filipetti, the French Ministry of Culture, Michel Cosnard, CEO of Inria, and Rémi Mathis, CEO of Wikimédia France. Inria currently hosts the project [19] and is the correspondent for the French chapter in DBpedia internationalization committee.

6.11. Co-Construction of Community Ontologies

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

PhD Thesis on Co-Construction of Community Ontologies and Corpus in a Limited Technological Environment.

To refresh the memory of people and revive many stories that accompany the creation and daily life of different Africa territories, the establishment of an online sociocultural encyclopedia was conceived. It will be an online platform based on a Geographical Information System (GIS) enriched by a semantic layer allowing access to different information.

In the last decade, we have seen the rise of two visions of the Web: on one side the Social Web or Web 2.0, which places users at the heart of the Web, they are no longer spectator but become editor of the content of Web pages. On another hand, the Semantic Web proposes knowledge representations (ontologies) that allow machines (software agent) to better understand data on the Web. Both aims were often opposed, but there are a lot of work trying to combine these visions.

In our work, oriented in this direction, we will try to create a new point of view of Community concept within the Web. Community is a group of people who share a common set of values and interests. This shift of view allows us to address a specific community as an atomic entity and focus this time on the sharing of knowledge between communities. The second challenge is to combine Social Web and Semantic Web technologies. Using Semantic Web in our social network, we have a semantic layer that provides access to various information contained in the network. Furthermore, the Semantic Web opens up a semantic approach to social network analysis, which also allows extracting new knowledge.

In this thesis, we study the implementation of an online platform to build and share the collective memory of citizens and revive many stories by using a semantic layer. Semantic Social Network Analysis will allow us to present data in eye-catching way and in another view. This platform will be updated by the actors and the citizens of these territories, and share their history and heritage through their "social network".

6.12. Semantic Wiki

Participants: Pavel Arapov, Michel Buffa.

[18] http://semanticpedia.org
[19] http://fr.dbpedia.org

A Wiki is a Web site that lets users create and edit content collaboratively. A Wiki engine is a programming base to create Wiki sites. In this PhD Thesis, our approach to the creation of a Wiki engine is to use an application on the Wiki pages and semantic meta-data. Our vision for the Wiki pages is that a Wiki page is an application. We do not work with static data on the Internet neither in a Wiki, now it is a Web application that contains the source code of the application interface, as well as the data for display. Application is able to retrieve and update data based on Linked Data principles on each page load, updating their sources as needed and showing only relevant information. This work is published [51], [52].

6.13. Discovery Hub

Participants: Nicolas Marie, Fabien Gandon, Damien Legrand.

Nicolas Marie is PhD student in collaboration with Alcatel-Lucent Bell Labs (Cifre). He is the project leader of Discovery Hub: a discovery engine on the top of DBpedia using real-time spreading activation.

We continue the CRE and CIFRE PhD Thesis (2011-2013) initially on Social objects, object-centered sociality, and object-centered social networks to propose mobile context-based notification application in a semantic and pervasive Web. The work evolved toward exploratory search, discovery and recommendation. Web growth, both in size and diversity, and users' growing expectations increase the need for innovative search approaches and technologies. Exploratory search systems are built specifically to help user in cognitive consuming search tasks like learning or investigation. Some of these systems are built on the top of linked data and use its semantic richness to provide cognitively-optimized search experiences. This work addresses the question of real-time linked data processing for exploratory search purposes. This live aspect offers advantages in terms of query expressivity and data dynamicity-handling.

To achieve this goal we propose a real-time semantic spreading activation algorithm (RTSA) which process linked data on-the-fly. This live aspect offers advantages in data dynamicity handling and query expressivity. Approximation strategies, algorithm behavior study and user evaluation related to RTSA algorithm are currently performed. The work includes a study of its behavior on DBpedia and a validation of its relevance through a user evaluation. Finally we also implemented a real deployment introducing the Discovery Hub prototype. It is an exploratory search engine offering advanced querying, browsing and explanation strategies for discovery purposes.

This algorithm is deployed in the Discovery Hub prototype [20], a discovery engine interfaced with services. Discovery Hub aims to help users to make numerous discoveries starting from its interests. The application works on DBpedia data including local version like fr.dbpedia.org (hosted by Inria/Wimmics). It also makes extensive use of the Corese/KGRAM Semantic Web Factory. Application front-end was designed and developed by Damien Legrand during an internship.

The application was presented during the Semanticpedia day, organized by official French language delegation [21].

Nicolas Marie is also active in the Web Science community [36].

6.14. Semantic Newsfeed Aggregation

Participant: Christophe Desclaux.

Christophe was this year in his last year of engineering school at Polytech UNS in the KIS speciality. During his end of course project he worked on the RSS feed aggregation using Named Entities Recognition. He presented his research project to the student contest *Boost Your Code* organized by Inria. The aim of the contest is to offer to a junior engineer a one year full time contract to work on an innovating OpenSource project. Christophe won the contest and is now part of the Wimmics team since november 2012. The ZONE project [22] provides a specialized tool for monitoring domain. ZONE semantically increases news for a better classification for the user. Christophe will work in collaboration with the team on documents clustering, natural language processing and RDF datastores.

[20] http://semreco.inria.fr/
[21] http://www.dglflf.culture.gouv.fr/Actualites/Programme_Semanticpedia.pdf
[22] http://zone-project.org

6.15. Linked Justifications

Participants: Rakebul Hasan, Fabien Gandon.

Semantic Web applications use inferential capabilities and distributed data in their reasoning. Users often find it difficult to understand how these applications produce their results. Hence, users often do not trust the results these applications produce. Explanation-aware Semantic Web applications provide explanations of their reasoning. Explanations enable users to better understand reasoning of these application. Users can use this additional information about reasoning to make their trust decisions.

The emergence of Linked Data offers opportunities for large-scale reasoning over heterogeneous and distributed data. Explaining reasoning over Linked Data requires explaining how these distributed data were produced. Publishing also the explanation related metadata as Linked Data enables such explanations. Justifications are metadata about how a given piece datum is obtained. We introduce the concept of Linked Justifications and provide guidelines to publish justifications as Linked Data in [67]. We published the *Ratio4TA*[23] (inte**r**linked justific**ation**s **fort**riple **a**ssertions) vocabulary to describe justifications. *Ratio4TA* extends W3C PROV Ontology[24] to promote interoperability.

In [89], [66], we analyze the existing explanation-aware Semantic Web systems. The existing systems inherit explanation features from explanation-aware expert systems. These explanations are targeted to expert users, such as knowledge engineers, with detailed information about all the execution steps of reasoners of these applications. Unlike the expert systems, users of Semantic Web applications have diverse background - from expert knowledge engineers who are interested in every details of the reasoning, to regular users who do not have any background in reasoning, logic, or ontologies. These non-expert users might feel overwhelmed with all the execution details of reasoners. To address this issue, we propose summarized and relevant explanations to users. Users can specify their explanation goals - types of information they are interested in. We take into consideration the explanation goals when we present explanations and summarize explanations. We use centrality and similarity matrices to summarize and provide relevant explanations.

6.16. Analyzing and Modeling Users, Communities and their Interactions in a Social Semantic Web Context

6.16.1. Analyzing and Modeling the Sharing and Articulation of Representations

Participants: Alain Giboin, Gessica Puri.

Comparing and Bridging Models of Representation Sharing Processes

Context: Follow-up to the RefCom joint research action of the GDR CNRS Psycho-Ergo, in collaboration with Pascal Salembier (UTT, France).

We continued our work on comparing and bridging models of representation sharing processes (see Edelweiss activity report 2011) in order to achieve mutual intelligibility between researchers working on such models. We extended at the European level the test and application of the grid we elaborated for collaboratively comparing and bridging the conceptualizations [86], [65].

Methods and tools for articulating developers, domain experts, users and usage analysts' representations

Reconciling informal and formal representations through the ECCO collaborative ontology editor

Context: Follow-up to the ANR project e-WOK HUB, in collaboration with Priscille Durville (a former Inria expert engineer, currently engineer at Xerox), Sandrine Grataloup and Olivier Morel (BRGM), Michel Perrin (ENSMP)

[23] http://ns.inria.fr/ratio4ta/
[24] http://www.w3.org/TR/prov-o/

In a new publication [81], we reported the method we designed and proposed to geologists and knowledge engineers to help them jointly define domain ontologies from textual documents. The method is instrumented through a collaborative ontology editor (ECCO) which integrates two tools for automatic analysis of natural language. ECCO allows articulating the informal representations formulated by geologists in natural language and the knowledge engineers' fully formalized ontology-based representations that can be processed automatically by a semantic search engine like Corese.

Frameworks and Toolsets

Context 1: Capitalizing methods and tools developed in the Wimmics (formerly Edelweiss) team

We started to integrate into a general framework the methods and tools developed in the Wimmics (formerly Edelweiss) team to help developers, domain experts, users and usage analysts reconcile their views in order to design user-adapted social semantic applications. A preliminary presentation of the framework was given in the interdisciplinary seminar mentioned in Section "Invited Talks".

Context 2: PhD thesis of Gessica Puri and Wimmics projects related to visualization and manipulation of links.

We are currently developing a "design thinking" toolset (including a framework) for helping developers think in terms of a user's point of view when they design and evaluate link visualization and manipulation applications such as graph visualization applications [90]. A first version of the toolset is being validated by developers of the team.

6.16.2. Scenario Modeling and User Modeling for system design and evaluation

Participants: Alain Giboin, Gessica Puri.

In the context of different design projects, we applied, adapted or renewed some of our scenario and user modeling methods.

Context 1: ISICIL project. In collaboration with Rubiela Silva (UNS & Ademe), Claire Prendleloup (Ademe), Mylène Leitzelman (Telecom ParisTech)

In order to evaluate the usefulness and usability of the ISICIL platform (see 6.2) for one of the communities of potential users of the platform (the French Environment and Energy Management Agency), we adapted in particular: (a) a technique for modeling collectives and their related scenarios to prepare testing situations as close as possible to real situations met by the potential users; and (b) an existing set of collective heuristics, i.e., heuristics formerly designed to evaluate groupware; see, e.g. [93]. We also proposed the technique of *Online collaborative sessions* supported by the ISICIL communication functionalities (chat and comments) as a writing-based version of the *think out loud* protocol. These techniques are partly described in [91], [92].

Context 2: PAL project, in collaboration with David Daney (Coprin), Rémi Barraquand (Prima), Nadine Mandran (Pôle d'Ingénierie Multidisciplinaire du Laboratoire d'Informatique de Grenoble)

In this work we applied and adapted techniques for articulating system-oriented scenarios (coming from robotics and computer vision specialists) and user-oriented scenarios. These scenarios were intended to guide the design of useful and usable services improving the autonomy and quality of life for elderly and fragile persons. An output of this work has been the organization of workshop where PAL researchers were invited to specify the scenarios motivating the development of their services.

Context 3: Projects *Socio-cultural encyclopedia of Senegalese communities* and *Global Warming Platform* – Papa Fary Diallo (Wimmics PhD student), Fatou Kamara and Moussa Lo (Université Gaston-Berger, Saint-Louis, Sénégal)

We started to apply techniques for modeling groups of users and their related scenarios to the design of social semantic applications aimed at (a) communities wanting to adapt these applications to their own culture, and (b) at communities including persons from different specialties (geographers, mathematicians) and laypersons which want to collaborate on global warming issues.

Context 4: Discovery Hub project, PhD thesis of Nicolas Marie, in collaboration with Florentin Rodio (Alcatel Lucent); also related to the PhD thesis of Gessica Puri.

In order to perform a user-oriented evaluation of the Discovery Hub recommender system (see 6.13 and [100]), we used techniques allowing to define more realistic scenarios of interaction with the recommender system from the user's point of view, and to analyze users' cognitive processes when they interact with the system (e.g., when they select recommended items or when they assess the quality of a recommendation with and without explanations).

6.16.3. Exploring eye-tracking techniques for evaluating information organization aspects in Web applications

Participants: Valériane Dusaucy, Alain Giboin.

Context: collaboration with Valériane Dusaucy (PhD student, University of Aix-Marseille & CIFRE Société Ausy) and Franck Ferront (ergonome, Société Ausy)

We designed an experiment to explore the potentiality of eye-tracking techniques for evaluating information organization aspects in Web applications from a user's point of view, and to compare it to other evaluation techniques such as heuristic evaluation. The experiment, which takes place in the Ubiquarium of the I3S Laboratory, is in progress.

ZENITH Project-Team

6. New Results

6.1. Data and Metadata Management

6.1.1. Uncertain Data Management

Participants: Reza Akbarinia, Patrick Valduriez, Guillaume Verger.

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

One of the areas, in which uncertainty management is important, is the integration of heterogeneous data sources, in the sense where usually there may be an uncertainty in the possible mappings between the attributes of the sources. Usually the human interaction is demanded to help the system in choosing the correct mappings. In [30], we propose a pay-as-you-go data integration solution that aims at preforming the data integration in a fully automated way. Our solution takes advantage of attribute correlations by using functional dependencies, and captures uncertainty in mediated schemas using a probabilistic data model. It allows integrating a given set of data sources, as well as incrementally integrating additional sources, without needing to restart the process from scratch. We implemented our solution, and compared it with a baseline approach. The performance evaluation results show significant performance gains of our solution in terms of recall and precision compared to the baseline approaches.

Another problem that arises in many applications such as data integration systems is that of Entity Resolution (ER). ER is the process of identifying tuples that represent the same real-world entity. It has been well studied in the literature for certain data, but it has not been deeply investigated for uncertain data. Existing proposals for the ER problem are not applicable to the above examples since they ignore probability values completely and return the most similar tuples as the solution. Furthermore, the semantics of the solution for the ERUD problem has not been clearly defined in the literature. In [31], we address the ERUD problem. We adopt the well-known possible worlds semantics for defining the semantics for the ERUD problem, and propose a PTIME algorithm for a large class of similarity functions, i.e. context-free. For the rest of similarity functions, i.e. context-sensitive, we use Monte-Carlo randomization for approximating the answer. We propose a parallel version of our Monte-Carlo algorithm using the MapReduce framework. To the best of our knowledge, this is the first study of the ERUD problem that adopts the possible world semantics and the first efficient algorithm for implementing it.

Another important problem in uncertain data management is the efficient processing of probabilistic queries. We have continued the development of our probabilistic database prototype, called ProbDB (Probabilistic Database) that deals with large-scale probabilistic data sharing. ProbDB divides each probabilistic query into two parts: probabilistic and deterministic (i.e. non probabilistic). The deterministic part is executed by the underlying RDBMS, and the rest of work is done by our probabilistic query processing algorithms that are executed over the data returned by the RDBMS.

6.1.2. Metadata Integration

Participants: Zohra Bellahsène, Emmanuel Castanier, Duy Hoa Ngo, Patrick Valduriez.

Our work on metadata integration encompassed ontology matching and open data source integration.

The major focus of our work in 2012 was to deal with large scale ontology matching and scalability. To improve the matching quality of YAM++, we designed a new IR-based measure to deal with terminological heterogeneity in real world ontologies. To deal with large ontology matching, we designed a method based on indexing concepts from their labels and comments. Our approach aims at reducing the search space when comparing the concepts of the input ontologies. For this purpose, we designed three filters: Description Filter, Context Filter and Label Filter. These methods make use of the Lucene search engine for indexing and searching the context of entities in the input ontologies. Another contribution lies on the Fast Semantic Filtering method, which refines the discovered mappings in the ontology matching task. The aim of the Semantic Filter is to detect and reject inconsistent mappings by exploring semantic information of entities in the input ontologies [45]. The originality of our method is to use a new structural indexing technique and a heuristic to generate relative disjointness axioms. At the 2012 competition of the Ontology Alignment Evaluation Initiative (http://oaei.ontologymatching.org), YAM++ was one of the best matchers, with very good results in all tracks. It obtained the first postision in the Large BioMed Track [55].

Integrating open data sources can yield high value information but raises major problems in terms of metadata extraction, data source integration and visualization of integrated data. In [34], [33], we describe WebSmatch, a flexible environment for Web data integration, based on a real, end-to-end data integration scenario over public data from Data Publica. WebSmatch supports the full process of importing, refining and integrating data sources and uses third party tools for high quality visualization. We use a typical scenario of public data integration which involves problems not solved by currents tools: poorly structured input data sources (XLS files) and rich visualization of integrated data.

6.1.3. High-dimensional data management

Participants: Mohamed Riadh Trad, Alexis Joly, Saloua Litayem.

High dimensional data hashing is essential for scaling up and distributing data analysis applications involving feature-rich objects, such as text documents, images or multi-modal entities (scientific observations, events, etc.). In this first research track, we first investigated the use of high dimensional hashing methods for efficiently approximating K-NN Graphs [47], particularly in distributed environments. We highlighted the importance of balancing issues on the performance of such approaches and show why the baseline approach using Locality Sensitive Hashing does not perform well. Our new KNN-join method is based on RMMH, a hash function family based on randomly trained classifiers that we introduced in 2011. We show that the resulting hash tables are much more balanced and that the number of resulting collisions can be greatly reduced without degrading quality. We further improve the load balancing of our distributed approach by designing a parallelized local join algorithm, implemented within the MapReduce framework. In other work [43], we address the problem of speeding-up the prediction phase of linear Support Vector Machines via Locality Sensitive Hashing. Whereas the mainstream work in the field is focused on training classifiers on huge amount of data, less efforts are spent on the counterpart scalability issue: how to apply big trained models efficiently on huge non annotated collections ? In this work, we propose building efficient hash-based classifiers that are applied in a first stage in order to approximate the exact results and alter the hypothesis space. Experiments performed with millions of one-against-one classifiers show that the proposed hash-based classifier can be more than two orders of magnitude faster than the exact classifier with minor losses in quality.

6.2. Data and Process Sharing

6.2.1. Hybrid P2P/cloud Architecture

Participants: Esther Pacitti, Patrick Valduriez.

Zenith adopts a hybrid P2P/cloud architecture. P2P naturally supports the collaborative nature of scientific applications, with 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 some of their data (a major requirement for our application partners). But for very-large scale data analysis or very large workflow activities, cloud computing is appropriate as it can provide virtually infinite computing, storage and networking resources. Such hybrid architecture also enables the clean integration of the users' own computational resources with different clouds.

In [24], we define Zenith's architecture with P2P data services and cloud data services. We model an online scientific community as a set of peers and relationships between them. The peers have their own data sources. The relationships are between any two or more peers and indicate how the peers and their data sources are related, e.g. friendship, same semantic domain, similar schema. The P2P data services include basic services (metadata and uncertain data management): recommendation, data analysis and workflow management through the Shared-data Overlay Network (SON) middleware. The cloud P2P services include data mining, content-based information retrieval and workflow execution. These services can be accessed through web services, and each peer can use the services of multiple clouds.

6.2.2. Social-based P2P Data Sharing

Participants: Reza Akbarinia, Emmanuel Castanier, Esther Pacitti, Didier Parigot, Patrick Valduriez, Guillaume Verger.

As a validation of the ANR DataRing project, we have developed P2PShare, a P2P system for large-scale probabilistic data sharing in scientific communities. P2PShare leverages content-based and expert-based recommendation. It is designed to manage probabilistic and deterministic data in P2P environments. It provides a flexible environment for integration of heterogeneous sources, and takes into account the social based aspects to discover high quality results for queries by privileging the data of friends (or friends of friends), who are expert on the topics related to the query.

Using the Shared-Data Overlay Network (SON), we have implemented a prototype of P2PShare that integrates three major DataRing services: ProbDB, a probabilistic database management service for relational data; WebSmatch, an environment for Web data integration; and P2Prec, a social-based P2P recommendation service for large-scale content sharing.

In [50], , we describe the demo of P2PShare's main services, e.g., gossiping topics of interest among friends, key- word querying for contents, and probabilistic queries over datasets.

6.2.3. View Selection in Distributed Data Warehousing

Participants: Zohra Bellahsène, Imen Mami.

Scientific data generate large amounts of data which have to be collected and stored for analytical purpose. One way to help managing and analyzing large amounts of data is data warehousing, whereby views over data are materialized [23]. At large scale, a data warehouse can be distributed. We have examined the problem of choosing a set of views and a set of data warehouse nodes at which these views should be materialized so that the full query workload is answered with the lowest cost. To address this problem, we extended our view selection method that we proposed for the centralized case. Thus, we modelled the distributed view selection problem as a Constraint Satisfaction Problem (CSP). Furthermore, we introduced the distributed AND-OR view graph, which can be seen as an extensive form of the AND-OR view graph to reflect the relation between views and communication network within the distributed scenario. The experiment results show that our approach provides better performance compared with the genetic algorithm in term of the solution quality (i.e., the quality of the obtained set of materialized views). We demonstrated experimentally that our approach provides better results in term of cost savings when the view selection is decided under space and maintenance cost constraints [44].

6.2.4. Scientific Workflow Management

Participants: Ayoub Ait Lahcen, Jonas Dias, Didier Parigot, Patrick Valduriez.

Scientific experiments based on computer simulations can be defined, executed and monitored using Scientific Workflow Management Systems (SWfMS). Several SWfMS are available, each with a different goal and a different engine. Due to the exploratory analysis, scientists need to run parameter sweep (PS) workflows, which are workflows that are invoked repeatedly using different input data. These workflows generate a large amount of tasks that are submitted to High Performance Computing (HPC) environments. Different execution models for a workflow may have significant differences in performance in HPC. However, selecting the best execution model for a given workflow is difficult due to the existence of many characteristics of the workflow that may affect the parallel execution.

In [36], we develop a study to show performance impacts of using different execution models in running PS workflows in HPC. Our study contributes by presenting a characterization of PS workflow patterns (the basis for many existing scientific workflows) and its behavior under different execution models in HPC. We evaluated four execution models to run workflows in parallel. Our study measures the performance behavior of small, large and complex workflows among the evaluated execution models. The results can be used as a guideline to select the best model for a given scientific workflow execution in HPC. Our evaluation may also serve as a basis for workflow designers to analyze the expected behavior of an HPC workflow engine based on the characteristics of PS workflows.

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

In the context of SON, we also proposed a declarative workflow language based on service/activity rules. In [27], [46], we present a formal approach that combines component-based development with well-understood methods and techniques from the field of Attribute Grammars and Data-Flow Analysis in order to specify the behavior of P2P applications, and then construct an abstract representation (i.e., Data-Dependency Graph) to perform analyzes on it. This formal approach makes it possible to infer a dependency graph for SON applications that provides for automatic parallelization.

6.2.5. Plants identification and classification from social image data

Participants: Hervé Goëau, Alexis Joly, Saloua Litayem.

This work is done in collaboration with the botanists of the AMAP UMR team (CIRAD) and with Inria team IMEDIA. Inspired by citizen sciences, the main goal of this trans-disciplinary work is to speed up the collection and integration of raw botanical observation data, while providing to potential users an easy and efficient access to this botanical knowledge. We therefore did continue working intensively on plants identification and classification [54], [37], [38], [26]. We first developed a new interactive method [37] for the visual identification of plants from social image data. Contrary to previous content-based identification methods and systems that mainly relied on leaves, or in few other cases on flowers, it makes use of five different organs and plant's views including habit, flowers, fruits, leaves and bark. Thanks to an interactive query widget, the tagging process of the different organs and views is as simple as drag-and-drop operations and does not require any expertise in botany. All training pictures used by the system were continuously collected during one year through a crowdsourcing application and more than 17K images are now integrated. System-oriented and human-centered evaluations of the application show that the results are already satisfactory and therefore very promising in the long term to identify a richer flora.
Besides, we did continue working on leaf-based identification notably through the organization of and participation to ImageCLEF plant identification evaluation campaign 2012 [54].
Finally we did apply one of our former work related to multi-source shared-nearest neighbors clustering to an original experiment aimed at evaluating if we were able to automatically recover morphological classifications built by the botanists themselves [38]. The results are very promising, since all clusters discovered automatically could be easily matched to one node of a morphological tree built by botanists.

6.3. Scalable Data Analysis

6.3.1. StreamCloud

Participants: Vincenzo Gulisano, Patrick Valduriez.

Recent years have witnessed the growth of a new class of data-intensive applications that do not fit the DBMS query paradigm. Instead, the data arrive at high speeds taking the form of an unbounded sequence of values (data streams) and queries run continuously returning new results as new data arrive. Examples of data streams are sensor data (e.g. in environmental applications) or IP packets (e.g. in a network monitoring application). The unbounded nature of data streams makes it impossible to store the data entirely in bounded memory. Current research efforts have mainly focused on scaling in the number of queries and/or query operators having overlooked the scalability with respect to the stream volume.

Current Stream Processing Engines do not scale with the input load due to single-node bottlenecks. Additionally, they are based on static configurations that lead to either under or over-provisioning. In [21], [22], we present StreamCloud, a scalable and elastic stream processing engine for processing large data stream volumes. StreamCloud uses a novel parallelization technique that splits queries into subqueries that are allocated to independent sets of nodes in a way that minimizes the distribution overhead. Its elastic protocols exhibit low intrusiveness, enabling effective adjustment of resources to the incoming load. Elasticity is combined with dynamic load balancing to minimize the computational resources used. We present the system design, implementation and a thorough evaluation of the scalability and elasticity of the fully implemented system.

6.3.2. Mining Uncertain Data Streams

Participants: Reza Akbarinia, Florent Masseglia.

Dealing with uncertainty by using probabilistic approaches has gained increasing attention these past few years. One of the main requirements for uncertain data mining is the ability to discover Probabilistic Frequent Itemsets (PFI). However, PFI mining, particularly in uncertain data streams, is very challenging and needs the development of new techniques, since approaches designed for deterministic data are not applicable in this context. In [29], we propose an efficient solution for exact PFI mining over data streams with sliding windows. Our proposal includes efficient solutions for updating frequentness probability of itemsets and thus fast extraction of PFI, whenever transactions are added or removed from the sliding window. To the best of our knowledge, this is the first efficient solution for data stream PFI mining. We have conducted an extensive experimental evaluation of our approach over synthetic and real-world data sets; the results illustrate its very good performance.

6.3.3. Detecting Rare Events in Massive Datasets

Participant: Florent Masseglia.

In this work, we consider that rare events are very small clusters typically representing less than 0.01% of the entire dataset. Finding these abnormal events allows to identify the emergence of pos- sible anomalies in their very early stages. Such a scenario is generally difficult to handle as it lies at the frontier between outlier detection and clustering and is characterized by a clear challenge to avoid false nega- tives. To address this challenge, we take a backward approach and pro- pose RARE, a framework that identifies and isolates the abnormal/rare regions. The dense regions are identified using a radius-limited density- driven variant of k-means and adjacent regions are merged to form new regions. These newly formed regions are gradually augmented as long as a density-driven condition is respected. When no more dense regions are observed, the remaining data is clustered and presented for further analysis to human experts. The framework is tested on a medical appli- cation and compared against human analysis. The experiments show that rare events that were missed during human analysis because of the multivariate character of the data can be discovered by our approach.

This work is funded by the labex NUMEV and a patent application involving Inria, CNRS, UM2 and INSERM has been filled.

6.3.4. Highly Informative Feature Set Mining

Participant: Florent Masseglia.

For many textual collections, the number of features is often overly large. As these features can be very redundant, it is desirable to have a small, succinct, yet highly informative collection of features that describes the key characteristics of a dataset. Information theory is one such tool for us to obtain this feature collection. In [48], we mainly contribute to the improvement of efficiency for the process of selecting the most informative feature set over high-dimensional unlabeled data. We propose a heuristic theory for informative feature set selection from high dimensional data. Moreover, we design data structures that enable us to compute the entropies of the candidate feature sets efficiently. We also develop a simple pruning strategy that eliminates the hopeless candidates at each forward selection step. We test our method through experiments on real-world data sets, showing that our proposal is very efficient.

6.3.5. Clustering Users with Evolving Profiles in Usage Streams

Participant: Florent Masseglia.

Existing data stream models commonly assume that users' records or profiles in data streams will not be updated once they arrive. In many applications such as web usage, however, the users' records/profiles may evolve along time. This kind of streaming transactions are referred to as bi-streaming data (*i.e.* the data evolves temporally in two dimensions, the flowing of transactions as with the traditional data streams, and the evolving of users' profiles inside the streams, which makes bi-streaming data different from traditional data streams). The two-dimensional evolving of bi-streaming data brings difficulties on modeling and clustering for exploring the users' behaviors. In [49], we propose three models to summarize bi-streaming data, which are the batch model, the Evolving Objects (EO) model and the Dynamic Data Stream (DDS) model. Through creating, updating and deleting user profiles, the models summarize the behaviors of each user as an object. Based on these models, clustering algorithms are employed to identify the user groups. The proposed models are tested on a real-world data set showing that the DDS model can summarize the bi-streaming data efficiently and effectively, providing better basis for clustering user profiles than the other two models.

6.3.6. Scalable Mining of Small Visual Objects

Participants: Pierre Letessier, Julien Champ, Alexis Joly.

Automatically linking multimedia documents that contain one or several instances of the same visual object has many applications including: salient events detection, relevant patterns discovery in scientific data or simply web browsing through hyper-visual links. Whereas efficient methods now exist for searching rigid objects in large collections, discovering them from scratch is still challenging in terms of scalability, particularly when the targeted objects are rather small. In this work [40], we formally revisit the problem of mining or discovering such objects, and then generalized two kinds of existing methods for probing candidate object seeds: weighted adaptive sampling and hashing based methods. We then introduce a new hashing strategy, working first at the visual level, and then at the geometric level. Experiments conducted on millions of images show that our method outperforms state-of-the-art.
This method was integrated within a visual-based media event detection system in the scope of a French project called the transmedia observatory. It allows the automatic discovery of the most circulated images across the main news media (news websites, press agencies, TV news and newspapers). The main originality of the detection is to rely on the transmedia contextual information to denoise the raw visual detections and consequently focus on the most salient trans-media events. This work was presented at ACM Multimedia Grand Challenge 2012 [39]. The movie presented during this event is available at http://www.otmedia.fr/?p=217.

COPRIN Project-Team

6. New Results

6.1. Interval analysis

6.1.1. A Contractor Based on Convex Interval Taylor

Participants: Gilles Trombettoni [correspondant], Bertrand Neveu.

Interval Taylor has been proposed in the sixties by the interval analysis community for relaxing continuous non-convex constraint systems. However, it generally produces a non-convex relaxation of the solution set. A simple way to build a convex polyhedral relaxation is to select a *corner* of the studied domain/box as expansion point of the interval Taylor form, instead of the usual midpoint. The idea has been proposed by Neumaier to produce a sharp range of a single function and by Lin and Stadtherr to handle $n \times n$ (square) systems of equations.

This paper presents an interval Newton-like operator, called `X-Newton`, that iteratively calls this interval convexification based on an endpoint interval Taylor. This general-purpose contractor uses no preconditioning and can handle any system of equality and inequality constraints. It uses Hansen's variant to compute the interval Taylor form and uses two opposite corners of the domain for every constraint.

The `X-Newton` operator can be rapidly encoded, and produces good speedups in constrained global optimization and constraint satisfaction. First experiments compare `X-Newton` with affine arithmetic[31], [19], [20]

6.2. Robotics

6.2.1. Robotics

6.2.1.1. Kinematics of wire-driven parallel robots

Participants: Laurent Blanchet, Jean-Pierre Merlet [correspondant].

The kinematics of wire robot is a complex problem because it involves both the geometrical constraints and the static equilibrium constraints as only positive tensions in the wire are possible. A major issue, that has not been addressed in the literature [16], [15], is that for a robot having n wires the forward kinematic problem (FK) (determining the possible pose(s) of the robot knowing the wire lengths, a problem that is crucial to solve for controlling the robot) cannot be solved by assuming that all n wires are under tension as the current pose of the robot may be such that only a subset of the wires may be under tension. Hence the FK problem has to be solved for **all** robots that may be derived from the initial one by removing 1 one to $n-1$ wires, each solving leading to a set of possible poses for the platform. Solving the FK for 1 wire is trivial, while for 6 wires the FK solving may be based on the already complex FK of parallel robot with rigid legs. For 2 wires it can be shown that the FK solutions can be found by solving a 12th order univariate polynomial, while for 3 wires we have shown last year by using an elimination procedure that the solutions are obtained by solving a 158th order polynomial. A very recent result of this year is that for 4 wires the order of this polynomial is 216, while no known result has been established for 5 wires (note that for 3 to 5 wires the size of the system of equations that has to be solved for the FK is larger than the one for the FK of 6-dof robot with rigid legs, a problem that has required 20 years to be solved).

Drawbacks of the elimination approach is that it does not take into account 1) that the solution should be mechanically stable, 2) that the wire tensions at the solution(s) must be positive. Hence, assuming that all solutions may be computed by the elimination approach, an a-posteriori analysis has to be performed to sort out the solutions that verify 1) and 2). We have proposed this year an efficient method to determine if a solution was mechanically stable [13]. But another major issue with the elimination method is that it leads to high order polynomial that cannot be safely numerically solved. To address this problem and 2) we are considering a numerical algorithm based on interval analysis, that consider also the tension as unknowns, hence allowing to search only for solution(s) with positive tensions [21].

Another issue for wire-driven parallel robots is the concept of redundancy. Having more wires than dof to be controlled is interesting for increasing the workspace of the robot. But it is believed that redundant wires may also be used to better distribute the load among the wires. Unfortunately we have shown for the $N-1$ robot (all N wires connected at the same point on the platform) with non elastic wires that whatever N there will be at most 3 wires under tension simultaneously [25] and consequently that tension management is not possible (with 3 wires the tensions is uniquely determined). If the wires are elastic, then tension management is possible but the positioning error is very sensitive to errors in the stiffness model [24]. Hence new method for tension management should be devised and we have explored some possibilities [23]. Still there is some magic in wire-driven parallel robots: in spite of all the uncertainties prototypes work quite well, a phenomenon which has been explained through a sensitivity analysis [24], [22].

Finally we address the management of modular robots, whose geometry can be adapted to various tasks and different objects to be manipulated, especially for very large scale robot [28], that may be used in industry for maintenance and logistics (see the Cablebot project in section 8.2.1).

6.2.1.2. Robot Calibration

Participants: Thibault Gayral, David Daney [correspondant], Jean-Pierre Merlet.

6.2.1.2.1. Experimental calibration of a high-accuracy space telescope

A collaborative work began in October 2010 with Thales Alenia Space on the calibration of the mechanical structure of a space telescope. Its architecture is based on a parallel manipulator (of the active wrist 6-PUS type, which has been patented by COPRIN) and is used to correct the relative position of two mirrors. The aim is to reach a micrometer accuracy in order to obtain a suitable quality of the images provided by the telescope. Thus, a complete model of the space telescope needs to be developed and validated through calibration. Since high velocity is not required in such an application, the dynamic effects can be neglected and only geometric and/or static calibration has to be considered. Moreover, measurements for calibration were performed in a clean room under controlled pressure, temperature and humidity conditions to minimize the influence of the non-geometric errors. Thus, two possible static inaccuracy sources were identified and modeled: one from the deformation of the mobile platform and the other resulting from the behavior of the flexure joints. Three incremental models of the flexure joints were developed and compared: a spherical joint model, a model issued from the beam theory and a stiffness model. Results of calibration using an accurate measurement system of photogrammetry showed that the flexure joints can be modeled by perfect spherical joints due to the small workspace of the telescope. Concerning the mobile platform deformation, two models were developed. Good accuracy results were obtained for both models. The developed models allowed us to explain how the model errors are directly accounted in the parameter identification during calibration. This resulted in different sets of identified parameters which all enable a good positioning accuracy. Those differences were explained and results of calibration allow a proper choice of the model of the mobile platform deformation. Considering this model, a positioning accuracy of some micrometers was finally reached after calibration with only position and orientation measurements of the mobile platform, which should allow the calibration of the telescope in space [33]. This is currently under study using interferometric measurements on the prototype of the space telescope.

6.2.1.2.2. Calibration of a cable-driven robot

To improve the accuracy of a cable manipulator, it is necessary to identify the uncertainties of its model. The cable robots, studied in the ANR funded project Cogiro (see section 8.1.1.2), are theoretically redundantly actuated: the number of powered wires is larger than the number of degrees of freedom of the manipulator (however see section 6.2.1.1 about the reality of this redundancy).

In 2011 an over-constrained prototype was self-calibrated (the identification of the parameters does not need additional external measurement), under some assumptions on the cable properties [17], [29]. We will apply our recent calibration methods on the large scale robot prototype developed for the Cogiro project at the very end of this year.

6.2.1.2.3. Cable properties

Quite often cable-driven robot analysis assume mass-less and non-elastic wires. We proposed a method based on interval analysis to judge the validity of this assumption for a particular robot in a specific workspace. Our aim is to use this method in order to determine a a region within the robot workspace for which the hypothesis is valid and consequently for which self calibration of the robot is possible. Indeed, the assumption on the cable properties is not acceptable over the full workspace of the large scale robot developed in the Cogiro project. Still a self-calibration is possible if calibration poses are chosen within a specific subpart of the workspace. A more efficient calibration approach is in progress with additional measures and a more complex model (static and elasticity). The results has been published in [18], [30].

6.2.1.3. Assistance robotics

Participants: David Daney, Claire Dune, Jean-Pierre Merlet [correspondant], Yves Papegay, Odile Pourtallier.

As mentioned earlier in the report we have started in 2008 a long term strategic move toward assistance robotics, with the objectives of providing low-cost, simple to control, robotized smart devices that may help disabled, elderly and handicapped people in their personal life, provide also assistance to family and caregivers while allowing doctors to get better and objective information on the health state of the end-user. Our credo is that these devices have to be adapted to the end-user and to its everyday environment (by contrast with the existing trend of focusing on a "universal" robot, to which the end-user and its environment have to adapt). As for cost reasons we intend to use only standard hardware adaptation has to be taken into account at the very early stage of the system design and uncertainties in the physical instances of our systems are also to be considered.

For validation purposes we have developed a flat in order to explore various full scale scenarii that cover a part of the daily life of an elderly, to develop specific assistance devices and to test them [1]. Our activity in this field is concentrated on transfer, manipulation, walking monitoring, rehabilitation and the use of virtual reality. We are also investigating how such complex environments with multiples smart agents, quite heterogeneous from a computing viewpoint, but that have to cooperate, may be programmed. All these topics are in accordance with the one of the large scale initiative PAL [2] of which we are an active member.

6.2.1.3.1. Transfer and manipulation

Participants: François Chaumette [Lagadic], Jean-Pierre Merlet, Rémy Ramadour.

For transfer operation we are using the `MARIONET-ASSIST` robot (see section 6.2.1.4) that is installed in our flat. Currently we use 4 winches with the wires connected at the same point on the platform, hence providing 3 translational degrees of freedom. This low-cost, low-intrusivity robot has proved to be very effective for transfer operation. Apart of transfer operation robot may be used for manipulation. Adding one or several low-cost cameras (the cost being here a fundamental constraint), visual-servoing control is used to provide a whole new set of useful services such as grasping objects in order to bring them to the end-user (if they are too heavy, too far, high or low), or cleaning the table after lunch. Using a parallel crane robot, we are able to cover a large workspace, the vision-control allowing us to obtain the precision required by the manipulation of daily-life objects. The collaborative implementation of the vision and the kinematic control of the robot gives us a way to make best use of the advantages of both parts, while overcoming their respective drawbacks.

Given a region where the object of interest belongs, the first step is to detect it in an evolutive environment. A segmentation is made, robust to luminance variations and perspective projections. The vision is then used to move the platform toward a desired position relatively to the target. In order to execute this task, some carefully chosen features are measured, allowing to estimate the incremental displacement required to move the end-effector to the desired place. We use the library ViSP for both the detection and the visual-control part.

[1] pictures of this assistive flat are available at http://www-sop.inria.fr/coprin/prototypes/main.html
[2] http://pal.inria.fr

Experimental results were obtained using a platform with 3 degrees of freedom and a single camera, grasping a single object and moving it from a place to another. We used for that basic image data such as 2D moments, allowing a fast computing and yet robustness in measurements. We currently address to generalize this manipulation to other configurations and to evaluate its robustness to calibration errors and other uncertainty sources. We also are looking for a global paradigm merging both the vision-based kinematic model and the mechanical one, which could significantly improve the efficiency of the experiment, while reducing the mathematical complexity behind each kinematic model considered on its own.

6.2.1.3.2. Walking monitoring

Participants: Claire Dune [Handibio], Jean-Pierre Merlet.

We use the walking aids `ANG-light` and `ANG-II` (see section 6.2.1.4) to monitor the trajectory of the walking aid. The on-board sensors of these aids allow to evaluate the step pattern, gait asymmetry,...during daily walking, hence providing an health monitoring system that is always available. `ANG-light` has been tested last year with 24 subjects that were themselves instrumented (accelerometers in the wrists and knees, pressure sensors in the shoes) and were asked to perform two different trajectories twice with/without the walking aid. The purposes were:

- to determine pertinent walking indicators
- to obtain a "gold" standard of these indicators for non pathological walking, taking into account the normal variability of the walking pattern
- to determine if indicators obtained with the walking aid may led to an accurate estimation of the indicators when the walking aid is not used

Several indicators have been determined after the analysis of these data. In a second phase the inclusion test of elderly people (30 subjects) is taking place at the CHU of Nice-Cimiez and will last until the first trimester of 2013. An analysis of the collected data, in close collaboration with the doctors, will allow to determine if the proposed walking indicators are pertinent.

Another interest of the walking aids `ANG` is that they allow to collect significant information for mobility during their daily use: slope and surface quality of the sidewalks and automatic detection of lowered kerbs with a ranking of their convenience. It will be interesting for a community to share such information that is collected by the community members. For that purpose we propose to use collective maps, such as OpenStreeMap, which allow for map annotation. To validate this concept we have used `ANG-light` to automatically annotate the map of the Inria Sophia site with pertinent information for walking aid and wheelchair users [3].

6.2.1.3.3. Rehabilitation

Participants: David Daney, Mandar Harshe, Sami Bennour, Jean-Pierre Merlet [correspondant].

The focus of our work is on analyzing knee joint motion during a walking activity. The main principle of the system is to observe relative motions of the collars attached to tibia and femur. The measurement of the motion of these collars is based on the wire actuated parallel robot architecture (using the `MARIONET-REHAB` robot, see section 6.2.1.4). To increase the reliability of our analysis, and decrease the influence of Skin Tissue Artifacts (STA), we also incorporate a passive wire measurement system, IR camera based motion capture system, accelerometers, and force sensors to measure human motions.

Measurements in the global frame and collar specific local frames give precise data to reconstruct collar (and thus, knee joint) motion. The system developed already incorporates the optical motion capture, inertial measurement units and the wire sensors for comprehensive coordinated measurements of the motion of the knee.We have performed preliminary trials on three subjects for walking motion.

In the past year, we worked on processing the data to obtain pose and orientation information of knee joint. Data obtained from the trials was analyzed and post-processing steps were implemented to reduce noise and errors. In order to perform sensor fusion, we implemented a probabilistic estimation based method to estimate the pose.

[3] see http://www-sop.inria.fr/coprin/prototypes/main.html#ang

The results from these analysis have allowed us to identify the merits of our approach and also helped us identify improvements that are needed. We have also identified the possible changes to our mathematical model that could allow use of interval analysis tools along with probabilistic estimation methods. We have identified the changes needed to the hardware setup that will help reduce the sensor noise and error. These changes once implemented will allow us to improve the usability of the system and also point us towards newer areas for further investigation, including, for example, effect of sensor placement, collar design, and interval based extended Kalman Filters for pose estimation[8], [9].

6.2.1.3.4. Virtual reality

Virtual reality has proved to be an effective mean for dealing with rehabilitation, provided that motion is added to the 3D visual feedback. The MARIONET-VR robot, together with our motion base (see section 6.2.1.4) may provide very realistic motion in an immersive room and we will start in 2013 a collaboration with the VR4I and REVES project-teams on this issue. A first task will be to install a moving walking treadmill in the immersive room at Sophia and to combine motion of the treadmill with 3D viewing.

6.2.1.3.5. Programming

In our opinion there will not be a single assistance device that will be able to offer all the required help that may be needed but rather numerous redundant smart agents that are able to perform very efficiently (at low cost and with a small energy consumption) a set of specific tasks. Such agents must be able to communicate and possibly will have to cooperate in some cases (e.g. after a fall). They will be heterogeneous from a computer view point as the used agents will change according to technological advances and to the trajectory of life of the end-users. If the manual programming of a single agent is possible (although quite complex for some of them) the overall system cannot be managed in that way: we need an unifying framework for this development. For that purpose we are currently investigating the use of HOP, a multi-tier programming language for the diffuse Web developed in the INDES project-team. Already one of our wire-driven parallel robot MARIONET-SCHOOL has been programmed in HOP and can be seen as a web resource.

6.2.1.4. Prototypes

Participants: Julien Alexandre Dit Sandretto, David Daney, Claire Dune, Jean-Pierre Merlet [correspondant].

Experimental works are a key point in the field of service robotics: it allows for validating concepts, getting feedback from the end-users and discovering new problems. We have extensively developed prototypes this year [4]

6.2.1.4.1. Wire-driven parallel robots

The MARIONET family is now constituted of

- MARIONET-REHAB (2004-): using up to 7 linear actuators it is mainly used for rehabilitation and health monitoring although it is also a very fast pick-and place robot
- MARIONET-CRANE (2008-): a very large 6-dof rescue crane, portable and autonomous that can be deployed in 10 minutes and has a lifting capability of 2 tons
- MARIONET-ASSIST (2011-): a robot deployed in our flat with up to 6 winches, that is used for transfer operation and health monitoring, with a lifting ability of about 2000 kg
- MARIONET-VR (2011-): using up to 6 linear actuators it will be used in the Sophia immersive room for simulation and rehabilitation. It allows to fully lift a person and has been physically installed in the immersive room this year although it is not fully operational
- MARIONET-SCHOOL (2012-): a set of very low-cost robots that are intended to be used for dissemination. They fit in a small suitcase and can be deployed on a table or over a full classroom. We believe that such robots may be used by roboticists but also by researchers from other domains that may use the motion of the robot(s) to illustrate scientific concepts. Currently we have 3 such robots (one using Lego components, one with step motors and one with servos)

[4] pictures and videos of our prototypes are available at http://www-sop.inria.fr/coprin/prototypes/main.html or on our YouTube channel http://www.youtube.com/user/CoprinTeam/videos?flow=grid&view=0

6.2.1.4.2. Walking aids

The *Assistive Navigation Guide* (ANG) family is based on commercially available Rollators. with several objectives (we mention only a few of them):

- fall prevention/detection: fall is a major problem for elderly
- mobility help: provide an on-demand mobility help
- gait pattern monitoring: we believe that being able to monitor the trajectory of the walking aid will provide useful information on the gait pattern of the user

For reaching these objectives we have developed two walking aids:

- `ANG-light`: a walking aid with encoders in the wheels, 3D accelerometer, gyrometer and GPS. These sensors allow to measure the trajectory of the walking aid and several features of the user's gait. This walking aid is currently being used at CHU Nice, see section 6.2.1.3.2 . A replica of `ANG-light` is currently under development at the Handibio laboratory of Toulon and will include force sensors in the handles to get measurement of muscular activities while walking.
- `ANG-II`: this aid is an evolution of the motorized walker `ANG`, with a lower weight and better integration

6.2.1.4.3. Other devices

As seen in the previous sections we have focused our work on mobility as it has been identified as a major priority during our two years interview period. Another priority is fall management: it is adressed with the `ANG`'s) but requires that the patient uses a walking aid. To obtain a better coverage we have developed an instrumented vest that includes an Arduino Lilypad microcontroller to monitor a 3D accelerometer and verticality sensors in order to detect a fall. This system may communicate an alarm through wifi, zigbee or even infrared signals. Apart of its low cost the system is washable and has a very low power consumption (we are considering energy harvesting to further increase its energy autonomy).

Two other important issues for assistance robotics is activities monitoring and patient localization. In cooperation with the STARS project-team an activity monitoring system within the Dem@care project. Its purpose is to log specif events such as taking a pen, setting on a kettle, ...using only simple sensors such as proximity and distance sensors, switches, that are more reliable and less complex than using a vision system. We have provided to the CHU a system that allows to monitor up to 208 different events. In the same manner we are working on a localization system of elderly in a room based only on distance sensors, that will be less intrusive and/or may complement the measurements of a vision system.

Rehabilitation is also part of our activities. Apart of the `MARIONET-REHAB` system we plan to investigate the use of Sophia immersive room for that purpose. To complement the motion that will be provided by the `MARIONET-VR` robot we have bought a 6-dof motion base from Servos with a nominal load of 150kg, that we have modified to accommodate our needs. We will use also two lifting columns with a load of 100 kg each, that will allow to manage motion for rehabilitation apparatus such as treadmill.

6.3. Miscellaneous results

6.3.1. Symbolic tools for modeling and simulation

Participant: Yves Papegay.

This activity is the main part of a long-term ongoing collaboration with Airbus whose goal is to directly translate the conceptual work of aeronautics engineers into digital simulators to accelerate aircraft design.

An extensive modeling and simulation platform has been designed which includes a dedicated modeling language for the description of aircraft dynamics models in term of formulae and algorithms, and a symbolic compiler producing as target an efficient numerical simulation code ready to be plugged into a flight simulator, as well as a formatted documentation compliant with industrial requirements of corporate memory [10].

Implementation of this platform is a modeling and simulation environment based on symbolic computation tools. It contains several components :

- a model editor, that makes it possible and easy to enter the whole set of equations describing large and complex industrial models,
- an highly interactive and modular evaluation workbench allowing to simulate the models and to visualize the results inside the modeling environment with the benefits for the designer of being able to directly use all its computational functionnalities.
- a C code generator which, using these models, automatically generates the numerical real-time simulation engines
- a technical documentation generator

To finalize the transfer of the technology demonstrated by our prototype to our industrial partner, an extensive collection of testing and corresponding improvements have been done in 2012. This step has delayed the delivery of the final version of our modeling and simulation environment to Airbus until November 2012.

6.3.2. Multi-agent aircraft design

Participant: Yves Papegay.

The modeling environment described in the previous section is used, in collaboration with other teams at Airbus, in the framework of the ID4CS project founded by ANR and dedicated to multi-agent optimization of large scale system.

Several models of aircraft engines and of aircraft have been developed as user cases for the project.

In 2012 we focused on automatic generation of agent code based on models that is ready for integration in the ID4CS platform prototype.

6.3.2.1. Equilibrium strategies for linked Electricity and CO2 markets

Participant: Odile Pourtallier.

In collaboration with M. Bossy (Inria -TOSCA Team) and N. Maïzi (CMA - Mines Paristech) O. Pourtallier has pursued the study of equilibrium model for coupled electricity and CO2 allowance exchange markets (see also Section 7.3). We have mainly focused on the determination of Nash equilibrium for the coupled electricity and carbon markets with the assumption that the producers maximize their market shares. Nash equilibrium have been obtained by using as an intermediary step a decreasing auction mechanism.

We have also pursued an indifference pricing methodology which is presented in more details in Inria -TOSCA Team section.

E-MOTION Project-Team

5. New Results

5.1. Perception and Situation Awareness in Dynamic Environments

5.1.1. Sensor Fusion for state parameters identification

Participants: Agostino Martinelli, Chiara Troiani.

5.1.1.1. Problem adressed and background

The general framework based on the new concept of continuous symmetry developed during the last two years (see [67] for a detailed description of this framework) has been extensively applied to investigate the visual inertial structure from motion problem. This problem was already considered in 2011. During 2012 more general results have been found. Special attention has been devoted to identify the conditions under which the problem has a finite number of solutions. Specifically, it has been shown that the problem can have a unique solution, two distinct solutions and infinite solutions depending on the trajectory, on the number of point-features and on their layout and on the number of camera images. The investigation has also performed in the case when the inertial data are biased, showing that, in this latter case, more images and more restrictive conditions on the trajectory are required for the problem resolvability.

5.1.1.2. Theorical results

The new results have been published on the journal of Transaction on Robotics [68], in a technical report [43] and submitted to the International Journal of Computer Vision. We have also considered the case of structured light. Specifically, we have considered a sensor assembling (from now on *aerial vehicle*) consisting of a monocular camera and inertial sensors. Additionally, a laser pointer is mounted on the aerial vehicle and it produces a laser spot. The laser spot is observed by the monocular camera and it is the unique point feature used in the proposed approach. We focus our attention to the case when the aerial vehicle moves in proximity of a planar surface and in particular when the laser spot belongs to this surface. We introduced two novel contributions. The former is the analytical derivation of all the observable modes, i.e., all the physical quantities that can be determined by only using the inertial data and the camera observations of the laser spot during a short time-interval. This derivation was based on the framework introduced in [67]. Specifically, it is shown that the observable modes are: the distance of the vehicle from the planar surface; the component of the vehicle speed, which is orthogonal to the planar surface; the relative orientation of the vehicle with respect to the planar surface; the orientation of the planar surface with respect to the gravity. The second contribution is the introduction of a simple recursive method to perform the estimation of all the aforementioned observable modes. This method is based on a local decomposition of the original system, which separates the observable modes from the rest of the system. The method has been validated by using synthetic data. Additionally, preliminary tests with real data are provided and more complete experiments are in progress. The presented approach can be integrated in the framework of autonomous take-off and landing, safe touch-down and low altitude manoeuvres even in dark or featureless environment. These results have been published in the iros conference [31]

5.1.1.3. Experimental results

In parallel to this theoretical activity an experimental activity has been carried out in order to deploy our technologies to industrial partners. To this regard, we had a collaboration with the company Delta Drone in Grenoble. In this framework we introduced a new method to localize a micro aerial vehicle (MAV) in GPS denied environments and without the usage of any known pattern. The method exploits the planar ground assumption and only uses the data provided by a monocular camera and an inertial measurement unit. It is based on a closed solution which provides the vehicle pose from a single camera image, once the roll and the pitch angles are obtained by the inertial measurements. Specifically, the vehicle position and attitude can uniquely be determined by having two point features. However, the precision is significantly improved by

using three point features. The closed form solution makes the method very simple in terms of computational cost and therefore very suitable for real time implementation. Additionally, because of this closed solution, the method does not need any initialization. We have implemented this method on the platform available in our lab. This is a *Pelican* from *Ascending Technologies* equipped with an Intel Atom processor board (*1.6 GHz, 1 GB RAM*) (figure 1).

Figure 1. AscTec Pelican quadcopter equipped with a monocular camera.

Our sensor suite consists of an Inertial Measurement Unit (*3-Axis Gyro, 3-Axis Accelerometer*) belonging to the Flight Control Unit (FCU) "AscTec Autopilot" , and a monocular camera (*Matrix Vision mvBlueFOX*, FOV: 130 deg). The camera is calibrated using the Camera Calibration Toolbox for Matlab by J.Y. Bouguet at caltech. The calibration between IMU and camera has been performed using the Inertial Measurement Unit and Camera Calibration Toolbox in [66]. The IMU provides measurements update at a rate of $100Hz$, while the camera framerate is $10Hz$. The Low Level Processor (LLP) of our Pelican is flashed with the *2012 LLP Firmware* and performs attitude data fusion and attitude control. We flashed the High Level Processor (HLP) with the asctec_hl_firmware [48]. The onboard computer runs linux 10.04 and ROS (Robot Operating System). We implemented our method using ROS as a middleware for communication and monitoring . The HLP communicates with the onboard computer through a FCU-ROS node. The communication between the camera and the onboard computer is achieved by a ROS node as well. The presented algorithms are running online and onboard at $10Hz$.

The scenario setup is shown in figure 3 . Since our lab is not yet equipped with a Motion Capture System, we used an ARToolKit Marker with the only aim of having a ground truth to evaluate the performance of our approach. The estimation of the camera pose provided by the marker is not used to perform the estimation. The marker is positioned such that it's reference frame is coincident with the configuration shown in figure 3 . The three features considered are the center of the three little balls in figure 3 . The use of three blob markers instead of natural features is only related to the need to get a ground truth. The information related to the pattern composed by the 3 features ($D = 0.25m$, $\gamma_1 = 60deg$, $\gamma 2 = 120deg$) is only used to evaluate the performance of our approach. The algorithm does not require any information about the features configuration.

Figure 4 and 5 show respectively the position and the attitude by using the proposed approach. The estimated values are compared with the ground truth obtained with the ARToolkit marker. From figure 4 we see that the difference between our estimates and the ground truth values is of the order of $2cm$ for x and y and less than $0.5cm$ for z. From figure 5 we see that the difference between our estimates and the ground truth values is of the order of $2deg$ for $Pitch$ and less than $0.5deg$ for $Roll$ and Yaw.

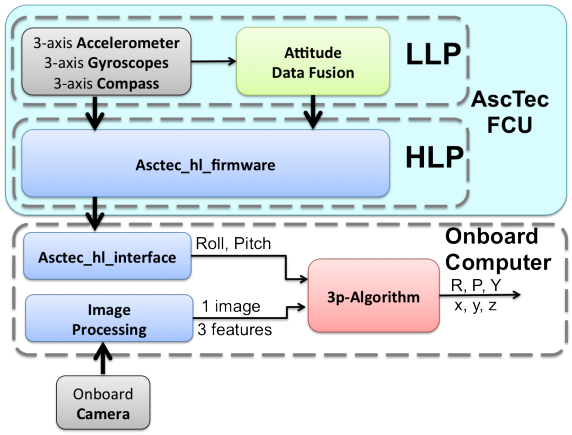


Figure 2. Our Pelican quadcopter: a system overview

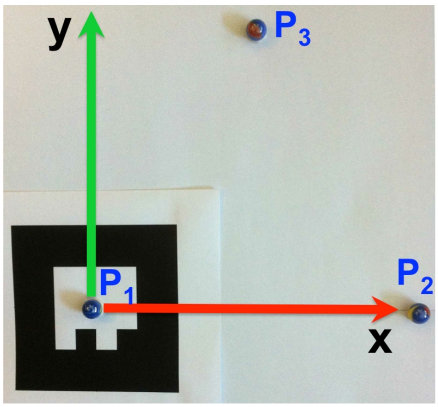


Figure 3. Scenario: The AR Marker and the 3 balls are used only with the aim to get a rough ground truth. The AR Marker provides the camera $6DOF$ pose in a global reference frame according to our conventions.

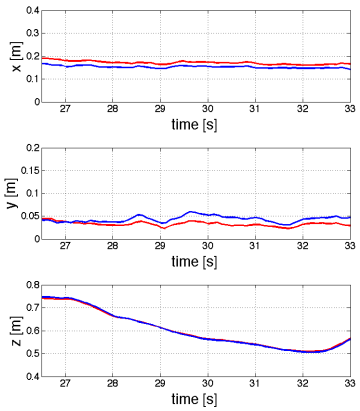


Figure 4. Estimated position, respectively x, y, z. The red lines represent the estimated values with the 3p-Algorithm, the blue ones represent a rough ground truth (from ARToolkit Marker).

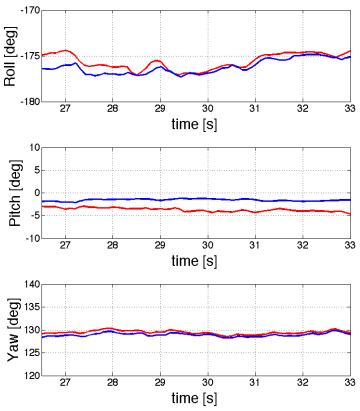


Figure 5. Estimated attitude, respectively $Roll$, $Pitch$, Yaw. The red lines represent the estimated values with the 3p-Algorithm, the blue ones represent a rough ground truth (from ARToolkit Marker).

We believe that the main source of error is due to the distortion of the lens, which is not fully compensated by the calibration. Note that this distortion also affects our ground truth. We plan to test our approach in an environment equipped with a Motion Capture System.

This method is currently under evaluation to be patented.

5.1.2. Visual recognition for intelligent vehicles

Participants: Alexandros Makris, Mathias Perrollaz, Christian Laugier.

We developed a generic object class recognition method. The state-of-the-art visual object class recognition systems operate with local descriptors and codebook representation of the objects. Various local features (e.g., gradient maps, edges) are used to create the descriptors. Then kernel based classifiers are commonly employed to classify the detected features in one of several object classes [50] [54]. The recognition of vehicles or pedestrians from sensors mounted on a moving platform is achieved by different approaches using various types of sensors, e.g., stereo camera, laser [61] [52]. The approaches that perform data fusion from various sensors have proven to be the more robust in a variety of road conditions [76].

Our work focuses on the development of an object class recognition system which follows the part based detection approach [65]. The system fuses intensity and depth information in a probabilistic framework. To train the system for a specific object class, a database of annotated with bounding boxes images of the class objects is required. Therefore, extending the system to recognize different object classes is straightforward. We apply our method to the problem of detecting vehicles by means of on-board sensors. Initially, depth information is used to find regions of interest. Additionally, the depth of each local feature is used to weight its contribution to the posterior of the object position in the corresponding scale. The votes are then accumulated in a 3d space-scale space and the possible detections are the local maxima in that space.

The novelty of our approach is the fusion of depth and intensity information to form a probabilistic part-based detector. Using depth information is beneficial for the robustness of the approach, because we avoid including many noisy detections resulting from false matches between features of different scales. The method is tested with stereo video sequences captured in an urban environment. Figure 6 shows some example detections. The proposed method detects cars in various scales, in cases with partial occlusions, and under significant background clutter.

Figure 6. Car detection examples. The new weighting strategy allows to better detect the partially occulded objects.

In 2012, we worked on two particular improvements of the method. First, we modified the weighting strategy in order to increase the detection of partially occulded objects. This approach effectively improves the detection results. Second, we consider replacing the current depth descriptor, which only integrates depth information, with a more advanced depth descriptor (e.g., the NARF descriptor). This work is still in progress, in collaboration with Dimitrios Kanoulas, PhD student in Northeastern University (USA).

In 2012, the full method for objects recognition has been submitted for publication in IEEE Transactions on Intelligent Transportation Systems.

5.1.3. Bayesian Motion Detection in Dynamic Environments

Participants: Qadeer Baig, Jander Perrollaz, Mathias Botelho, Christian Laugier.

5.1.3.1. Introduction

Bayesian Occupancy Filter (BOF) [51] is a grid based perception framework that we use for environment monitoring. In this representation this framework estimates the probability of occupancy as well as velocity of each cell of this grid using sensor data. Output of this framework is used by Fast Clustering Tracking Algorithm (FCTA) [69] to cluster objects and to track them. An important point is that BOF estimates cell velocities without motion information of the ego vehicle, so these are relative velocities. Since no motion information are used, the static objects observed from the moving ego vehicle are also tracked, this results into many false moving objects. Although many of these false positives can be removed by tuning parameters of FCTA, however, this usually is a time consuming task. We note that the number of false can be reduced as well as dependence on FCTA parameters can be relaxed if we can separate the input to BOF into static and dynamic parts. Adding these motion information with cells will allow BOF to calculate velocity information for moving cells only and FCTA will also ignore the static cells while clustering step resulting into faster calculations and better track. In this context we have developed a very fast motion detection technique to separate BOF input into static and dynamic parts. The integration of this module with BOF and FCTA has helped us remove about 78% of the false positives. This technique is summarized next.

5.1.3.2. Fast Motion Detection

In this section we summarize the technique that we have developed to find moving parts of the environment. This motion detection module is situated in the processing chain just before the BOF. The input to this module consists of an occupancy grid generated by the fusion module. And the output of this module is used by both BOF and FCTA modules.

The objective of this module is to separate the input occupancy grid into two parts: cells belonging to static objects and cells belonging to moving objects. The main idea of this separation between static and dynamic parts, consists of keeping a track of how many times a cell is observed as free and how many times it is observed as moving. However to realize this concept we must solve the localization problem. We solve this problem using velocity and rotation information given by MTi-G XSens unit. This allows us to map cells between two input grids OG_{t-1} and OG_t at time $t-1$ and t as shown in figure 7 .

We use two sets of $Free$ and $Occupied$ counter arrays. One set is initialized from new input grid at time t whereas other set keeps updated counts until time $t-1$. Then after above transformation betweens cells of grids OG_{t-1} and OG_t newly initialized set of arrays is updated from arrays at time $t-1$, resulting in incremented counts for overlapping areas between two grids. Finally following decision function is used to separate cells of current input grid OG_t into static and dynamic parts and results are stored in a motion grid.

$$MotionGrid_t[i] = \begin{cases} 1, & OG_t[i] > 0.5 \text{ and} \\ & FreeCount_t[i] > 2 * OccupiedCount_t[i] \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

This technique being simple is quite robust and efficient and does not oblige us to solve the complete SLAM problem. This work is published as [19] and [20].

5.1.3.3. Integration within the BOF framework

We have updated the BOF implementation to take into account the motion detection results. The motion grid is used as an input for updating the BOF. If the input motion grid tells that a cell belongs to a static object, then during prediction and update cycles of BOF the cell's velocity distribution over the velocity range is set to uniform for all discrete velocity values. This essentially means that no velocity information for a given cell

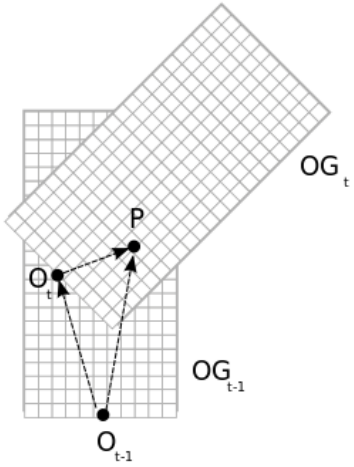


Figure 7. Position of the grid at time instants $t-1$ and t. Vehicle undergoes a motion of $u_t = (\nu_t, \omega_t)$ to move from O_{t-1} to O_t. We need to find the position of point P of grid OG_{t-1} in grid OG_t.

is available and the cell is labeled as static in the current BOF implementation. However, if the cell has been detected as belonging to a moving object, then the velocity distribution prediction and the update cycle are carried out normally. In formal terms this change in the parametric form of dynamic model can be stated as:

$$P(A_i^t|A_i^{t-1}) = \begin{cases} (1-\epsilon)P(A^{t-1}_{A_i^{t-1}}) + \epsilon/\|A_i\| \\ \quad \text{if } MotionGrid_t[i] > 0 \\ \\ 1/\|A_i\| \quad \text{otherwise} \end{cases}$$

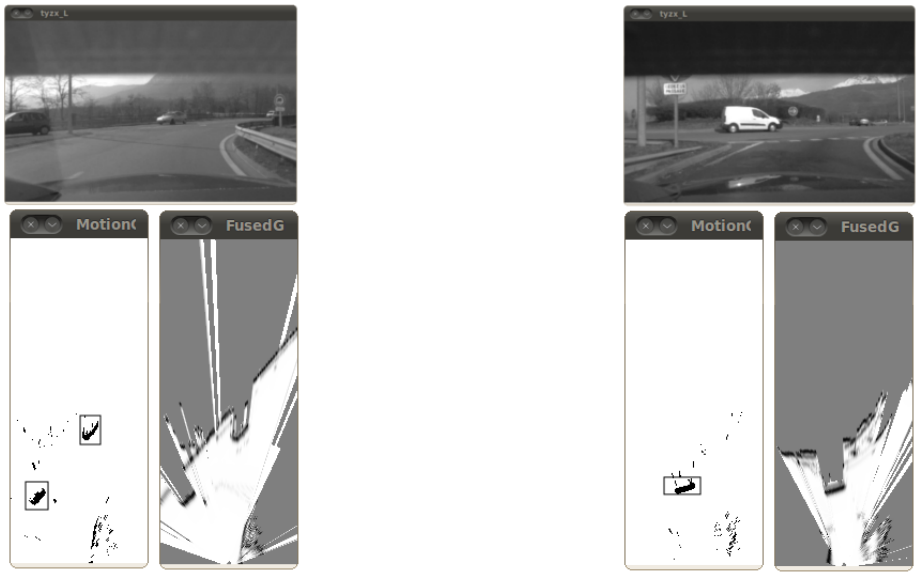
where A_i^t is the set of antecedents of cell i at time t and ϵ is a parameter of BOF, modelling the prediction error probability.

5.1.3.4. Integration with FCTA

We have also updated the FCTA implementation to take into account the motion detection results. The cells which do not possess the velocity information are now ignored during the clustering step. While generally most of the areas belong to static objects and are detected as static by the motion detection module, two main advantages are expected from this strategy: (i) the clustering stage of the algorithm is highly accelerated by the reduction of hypotheses, and (ii) the false moving clusters are ignored because they are not considered for clustering, even with the relaxed FCTA parameters.

5.1.3.5. Results

Some qualitative results of motion detection module are shown in figure 8 , (rectangles around the objects are drawn manually to highlight them). As expected, the moving objects are properly detected. For example, figure 8 (left) shows the motion detection scenario of two cars, and the car moving around a roundabout has been successfully detected in figure 8 (right). Some noise is also visible on the results, mainly due to two causes: first, the uncertainty on the IMU measurements along with the circular motion model may result in some errors in the estimation of the motion; second, the decision function is too rough for taking correct decisions in every situation. The results would benefit from replacing this function by a probabilistic model.



a) b)//

Figure 8. Left: Motion detection results of two cars. Top, scenario, bottom right input fused grid, bottom left resulting motion grid. Right: Motion detection results of a car on a roundabout. Some noise due to sensor uncertainty is also visible

The tracking results of FCTA are highly sensitive to its parameters values. There are less false positives when strict parameters (large thresholds) are used, however, a large number of the true tracks may be missed, resulting in numerous miss detections -note that since the focus of this work is to detect moving objects, we consider in this part that detections belonging to the static environment are false alarms-. The relaxed parameters (small thresholds) provide less miss detections, however, a large number of false tracks are detected. While finding the appropriate set of parameters can be a challenging task, our implementation of the motion detection module with relaxed parameters represents a trade-off.
The following statistics with a dataset duration of about 13 minutes give an insight into the improvements gained with this implementation. When the motion detection module is not used, 22303 tracks are detected. The activation of the motion detection module with all other parameters being equal provides to detect 4796 tracks. This example shows the advantage of the motion detection module because it allows us to remove most of the false tracks while leaving most of the true tracks. Some qualitative FCTA tracking results with and without motion detection module activated (with all other parameters being same) are shown in figures 9 and 10 . Red rectangles are the detected tracks by FCTA in the shown scenario. We clearly see that most of the false positives have been removed.

5.1.3.6. Conclusion

In this section we have presented a fast technique to find moving objects from laser data and its integration with Bayesian Occupancy Filter (BOF) and Fast Clustering-Tracking Algorithm (FCTA). We have seen that after this integration we were able to remove a significant number of false alarms, this has also relaxed the dependence of results on the FCTA parameters.

We plan to change the rather ad hoc decision module that is currently based on occupied and free counter values to a more formal probabilistic function that also takes into account the uncertainty effects on the neighboring cells to accommodate the localization errors. We are also working on extending the tracking module from single motion mode to multiple motion modes.

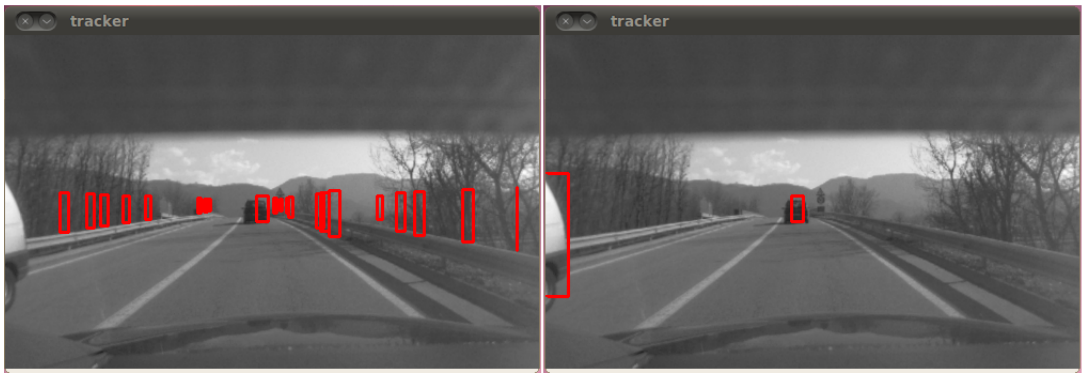


Figure 9. Tracking results of a car. Left, FCTA results without motion detection module activated. Right, same scenario but with motion detection module activated.

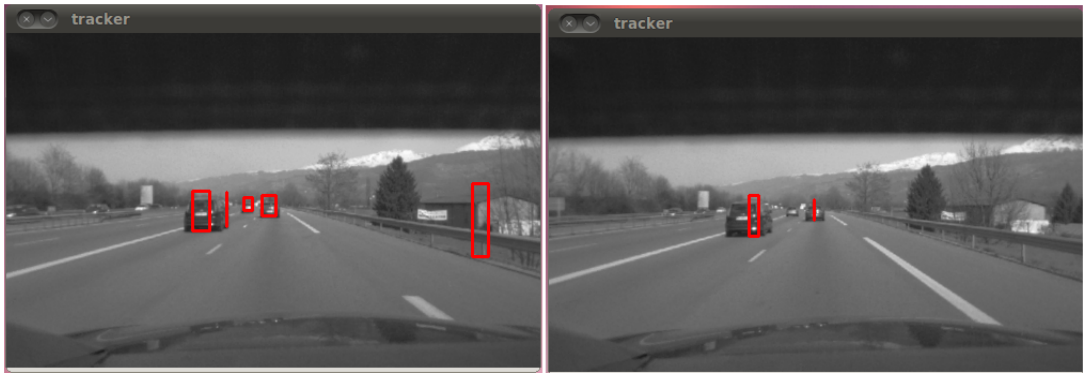


Figure 10. Tracking results of two cars on highway. Left, FCTA results without motion detection module activated. Right, same scenario but with motion detection module activated.

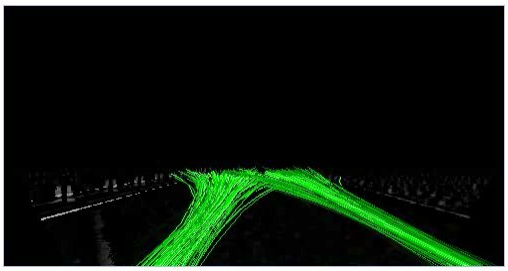
5.1.4. Vision-based Lane Tracker

Participants: Mathias Perrollaz, Amaury Nègre.

For perception in road structured environment the detection of the lane markers and its localization provide an interesting information to predict drivers behaviors and to evaluate collision risks. We currently develop a real time road lane detection and tracking application using camera's image information. The tracking application estimates simultaneously the road plane orientation, the lane curvature and the camera position by using a Monte-Carlo particle filter. With this method, the parameter distribution is represented by a set of particles (see Fig 11 .a) that are sequentially updated using the vehicle dynamic model, evaluated by a ridge extraction (Fig 11 .b) and sampled considering the evaluation result. The average of the particles, displayed on Fig 11 .c) provides a good estimation of the lane state.

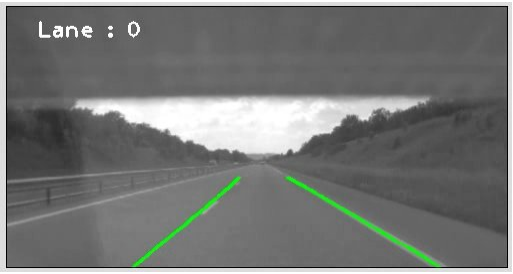
To obtain real-time performance, we implemented the whole process on GPU using the nVidia Cuda toolkit.

The output of this application has been mainly used to predict lane change behaviour 5.2.1 and to risk estimation applications.



(a) Particles cloud

(b) Ridges extraction



(c) Lane state estimation

Figure 11. Visual Particle based lane tracking. (a) The Lane state is estimated by a particles set which is recursively updated, evaluated and resampled. (b) A ridge image is compute to estimate each particle. (c) The average of the particle state provides a good estimation of the lane.

5.1.5. Experimental platform for road perception

Participants: Nicolas Vignard, Mathias Perrollaz, Amaury Nègre.

5.1.5.1. Experimental platform material description

Our experimental platform is a Lexus LS600h car shown in Figure 12 . The vehicle is equipped with a variety of sensors including two IBEO Lux lidars placed toward the edges of the front bumper, a TYZX stereo camera situated behind the windshield, and an Xsens MTi-G inertial sensor with GPS.

Figure 12. Lexus LS600h car equipped with two IBEO Lux lidars, a TYZX stereo camera, and a n Xsens MTi-G inertial sensor with GPS.

The stereo camera baseline is 22 cm, with a field of view of 62°. Camera resolution is 512x320 pixels with a focal length of 410 pixels. Each lidar provides four layers of up to 200 impacts with a sampling period of 20 ms. The angular range is 100°, and the angular resolution is 0.5°. The on-board computer is equipped with 8GB of RAM, an Intel Xeon 3.4 GHz processor and an NVIDIA GeForce GTX 480 for GPU. IMU data contains accelerations, velocity, GPS position and steering angle. The experiments are conducted in various road environements (country roads, downtown and highway), at different time of the day, with various driving situations (light traffic, dense traffic, traffic jams). The datasets are acquired online and are used for testing of our sensor fusion and risk assessment algorithms.

5.1.5.2. Migration from Hugr to ROS middleware

Our platform described in 5.1.5.1 previously used a middleware named Hugr. Middlewares bring an abstraction layer between the sensors drivers and the processing modules. We also used this middleware to share information with modules and applications. Using a middleware facilitates and normalises the communication between modules.

Hugr has been developed by inria for the Cycab project and a team was built to add functionalities and maintain this new middleware. However, now the team has to work on other projects and it is becoming increasingly difficult to allocate resources to maintain this middleware. Given this and some other technical issues [49], we have decided to change our robotic middlware.

We find that many different middleware (AROCAM, RTMaps, ROS,$\cdots$) are being used in the robotic community [53]. Among these, Robotic Operating System (ROS) is increasingly becoming a research standard in robotics. The reason being: an important community, a lot of tools and sharing work and development. The primary goal of ROS is to develop faster robotics applications. However, before moving to ROS we also did an extensive research on the comparison between Hugr and ROS [49], that supported our this decision.

Because of this middleware change, we had to reimplement all the perception process from drivers to applications. In this regard, we have implemented the following drivers:

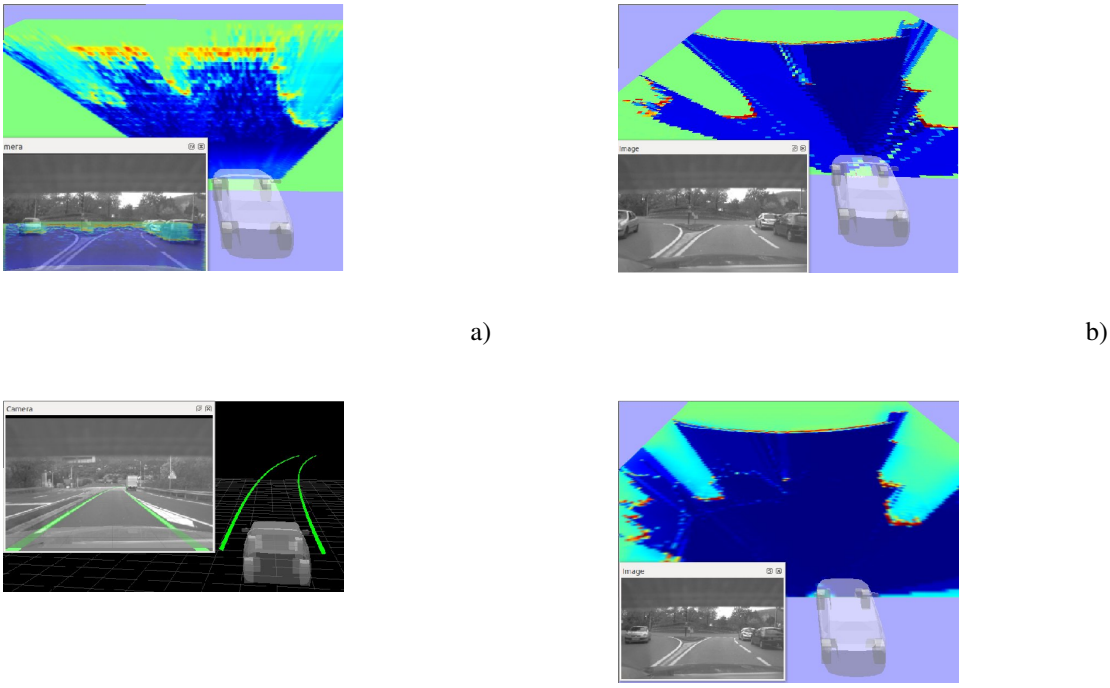
- the IBEO Lux lidar
- the TYZX camera
- the CAN bus
- the Xsens MTi-G (inertial sensor with GPS)

However for the Xsens MTi-G, we found an existing driver that we modified to add the GPS functionality http://www.ros.org/wiki/lse_xsens_mti.

Furthermore, we have also migrated the following modules:

- a module that fuses lidar data into an occupancy grid
- a module that generates occupancy grid from the stereo camera
- the Bayesian Occupancy filter (BOF) module
- the lane tracker

Some result images of occupancy grids and data from the lane tracker after this migration to ROS are shown below 13 . Finally, we have created a public repository at http://gforge.inria.fr that share our developments (both drivers and modules).



a) b)

c) d)

Figure 13. a) occupancy grid from the stereo camera. b) occupancy grid from the lidar. d) lanes detected by lane tracker. e) occupancy grid from the BOF.

5.1.5.3. Disparity space approach for a vision based occupancy grid

Participants: Mathias Perrollaz, Anne Spalanzani, John-David Yoder, Amaury Nègre, Christian Laugier.

To use sensors in the BOF framework, it is essential to develop an associated probabilistic sensor model that takes into consideration the uncertainty over measurements. In 2009, we proposed such a sensor model for stereo-vision [72]. The originality of the approach relied on the decision to work in the disparity space, instead of the classical Cartesian space. In 2010, we improved our sensor model, in order to mimic some features of the sensor models used for range finders. Particularily, we worked on managing visible/occluded areas of the scene [74], and on including the information from the road/obstacle segmentation of the disparity image [73]. Our approach was also designed to allows highly parralel computation of the occupancy grid. A. Nègre implemented the approach on GPU using NVIDIA CUDA to enhance the performance. The complete

processing of the stereo data can now be done in 6 ms, while more than 150 ms were necessary with the CPU implementation. The complete approach for occupancy grid computation using stereovision has been publish in 2012, in [13].

5.1.6. Software and Hardware Integration for Embedded Bayesian Perception

Participants: Mathias Perrollaz, Christian Laugier, Qadeer Baig, Dizan Vasquez.

The objective of this recently started research work is to re-design in a highly parallel fashion our Bayesian Perception approach for dynamic environments (based on the BOF concept), in order to deeply integrate the software components into new multi-processor hardware boards. The goal is to miniaturize the software/hardware perception system (i.e., to reduce the size, the load, the energy consumption and the cost, while increasing the efficiency of the system).

This work has been started in 2012 in cooperation with CEA-LETI DACLE laboratory. During 2012, we have worked on the definition of the software/hardware architecture and we have started to re-think some components of the lower layer of the BOF software module.

The work plan has been split in two three-year-long phases, respectively leading to address a first level of integration based on mobile technologies, and a second level of integration, based on a more dedicated hardware architecture (and maybe to a SOC).

Two cooperative projects have been prepared and submitted this year for supporting this promising research: the “Permobile” project (FUI), involving industrial companies and user, and the “Perfect” project (IRT-Nano) involving the CEA-LET LIALP lab and ST-Microelectronics. Permobile is focusing on the first integration objectives (3 years) and has been recently submitted. Perfect is focusing onto the second integration objectives (6 years) and the development of integrated open platforms in the domain of transportation (vehicle and infrastructure) and in a second step in the domain of health sector (mobility of elderly and handicapped people, monitoring of elderly people at home...).

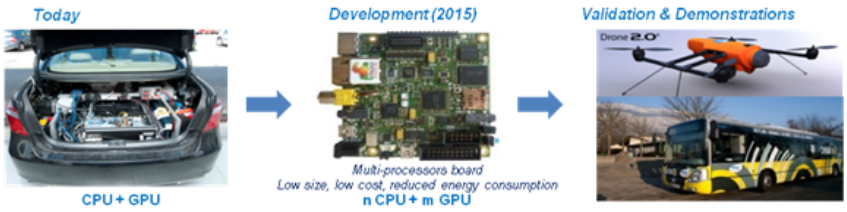


Figure 14. First objective for software/hardware of the BOF: developing and using multiple processor board from mobile technologies. The approach will be validated with real demonstrators.

5.2. Dynamic Change Prediction and Situation Awareness

5.2.1. Vision-based Lane Change Prediction

Participants: Puneet Kumar, Mathias Perrollaz, Stephanie Lefevre, Amaury Nègre, Maiwen Gault.

Predicting driver’s behaviors is a key component for future Advanced Driver Assistance Systems (ADAS). In 2012, we have proposed a novel approach for lane change prediction, using only information from a vision sensor embedded into the car. The idea is to predict in advance if our vehicle is about to change lane. Then this information can be used to properly help the driver, for instance by detecting inconsistencies with the turn lights signals.

As an input, the method uses visual data from a camera embedded into the car. A multiple-size ridge filter is used to extract low level features from the image (white markings on black road). Then road lanes are estimated and tracked over time using a particle filter. This process allows parallel computing, and thus works in real time on GPU.

Figure 15. Vision-based tracking of the road markings. From left to right: particles generated by the particle filter, low level features extracted using the ridge filter, and estimated lane.

The road markings are used to estimate the position and heading angle of our car with respect to the lane, as well as the derivatives of these variables. This information is then used as a vector of features for a classifier. The used classifier is a multi-class Support Vector Machine (SVM). The three possible classes are "no lane change" (NL), "right lane change" (CR) and "left lane change" (CL). The classifier has been trained using real data of 180 lane changes on highway, manually annotated. The output of the classification is then converted into a set of probabilities using a generalized Bradley-Terry model.

The classifier provides a very short term classification, which can contain many errors. The longer term integration of the time information is obtained by feeding the classification results into a Bayesian Filter (BF). The posterior output of the filter provides the probability distribution over possible behaviors (NL, CR, CL), hence providing the lane change prediction.

Real-world data from our vehicle is used for the purpose of training and testing. Data from different drivers on different highways were used for the robustness evaluation of the overall approach. The proposed method show promising results, because it is able to predict driver's intention to change lane 1.3 seconds (average) in advance, with maximum prediction horizon of 3.29 seconds. We are now working on a real time implementation of this approach, to demonstrate its use on real situations (e.g., for warning the driver while driving on the highway).

5.2.2. Risk estimation at road intersections for connected vehicle safety applications

Participants: Stéphanie Lefèvre, Christian Laugier.

Intersections are the most complex and dangerous areas of the road network. Statistics show that most road intersection accidents are caused by driver error and that many of them could be avoided through the use of Advanced Driver Assistance Systems. In this respect, vehicular communications are a particularly promising technology. The sharing of information between vehicles over wireless links allows vehicles to perceive their environment beyond the field-of-view of their on-board sensors. Thanks to this enlarged representation of the environment in time and space, situation assessment is improved and dangerous situations can be detected earlier.

A PhD was started on this topic in 2009, in collaboration with Renault. It tackles the problem of risk estimation at road intersections from a new perspective: a Bayesian framework is proposed for reasoning about traffic situations and collision risk at a semantic level instead of at a trajectory level. While classic approaches estimate the risk of a situation by predicting the future trajectories of the vehicles and looking for intersections between them, here dangerous situations are detected by estimating the intentions of drivers and looking for conflicts between them. This novel approach to risk assessment is very relevant in the context of road traffic, as it takes into account the fact that the road network is a highly constrained environment regulated by traffic

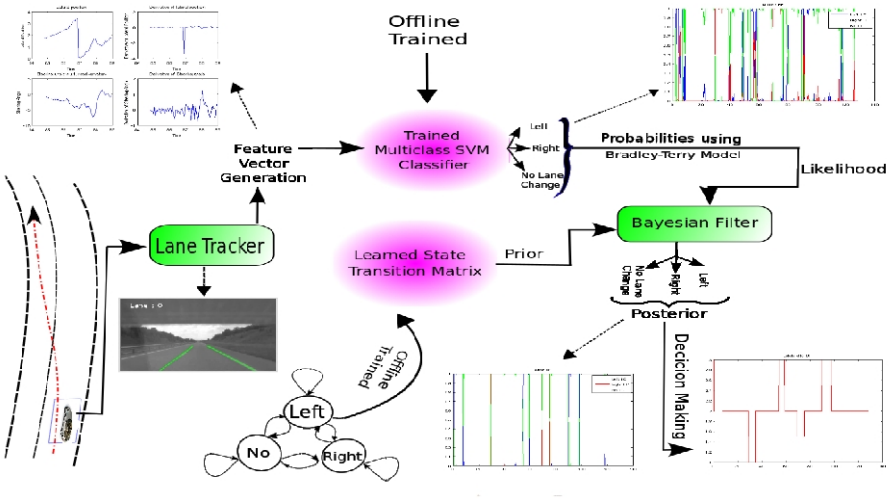


Figure 16. General architecture of the lane change prediction module.

rules. The proposed approach relies on the estimation of drivers' intentions, and the main difficulty lies in the presence of uncertainties in the estimation process: uncertainties inherent to sensor data, and ambiguities when linking vehicle behavior with driver intention. In this work the information about the state of other vehicles is obtained via vehicle-to-vehicle communication, but the proposed framework for reasoning on traffic situations and risk is general and can be applied with other types of sensors, e.g., the ones presented in 5.1.3 .

The focus of the first year (2010) was on estimating a driver's intended maneuver at an intersection (go straight, turn left, etc.) based on the current state of the vehicle (position, orientation, turn signal state) and on contextual information extracted from the digital map. The idea was to use the information on the geometry of the road network and on the connectivity between lanes to build a statistical model of the relationship between a vehicle's state and the driver's intended maneuver. The proposed solution is based on a Bayesian Network and on geometric functions which automatically extract the characteristics of the intersection from a digital map. This approach was designed and implemented during a 3-month internship in the Stanford Artificial Intelligence Laboratory, in collaboration with Sebastian Thrun's Driving Group.

During the second year (2011) we augmented the Bayesian Network with a filtering process so that new measurements could be recursively used to estimate the driver's intentions. The new version of the motion model explicitly models the influence of traffic rules on the behavior of a vehicle. While state-of-the-art approaches usually assume independence between vehicles, the proposed motion model takes into account the mutual influences between the maneuvers performed by the vehicles in the scene. These improvements were carried out by introducing two new variables in the Bayesian Network. The ***"Intention to stop"*** corresponds to the driver's intention to come to a halt at the intersection. The ***"Expectation to stop"*** corresponds to whether or not the traffic rules expect the driver to come to a halt at the intersection. The former is assumed to be dependent on the previous intention of the driver and on the current expectation. The latter is assumed to be dependent on the rules applying at the intersection and on the previous situational context, i.e., the state of the other vehicles in the scene. With this model it is possible to infer what a driver intends to do and what a driver is expected to do from the successive measurements of the pose, speed, and turn signals of the vehicles in the scene. Risk can then be computed based on the probability that intention and expectation do not match.

The focus of this year (2012) was on the evaluation of the performance of the algorithm. The proposed approach was validated in field trials using passenger vehicles equipped with vehicle-to-vehicle wireless communication modems, and in simulation. Our simulations assumed ideal perception and communication, and considered typical accident scenarios at a two-way-stop cross intersection. The tested maneuvers included crossing maneuvers, merging maneuvers, and left turn across path maneuvers (see Figure 17). A total of 240 instances of these scenarios were simulated, with both priority violations and stop violations as accident causes. The same number of instances were simulated for non-dangerous situations, by enforcing a 3 seconds safety distance between the vehicles at all times. An analysis of the collision prediction horizon led to the following conclusions:
1. There were no false alarms in non-dangerous situations, and no missed detection in the dangerous scenarios.
2. For merging and crossing maneuvers, the proposed algorithm was able to predict collisions at least 1.5 s before they occurred.
3. For left turn across path maneuvers, the proposed algorithm was able to predict collisions at least 0.6 s before they occurred.
4. Accidents caused by stop violations were detected on average 1 s earlier than the ones caused by priority violations.
Different accident avoidance strategies were tested: warning the driver of the vehicle with right-of-way, warning the driver of the other vehicle, applying autonomous braking on the vehicle with right-of-way, and applying autonomous braking on the other vehicle. It was found that the ability of each strategy to avoid an accident varies a lot with the situation. For example, the "autonomous braking on the vehicle with right-of-way" can avoid the accident in 91% of cases for stop violations, but only in 34% of cases for priority violations. "Warning the driver of the vehicle with right-of-way" can avoid the accident in 1% of cases for priority violations, while for the same scenarios "autonomous braking on the other vehicle" can avoid the accident in 99% of cases. These results were published at the conference IEEE IROS'12 [22], and as an Inria Research Report [41]. Field trials were conducted using two vehicles equipped with off-the-shelf vehicle-to-vehicle wireless communication modems. Six different drivers took part in the experiments to recreate realistic dangerous and non-dangerous situations at a T-shaped give-way intersection (see Figure 17). The risk estimation algorithm was run online in one of the vehicles, and triggered a warning for the driver when it detected a dangerous situation (see Figure 18). In the 120 tests, there were no false alarms and no missed detections. The warning was always triggered early enough that accidents were avoided by performing an emergency braking. The field trials proved that the proposed approach can operate with success in real-life situations and trigger warnings in real time. They also showed the robustness of the algorithm, since the experiments were carried out with several drivers, a positioning system with a precision of 2 meters (standard deviation) and challenging wireless communication conditions. These results were published at the conference IEEE IV'12 [23], where the paper received the Best PhD Student Paper award.

The PhD was successfully defended in October 2012 [9]. A patent application was filed with Renault in October 2012 [45]. This work will be continued within the Inria@SiliconValley program, in collaboration with the University of Berkeley, California. Ms Lefevre will conduct further research on this topic as a post-doctoral researcher at Berkeley starting January 2013.

5.2.3. *Guidance for Uncertain shooting domain*

Participant: Emmanuel Mazer.

This study is made in collabortion with MBDA (Monsieur Le Menec) and Probayes (Monsieur Laurent Saroul) under the ITP framework financed by the british MOD and the french DGA

Contex This project relates to the use of lock after launch missiles, both long range anti aircraft missiles such as Meteor, or air to ground strike weapons employing for example IIR or Semi Active Laser (SAL) guidance. In both cases, a target is ultimately recognized and tracked by means of a seeker which detects a characteristic signal above the noise. This could be the target reflections of a radar beam, or the spot from a designating laser.

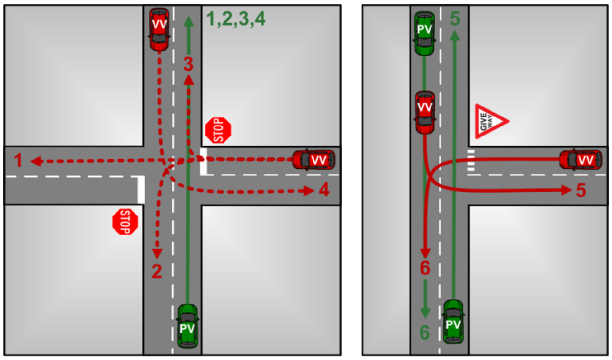


Figure 17. Scenarios tested in simulation (left) and during field trials (right).

Figure 18. Online execution of the algorithm during the field trials: warning the driver of an upcoming collision with a vehicle on the left.

However, a missile is often launched at a target range which is greater than its seeker acquisition range, although within the kinematics No Escape Zone (NEZ). It is provided with targeting geometry before launch, and maybe (via a data link) during the first part of the trajectory. However, it must fly for some period in inertial mode, and during this time the target may manoeuvre. Also, errors build up due to the imperfections in the inertial navigation system. This means that the target bearing becomes increasingly uncertain whilst the range reduces. It may be necessary to scan the seeker to acquire the target. If the scan is not matched to the possible manoeuvres, the target may escape detection. But if the scan is large, the acquisition range will be reduced, because of the reduction in search time per solid angle. As the target is acquired later, the missile's terminal manoeuvre will be more severe, and as a result the range assumed for the original kinematics NEZ may have been too optimistic. Equivalently, it is possible to be too pessimistic about the target uncertainty, hence to scan too much, and acquire the target so late that there is no longer the manoeuvre capability to reach it. Present Weapon systems optimize the probability of successful interception assuming either Gaussian uncertainties, or worst case uncertainties.

Objectives and achievments of the GUS-D system

These considerations lead to the concept of a stochastic approach for computing a probabilistic, adaptive NEZ. Probabilistic NEZ depend on the uncertain target behaviour. The uncertainties we propose to deal with are also related to the missile Inertial Navigation System (INS) precision, to sensor errors and to misalignments. Moreover, the uplink management, i.e., when to evade and breakdown the link between the launching platform and the in-flight missile plays a major role on the target localization accuracy and by the way to the size of the NEZ. Finally, there is uncertainty in the target radar cross section, which has a big effect on the seeker acquisition range. The purpose of this study is better tactical advice to the pilot about launching decision and how long maintain the uplink, and where appropriate, better matching of seeker scan strategies to target behaviours. These decisions have impacts on the probability of combat success; i.e., not only to hit the target but also on the probability to survive, as the opponent aircraft or ground threat may launch similar weapons.

The project focuses predominantly on Air to Air systems. The Meteor scan strategy has been studied deeply and is no longer critical for the engagement of fighter jets, but an objective of the study is to extend the strategy to the engagement of targets of much lower radar cross section, where the acquisition range is significantly shorter. Nevertheless all the issues apply also to Air to Ground weapon systems.

The GUS-D system is limited to one to one engagement scenario:

- one aircraft and its missile
- opponent aircraft and its missile

The main functionality of the GUS-D system is then to provide to the user a probability of successful target interception given the current engagement conditions, and the uncertainties on the target properties and behaviours.

5.3. Human Centered Navigation in the physical world

5.3.1. Goal oriented risk based navigation in dynamic uncertain environment

Participants: Anne Spalanzani, Jorge Rios-Martinez, Arturo Escobedo-Cabello, Procopio Silveira-Stein, Alejandro Dizan Vasquez Govea, Christian Laugier.

Navigation in large dynamic spaces has been adressed often using deterministic representations, fast updating and reactive avoidance strategies. However, probabilistic representations are much more informative and their use in mapping and prediction methods improves the quality of obtained results. Since 2008 we have proposed a new concept to integrate a probabilistic collision risk function linking planning and navigation methods with the perception and the prediction of the dynamic environments [57]. Moving obstacles are supposed to move along typical motion patterns represented by Gaussian Processes or Growing HMM. The likelihood of the obstacles' future trajectory and the probability of occupation are used to compute the risk of collision. The proposed planning algorithm, call RiskRRT (see Figure20 for an illustration), is a sampling-based partial planner guided by the risk of collision. Results concerning this work were published in [58] [59] [60]. In

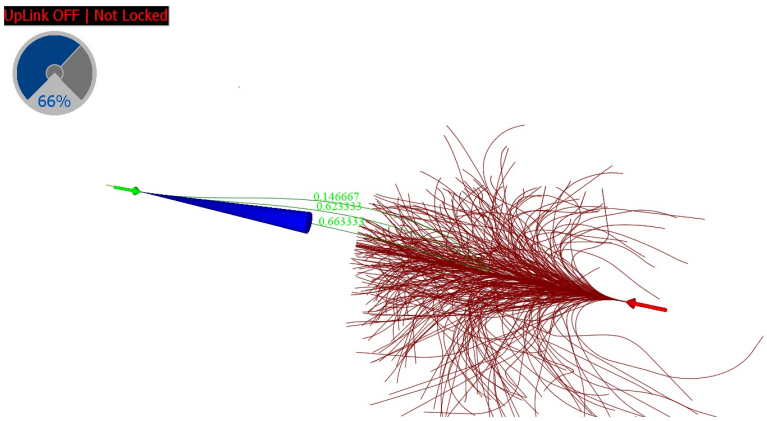


Figure 19. This Figure shows (up left corner) the "probability meter" which indicates the probability to intercept the target if the uplink is to be shutoff now. The 3D representation of the scene is displayed. The red filaments are future possible trajectories obtained with a Markov Process (MBD-UK). The blue cone modelizes the detection cone of the seeker. The green numbers indicates the probabilities to lock the target with the corresponding trajectory.

2012, We continue to work on developing probabilistic models and algorithms to analyze and learn human motion patterns from sensor data (e.g., tracker output) in order to perform inference, such as predicting the future state of people or classifying their activities. Our work has been published in the Handbook of Intelligent Vehicles [40]. We obtained some preliminary results on our robotic wheelchair combining RiskRRT with some social conventions described in section 5.3.2 . This approach and experimental results have been published at ISER 2012 [32].

This algorithms is used in the work presented in the next three sections, work conducted under the large scale initative project PAL.

5.3.2. Socially-aware navigation

Participants: Jorge Rios-Martinez, Anne Spalanzani, Alessandro Renzaglia, Agostino Martinelli, Christian Laugier.

Our proposal to endow robots with the ability of socially-aware navigation is the Social Filter, which implements constraints inspired by social conventions in order to evaluate the risk of disturbance represented by a navigation decision.
The Social Filter receives from the perception system a list of tracked humans and a list of interesting objects in the environment. The interesting objects are designated manually according to their importance in a particular context, for example, an information screen in a bus station. After the process of such data, the Social Filter is able to output the risk of disturbance relative to people and interesting objects, on request of the planner and the decisional system. Thus, the original navigation solutions are "filtered" according to the social conventions taken into account. Notice that the concept of social filter is built as a higher layer above the original safety strategy, the planner and the decisional system are responsible to include the new constraints.
The on-board Kinect attached to our robotic platform was used to track people and to detect interactions. The Kinect sensor permits to get the position and orientation of the torso for each identified human. That information is passed to the Social Filter. Result images can be seen in Figure 21 .

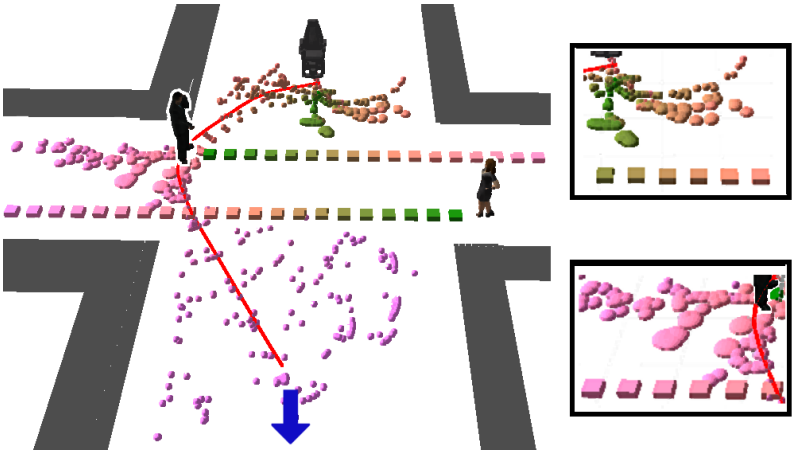
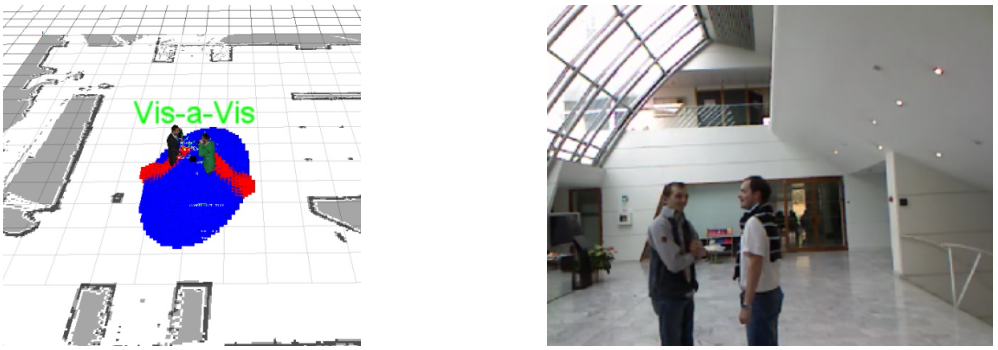


Figure 20. Predictive navigation example. RiskRRT selected a plan (red line) to the goal (blue arrow). The chosen path leads the robot to pass by the back of the first person, and then reduces the speed to let the second person to pass as well. With this strategy, the robot minimizes the risk of collision and the discomfort caused for the two pedestrians. Once second person has passed, the algorithm choses a straighter path to the goal. Frames at the right of the figure show that estimated risk is bigger at future positions of the wheelchair (circles) which are close to predicted positions of pedestrians (squares).



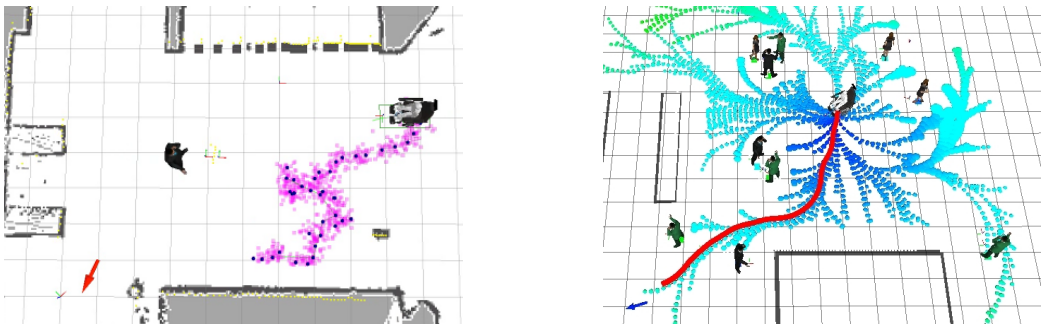
(a) (b)

Figure 21. Interaction detected with Social Filter from Kinect input for a pair of humans. Torso direction is used to estimate the main focus of interest.

In the context of socially-aware robot navigation in dynamic environments, as part of Jorge Rios-Martinez PhD thesis (to be defended in january 2013), two techniques have been proposed: one considering optimization-based navigation presented in [26] and the other a Risk-based navigation approach, previously presented in [75].

The **optimization-based navigation strategy**, done in collaboration with A. Renzaglia, is based on the Cognitive-based Adaptive Optimization (CAO) approach applied to robots [10]. We formulate the problem of socially-aware robot navigation as an optimization problem where the objective function includes, in addition to the distance to goal, information about comfort of present humans. CAO is able to efficiently handle optimization problems for which an analytical form of the function to be optimized is unknown, but the function is available for measurements at each iteration. A model of social space, contained in the Social Filter module, was integrated in order to work as a "virtual" sensor providing comfort measures. Figure22 a) shows an image of the method implementation on ROS [2] framework.

Social Filter models of social conventions were combined with RiskRRT [56] by including the knowledge of human management of space (Personal Space, interaction space, activity Space). The particular considered interaction was the conversation between pedestrians which was missed in the most part of related works. The approach presented shows a way to take into account social conventions in navigation strategies providing the robot with the ability to respect the social spaces in its environment when moving safely towards a given goal. Due to the inclusion of our social models, the risk calculated for every partial path produced by RiskRRT algorithm is given by the risk of collision along the path and the risk of disturbance to human spaces.



(a) (b)

Figure 22. Results of socially-aware navigation approaches. In a) the optimization-based navigation solution avoids a region were the discomfort for the human would be higher. In b) the Risk-Based navigation technique explores the space and decides to follow a path avoiding social spaces minimizing the risk of disturbance. The goal in each case is signaled by an arrow.

One last work was presented in [25], where the socially-aware navigation based on risk was integrated with a model of human intention estimation (presented in section 5.3.4 . Results exhibited emerging behavior showing a robotic wheelchair interpreting facial gesture commands, estimating the intended goal and autonomously taking the user to his/her desired goal, respecting social conventions during its navigation.

5.3.3. Navigation Taking Advantage of Moving Agents

Participants: Procopio Silveira-Stein, Anne Spalanzani, Christian Laugier.

[2] http://www.ros.org

Following a leader in populated environments is a form of taking advantage of the motion of the others. A human can detect cues from other humans and smartly decide in which side to pass. Humans can also easily predict the motion of the others, changing his/her path to accommodate for conflictive situations, for example. Imitating the motion of a human can also improve the social acceptance of robots and so on.

The best leader is the one whose goal is close to the robot's one. To implement that, the Growing Hidden Markov Model (GHMM) technique is used [79]. This technique provides at the same time a capability to learn and modeling typical paths, as well as learning and predicting goals associated to paths, making it ideal for the proposed approach of leader election.

Once a leader is chosen, the robot starts to track his/her path and follow it, using the RiskRRT algorithm presented in section 5.3.1 . This algorithm takes into account the risk of collision with other agents, guaranteeing that the robot can avoid collisions even if its leader is lost or occluded.

Some results can be seen in the following experiments, where real human data was used together with a robot simulator.

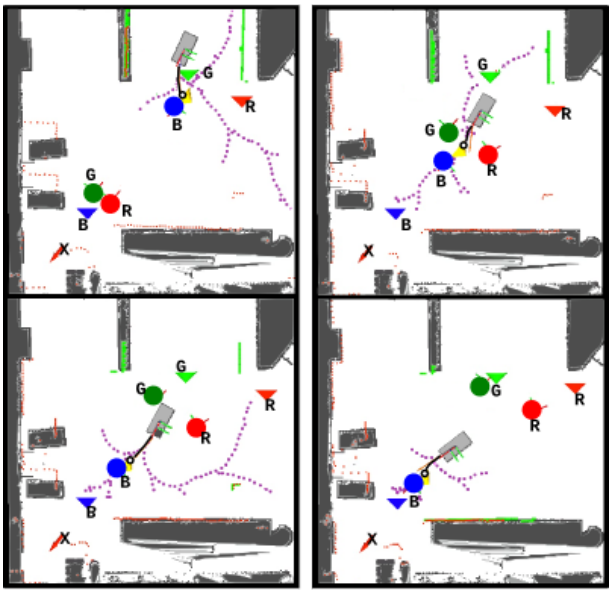


Figure 23. Robot navigation following a leader seamlessly avoid other incoming agents.

In Figure 23 , the experiment demonstrates one of the advantages of following a leader to improve the robot's navigation capabilities. The direct path to the robot's goal is obstructed by two incoming humans. Normally an algorithm suited for dynamic environment would create a detour as future humans' position would conflict with the robot straight trajectory. However, as the robot is following a leader, it does not reason about the other agent's future position. Therefore, the leader knows that people will give room for he/she to pass, and the robot profits from it.

Next step will be to use this technics will navigating in a crowd, task that a common planning strategy could hardly do.

5.3.4. Autonomous Wheelchair for Elders Assistance

Participants: Arturo Escobedo-Cabello, Gregoire Vignon, Anne Spalanzani, Christian Laugier.

The aging of world's population is bringing the need to provide robotic platforms capable to assist elder people to move [77]. It is necessary that such transportation is reliable, safe and comfortable. People with motor disabilities and elders are expected to benefit from new developments in the field of autonomous navigation robotics. Autonomously driven wheelchairs are a real need for those patients who lack the strength or skills to drive a normal electric wheelchair. The services provided by this kind of robots can also be used to provide a service of comfort, assisting the user to perform difficult tasks as traversing a door, driving in a narrow corridor etc. Simple improvements of the classical powered wheelchair can often diminish several difficulties while driving. This idea of comfort has emerged as a design goal in autonomous navigation systems, designers are becoming more aware of the importance of the user when scheming solution algorithms. This is particularly important when designing services or devices intended to assist people with some disability.

In order for the robot to have a correct understanding of the intention of the user (when moving around) it is necessary to create a model of the user that takes into account his habits, type of disability and environmental information. The ongoing research project is centered in the understanding of the intentions of the user while driving an autonomous wheelchair, so that we can use this information to make this task easier.

In 2011 a robotic wheelchair was set up as experimental platform. Some basic functions were included as the mapping of the environment using a Rao-Blackwellized Particle Filter [62], localization using an Adaptive Monte Carlo Localization approach (AMCL) [78], global planning using an A* algorithm [63] and local reactive planning using the Dynamic Window Algorithm [55]. Alongside some work was done with the kinect sensor in order to detect and track people. This behaviour was aimed to bring assistance not only to the user but also to the caregiver by allowing him to move more freely. The software implementation of the related approaches was done on the basis of the ROS middleware.

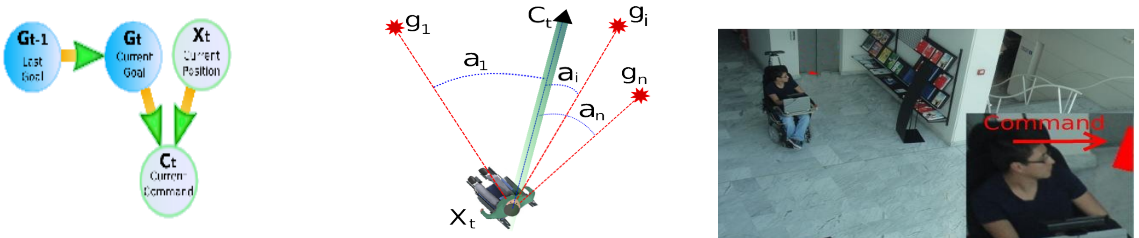
During 2012 the work was centered in the improvement of the usability of the system around three main axes:

- User intention estimation: A review of the state of the art in user's intention estimation algorithms was made and a new model to infer the intentions of the user in a known environment was presented [46],[47]. The algorithm models the intention of the user as 2D topological goals in the environment. Those places are selected according to how frequently they are visited by the user (user habits). The system was designed so that the user can give orders to the wheelchair by using any type of interface, as long as he can show the direction of the intended movement (joystick, head tracking, brain control, etc). As shown ni figure24 , the chosen approach uses a Bayesian model to model and infer the intentions. The main contribution of this work is to model the intention of the user as topological goals instead of normal trajectory-based methods, therefore the model is simpler to deal with. Current research is being done to understand which information is important to take into account in order to do better estimations of the user's intention. In particular, the movements of the head are considered by the proposed inference method.

  The navigation is performed using the human-aware planning algorithm developed by the team which integrates a notion of social conventions and avoidance of dynamic obstacles to prevent uncomfortable situations when the wheelchair is navigating among humans (see section 5.3.2 for details)
- Interfaces: People with motor disabilities and elders often have problems using joysticks and other standard control devices. Under this consideration our experimental platform was equipped with different types of user-interfaces to provide a multimodal functionality as described in [47]. A face pose interface allows to control the wheelchair's motion by changing the face direction, while voice recognition interface is used to guarantee an adequate control of the wheelchair for those commands that otherwise would be difficult to give by only using the face (Stop, start, etc). The use of a touch screen control is also possible.
- Multimodal control: The wheelchair can be controlled in semi-autonomous mode employing the user's intention estimation module, described later, or in manual mode in which the user is in charge of driving by him self.

  In manual mode the user controls the wheelchair's angular speed moving her head while the linear speed is controlled with vocal commands (faster, slower, break, etc).

In semi-autonomous mode the user shows the direction to his/her desired destination facing towards it. Whenever a new command is read from the face pose estimation system. The user's intention module computes the goal with the highest posterior probability. The navigation module receives the map of the environment, the list of humans present in the scene and the currently estimated goal to compute the necessary trajectory to the goal.



(a) (b) (c)

Figure 24. ***left****:User's intention model. The Bayesian network used to estimate the current user's intended goalG_t, The current position X_t and the user command C_t are used as evidence. G_t is dependent on the value of the last estimation G_{t-1}.* ***center****: Experimental evaluation of the user's intention module. The probability value for a given command C_t (big arrow) is proportional to the angle a_i formed respect to each goal g_i in the environment.* ***right:*** *The user is looking to the left (in the direction of his desired goal). Once that the user's intention estimation system computes the goal with the highest probability, the autonomous navigation module plans the path and controls the movement of the wheelchair to take the user to the destination.*

5.3.5. Multi-Robot Distributed Control under Environmental Constraints

Participants: Agostino Martinelli, Alessandro Renzaglia.

This research is the follow-up of a study begun three years ago in the framework of the European project sFly. The problem addressed is the deployment of a team of flying robots to perform surveillance coverage mission over an *unknown* terrain of complex and non-convex morphology. In such a mission, the robots attempt to maximize the part of the terrain that is visible while keeping the distance between each point in the terrain and the closest team member as small as possible. A trade-off between these two objectives should be fulfilled given the physical constraints and limitations imposed at the particular application. As the terrain's morphology is unknown and it can be quite complex and non-convex, standard algorithms are not applicable to the particular problem treated in this paper. To overcome this, a new approach based on the Cognitive-based Adaptive Optimization (CAO) algorithm is proposed and evaluated. A fundamental property of this approach is that it shares the same convergence characteristics as those of constrained gradient-descent algorithms (which require perfect knowledge of the terrain's morphology and optimize surveillance coverage subject to the constraints the team has to satisfy). Rigorous mathematical arguments and extensive simulations establish that the proposed approach provides a scalable and efficient methodology that incorporates any particular physical constraints and limitations able to navigate the robots to an arrangement that (locally) optimizes surveillance coverage.

Special focus has been devoted to adapt this general approach in order to deal with real scenarios. Specifically, this has been carried out by working in collaboration with the ETHZ (Zurich). To this regard, the approach has been adopted in the framework of the final demo of the sFly project. The demo simulates a search and rescue operation in an outdoor GPS-denied disaster scenario. No laser, no GPS, and Vicon or other external cameras are used for navigation and mapping, but just onboard cameras and IMUs. All the processing runs onboard, on a Core2Duo processing unit. The mission consists of first collecting images for creating a common global

map of the working area with 3 helicopters, then engaging positions for an optimal surveillance coverage of the area, and finally detecting the transmitter positions.

The results of this research have been published in two journals, [14], [15], and on the thesis of A. Renzaglia, [10].

5.4. Bayesian Modelling of Sensorimotor Systems and Behaviors

Results described in this section were done in collaboration with the LPPA collège de France.

5.4.1. Bayesian based decision making in multi-player video games

Participants: Gabriel Synnaeve, Pierre Bessière.

The problem addressed in this work is the autonomous replacement of a human player. It is the continuation of last year's work on the same topic as well as a follow-up of previous E-Motion Ph.D Ronan Le Hy [64]. This year, we focused on real-time strategy (RTS) games, in which the players have to build an economy, advance technology, produce and control an army to kill the opponents. From a research point of view, multi-player games are interesting because they stand for a good in-between of the real world and simulations. The world is finite and simulated (no sensors problems) but we didn't wrote the simulation and the other players are humans (or advanced robots in the case of AI competitions).

This year's research work focused on tactical prediction and decision-making as well as armies composition adaptation. For the tactical model, the idea is to have (most probably biased) lower-level heuristics from units observations, which produce information exploitable at the tactical level, and take advantage of strategic inference too. We abstract space into automatically extracted choke points and regions of StarCraft maps from a pruned Voronoi diagram (using [71]). We then assign different scores to each of these regions and learn the influence of these scores on different attack types and locations. To do that, we set up a huge data-set of professional player's games, whose game state was extracted [29]. This work was accepted for publication at Computational Intelligence in Games (IEEE CIG) 2012 in Grenada [30] and was presented at the Computer Games Workshop of the European Conference of Artificial Intelligence (ECAI) 2012 [28].

Another focus of work this year was on army composition adaptation. RTS games unit types combinations in armies can be seen as complex (soft max) rock-paper-scissors games. Our analysis boiled to down army compositions encoded as clusters (we used a Gaussian Mixtures Model) of "classic" combinations (because of economy and technology constraints during the game). This work was published at the AI in Adversarial Real-Time Games workshop of AAAI AIIDE 2012 [29].

On top of the research/evaluation implementation, we also implemented it in our StarCraft: Broodwar's bot implementation BroodwarBotQ. With this bot, we took part in AIIDE and CIG conferences AI tournaments placing respectively 4th (out of 10) and 6th (out of 10). Gabriel Synnaeve defended his thesis on October 24th 2012.

5.4.2. Bayesian modelling to implement and compare different theories of speech communication

Participants: Raphael Laurent, Pierre Bessière, Julien Diard, Jean-Luc Schwartz.

A central issue in speech science concerns the nature of representations and processes involved in communication. The search for phoneme or syllable specific invariants led to three major sets of approaches: motor, auditory and perceptuo-motor theories. They have been widely argued for and against, but the theoretical debate appears to be stagnating. It is our belief that computationnal models designed within a rigorous mathematical framework may allow to put forward new arguments to support either theory, and new ideas for experiments to be carried out on human subjects.

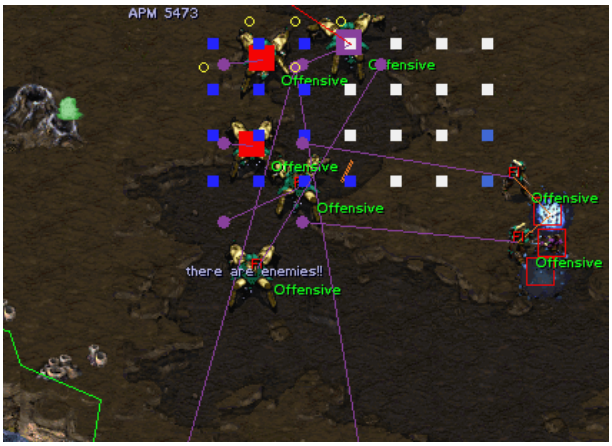


Figure 25. Units movement debugging output during a StarCraft game of the BroodwarBotQ bot. Considering the unit in the upper middle of the picture, white squares represent the highest probabilities of directions, while the darker the blue, the lower the probability to go there. The unit controller searches both to minimizes collisions and stay in range of the enemy targets (bottom right).

We have designed an integrative Bayesian model which allows to study auditory, motor and perceptuo-motor aspects of speech production and perception. In 2011, this model was used to work on purely theoretical simulations where we studied with diverse paradigms the decrease in the performances predicted by the different theories due to communication noise. This work led to the proof of an indistinguishability theorem : given some hypotheses on the learning process, purely motor and purely auditory models have identical answers to perception tasks. Thanks to VLAM, a vocal tract simulation tool which allows to map articulatory parameters to acoustic signals, we tested our model on vowel perception tasks. The results of both these studies are detailed in [70].

In 2012, we worked on a much more complex version of the model, which mas made able to deal with plosive syllable production and perception. A first version of this model was tested on perception tasks on evaluation corpora with more and more variability compared to the learning corpus. This showed a really high robustness of the purely motor model, which contained more information that it is the case in practise, due to unrealistic learning methods. That's why the work was then focused on more realistic learning algorithms, where speech motor gestures are unsupervisedly learned through imitation, by generating motor gestures trying to reach auditory targets, and memorising the acoustics corresponding to these motor commands.

5.4.3. Bayesian programming : book and software

Participants: Emmanuel Mazer, Pierre Bessière.

5.4.3.1. A need for a new computing paradigm

Bayesian probability theory is a mathematical alternative to logic.

However, we want working solutions to incomplete and uncertain problems. Consequently, we require an alternative computing framework based on Bayesian probabilities.

To create such a complete computing Bayesian framework, we require a new *modeling methodology* to build probabilistic models, we require new *inference algorithms* to automate probabilistic calculus, we require new *programming languages* to implement these models on computers, and finally, we will eventually require new *hardware* to run these Bayesian programs efficiently.

Our ultimate goal is a Bayesian Computer. The purpose of this book is to describe a formalism and a computer langage as first steps in this direction.

5.4.3.2. Outline of the book

Its purpose is to introduce the fundamental concepts of Bayesian Programming, to present the novelty and interest of the approach, and to initiate the reader to the Bayesian modeling. Numerous simple examples of applications are presented in different fields.

It is divided in three parts, chapters *2 Basic-Concepts* to *6: Bayesian-Program* which presents the principles of Bayesian Programming, chapters *7 : Information-Fusion* to *11 : Bayesian-Programming-Iteration* which offer a cook book for the good practice of probabilistic modeling and *12 : Bayesian Programming Formalism* to *16 Frequently Asked Question* which revisit the Bayesian inference and learning problems with the help of the presented formalism.

A fist version of the book will be send to the reviewer selected by the editor before the end of 2012

5.4.3.3. Distributed Software

One way to read this book and learn bayesian programming is to run and modify the programs given as example. A Python package "pypl" based on Probt bindings is made available with book.

The source code of the examples as well as the Python package can be downloaded free of charge.

Many examples in the book are given with parts of real corrresponding programs which could be run using the distributed package. They are given under the following format

The figure 26 has been generated using the program "chapter7/invpgm.py". The following instruction allows to get to most probable value for the heading H given the readings.

```
PH=PHkB0B1.instantiate(sensor_reading_values)
best=PH.compile().best()
```

Bayesian programs are also used to generate the illustration of the book such as this one which illustrating the navigation based on sensor fusion.

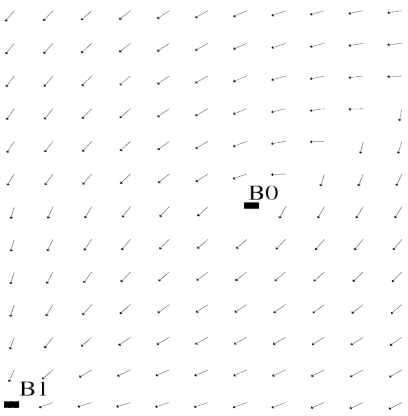


Figure 26. The vector field corresponding to $\max_h P\left(H = h | b_0 \wedge b_1 \wedge \pi\right)$

FLOWERS Project-Team

6. New Results

6.1. Autonomous and Social Skill Learning and Development

6.1.1. Active Learning and Intrinsic Motivation

6.1.1.1. Active Learning of Inverse Models with Goal Babbling

Participants: Adrien Baranes, Pierre-Yves Oudeyer.

We have continued to elaborate and study our Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity (SAGG-RIAC) architecture as an intrinsically motivated goal exploration mechanism which allows active learning of inverse models in high-dimensional redundant robots. Based on active goal babbling, this allows a robot to efficiently and actively learn distributions of parameterized motor skills/policies that solve a corresponding distribution of parameterized tasks/goals. The architecture makes the robot sample actively novel parameterized tasks in the task space, based on a measure of competence progress, each of which triggers low-level goal-directed learning of the motor policy parameters that allow to solve it. For both learning and generalization, the system leverages regression techniques which allow to infer the motor policy parameters corresponding to a given novel parameterized task, and based on the previously learnt correspondences between policy and task parameters.
We have conducted experiments with high-dimensional continuous sensorimotor spaces in three different robotic setups: 1) learning the inverse kinematics in a highly-redundant robotic arm, 2) learning omnidirectional locomotion with motor primitives in a quadruped robot 1718 , 3) an arm learning to control a fishing rod with a flexible wire. We show that 1) exploration in the task space can be a lot faster than exploration in the actuator space for learning inverse models in redundant robots; 2) selecting goals maximizing competence progress creates developmental trajectories driving the robot to progressively focus on tasks of increasing complexity and is statistically significantly more efficient than selecting tasks randomly, as well as more efficient than different standard active motor babbling methods; 3) this architecture allows the robot to actively discover which parts of its task space it can learn to reach and which part it cannot.
This work was published in the journal Robotics and Autonomous Systems [22].

6.1.1.2. Exploration in Model-based Reinforcement Learning

Participants: Manuel Lopes, Tobias Lang, Marc Toussaint, Todd Hester, Peter Stone, Pierre-Yves Oudeyer.

Formal exploration approaches in model-based reinforcement learning estimate the accuracy of the currently learned model without consideration of the empirical prediction error. For example, PAC-MDP approaches such as R-MAX base their model certainty on the amount of collected data, while Bayesian approaches assume a prior over the transition dynamics. We propose extensions to such approaches which drive exploration solely based on empirical estimates of the learner's accuracy and learning progress. We provide a "sanity check" theoretical analysis, discussing the behavior of our extensions in the standard stationary finite state-action case. We then provide experimental studies demonstrating the robustness of these exploration measures in cases of non-stationary environments or where original approaches are misled by wrong domain assumptions. [46]. Furthermore, we studied how different exploration algorithms can be combine and selected at runtime. Typically the user must hand-tune exploration parameters for each different domain and/or algorithm that they are using. We introduced an algorithm called leo for learning to select among different exploration strategies on-line. This algorithm makes use of bandit-type algorithms to adaptively select exploration strategies based on the rewards received when following them. We show empirically that this method performs well across a set of five domains In contrast, for a given algorithm, no set of parameters is best across all domains. Our results demonstrate that the leo algorithm successfully learns the best exploration strategies on-line, increasing the received reward over static parameterizations of exploration and reducing the need for hand-tuning exploration parameters [42].

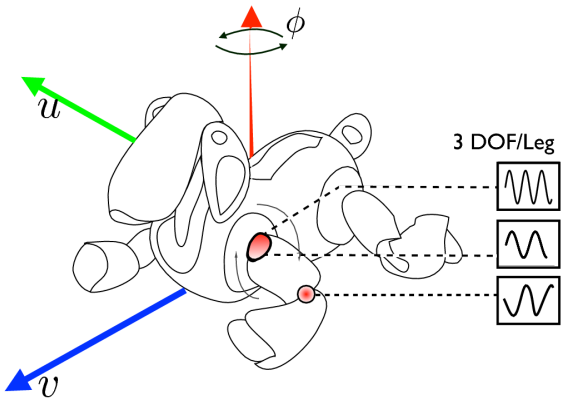


Figure 17. Experimenting SAGG-RIAC for learning an inverse model for omnidirectional locomotion of a quadruped robot. The quadruped robot is controlled using 24 dimensional motor synergies parameterized with 24 continuous values : 12 for the amplitudes and 12 others for the phases of a sinusoid tracked by each motor. Experiments consider a task space u, v, α which corresponds to the 2D position and orientation of the quadruped.

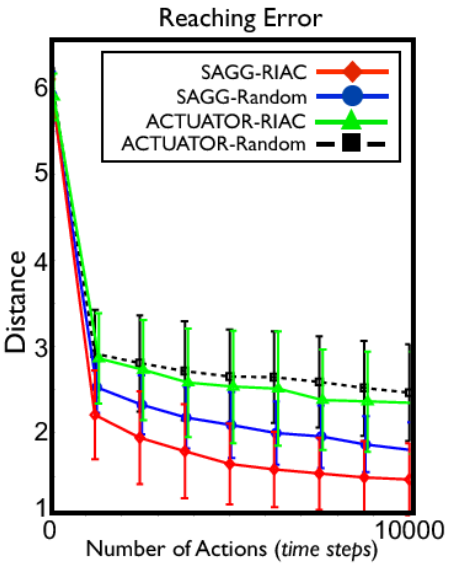
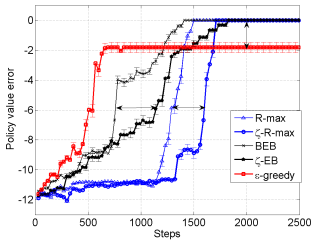
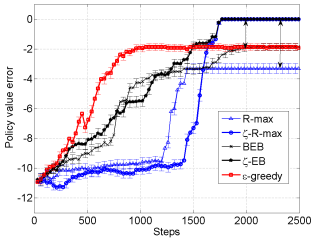
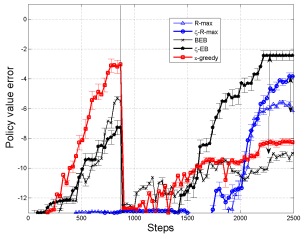


Figure 18. Evolution of the quality of the learnt inverse model for the quadruped robot experiment, depending on various exploration strategies (measured as mean error over a set of uniformly distributed goals generated independantly from learning trials).



(a) *Experiment 1—Correct Assumptions*





(b) *Experiment 2—Violated Assumptions* (c) *Experiment 3—Change in Dynamics*

Figure 19. Experiments: (a) Like Rmax and BEB with correct assumptions, our algorithms ζ-Rmax and ζ-EB based on an empirical estimation of the learning progress converge to the optimal policy without relying on these assumptions, but take a small extra amount of time. (b) When their assumptions are violated, Rmax and BEB fail to converge, while ζ-Rmax and ζ-EB don't rely on these assumptions and again find the optimal policy. (c) In contrast to existing methods, ζ-Rmax and ζ-EB can cope with the change in transition dynamics after 900 steps and refocus their exploration.

6.1.1.3. The Strategic Student Approach for Life-Long Exploration and Learning

Participants: Manuel LOPES, Pierre-Yves OUDEYER.

We introduced and formalized a general class of learning problems for which a developmental learning strategy is shown to be optimal. This class of problems can be explained using the strategic student metaphor: a student has to learn a number of topics (or tasks) to maximize its mean score, and has to choose strategically how to allocate its time among the topics and/or which learning method to use for a given topic. We show that if the performance curves are sub-modular, then a strategy where time allocation or learning method are chosen in a developmental manner is optimal. We argue that this optimal developmental trajectory can be automatically generated by greedy maximization of learning progress. This optimal strategy amounts to creating a structured developmental exploration where typically easy tasks are first explored, and then progressively more complicated ones are explored. Furthermore, this result holds independently of the nature of the topics and the learning methods used. Then, we show an algorithm, based on multi-armed bandit techniques, that allows empirical online evaluation of learning progress and approximates the optimal solution. Finally, we show that the strategic student problem formulation allows to view in a common framework many previous approaches to active and developmental learning [47].

6.1.1.4. Active Inverse Reinforcement Learning through Generalized Binary Search

Participants: Manuel Lopes, Francisco Melo.

We contributed the first aggressive active learning algorithm for nonseparable multi-class classification. We generalize an existing active learning algorithm for binary classification [107] to the multi-class setting, and identify mild conditions under which the proposed method provably retains the main properties of the original algorithm, namely consistency and sample complexity. In particular, we show that, in the binary case, our method reduces to the original algorithm of [107]. We then contribute an extension of our method to multi-label settings, identify its main properties and discuss richer querying strategies. We conclude the paper with two illustrative application examples. The first application features a standard text-classification problem. The second application scenario features a learning from demonstration setting. In both cases we demonstrate the advantage of our active sampling approach against random sampling. We also discuss the performance of the proposed approach in terms of the derived theoretical bounds.

6.1.1.5. Towards high-dimensional and cumulative task space active exploration

Participant: Benureau Fabien.

One direction of research of the team has been on intrinsic motivation in the context of autonomous learning. Building on the PhD work of Adrien Baranes, the efforts have concentrated on creating algorithms capable to handle high-dimensional spaces and manage context with multiple tasks. The goal is for the learner to be able to autonomously create collection of reusable skills. In this context, two main research efforts have been led this year.

A typical robot is made of chains of joints. We can take advantage of the fact that joints earlier in the chain have more impact that joints further down. Given sensory feedback on the middle of the chain, an algorithm can use this information to boost learning speed and divide the learning space in subsets of smaller dimensions. We wanted to adapt this idea to high dimensional space, and specifically to the interaction with objects; a robotic arm that has already learned an inverse model of its kinematic could reuse this knowledge learn about the mapping between the position of the end-effector and the displacement of an object it is manipulating. Experiments were conducted, but they lead to the conclusion that such an approach, while effective in some specific setting, relies too heavily on a good representation of the end effector position and motion, which, in some cases, requires sensory space of higher dimension that the motor space, thus defeating the purpose. This approach was not found to be robust enough for the type of robotic context our lab is pursuing.

The SAGG-RIAC architecture is an efficient but complex architecture which implementation cannot be easily summarized in a few lines of pseudo-code. This is problematic because it reduces the ability of other research groups to implement and reuse our algorithms for their own work. An effort was started this year to create an implementation of SAGG-RIAC that would be more robust and simpler. The main idea was to use kernels rather than bins to estimate in interest in SAGG-RIAC. This approach led to very promising results, notably in

its ability to handle unbounded sensory spaces. We aim at publishing the result of this work in 2013, together with a publicly available implementation of our algorithms with easy to run examples for dissemination of active learning architectures elaborated in the team. This work will also be reused in the participation of the lab into the MaCSi project.

6.1.2. Learning and optimization of motor policies

6.1.2.1. Off-Policy Actor-Critic

Participants: Thomas Degris, Martha White, Richard Sutton.

Actor–critic architectures are an interesting candidate for learning with robots: they can represent complex stochastic policies suitable for robots, they can learn online and incrementally and their per-time-step complexity scales linearly with the number of learned weights. Moreover, interesting connections have been identified in the existing literature with neuroscience. Until recently, however, practical actor–critic methods have been restricted to the on-policy setting, in which the agent learns only about the policy it is executing.

In an off-policy setting, on the other hand, an agent learns about a policy or policies different from the one it is executing. Off-policy methods have a wider range of applications and learning possibilities. Unlike on-policy methods, off-policy methods are able to, for example, learn about an optimal policy while executing an exploratory policy, learn from demonstration, and learn multiple tasks in parallel from a single sensory-motor interaction with an environment. Because of this generality, off-policy methods are of great interest in many application domains.

We have presented the first actor-critic algorithm for off-policy reinforcement learning. Our algorithm is online and incremental, and its per-time-step complexity scales linearly with the number of learned weights. We have derived an incremental, linear time and space complexity algorithm that includes eligibility traces and empirically show better or comparable performance to existing algorithms on standard reinforcement-learning benchmark problems. This work was presented by Degris et al. [38] and was reproduced independently by Saminda Abeyruwan from the University of Miami.

6.1.2.2. Auto-Actor Critic

Participant: Thomas Degris.

As mentioned above, actor–critic architectures are an interesting candidate for robots to learn new skills in unknown and changing environments. However, existing actor–critic architectures, as many machine learning algorithms, require manual tuning of different parameters to work in the real world. To be able to systematize and scale-up skill learning on a robot, learning algorithms need to be robust to their parameters. The Flowers team has been working on making existing actor–critic algorithms more robust to make them suitable to a robotic setting. Results on standard reinforcement learning benchmarks are encouraging. This work will be submitted to international conference related with reinforcement learning. Interestingly, the methods developed in this work also offer a new formalism to think about different existing themes of Flowers research such as curiosity and maturational constraints.

6.1.2.3. Relationship between Black-Box Optimization and Reinforcement Learning

Participant: Freek Stulp.

Policy improvement methods seek to optimize the parameters of a policy with respect to a utility function. There are two main approaches to performing this optimization: reinforcement learning (RL) and black-box optimization (BBO). In recent years, benchmark comparisons between RL and BBO have been made, and there has been several attempts to specify which approach works best for which types of problem classes.

We have made several contributions to this line of research by: 1) Defining four algorithmic properties that further clarify the relationship between RL and BBO. 2) Showing how the derivation of ever more powerful RL algorithms displays a trend towards BBO. 3) Continuing this trend by applying two modifications to the state-of-the-art PI^2 algorithm, which yields an algorithm we denote PI^{BB}. We show that PI^{BB} is a BBO algorithm, and, more specifically, that it is a special case of the state-of-the-art CMAES algorithm. 4) Demonstrating that the simpler PI^{BB} achieves similar or better performance than PI^2 on several evaluation tasks. 5) Analyzing why BBO outperforms RL on these tasks. These contributions have been published on HAL [69], and have been submitted to JMLR.

This work has also resulted in the novel PI^2-CMA, PI^2-CMAES algorithms, which are presented in [63], [60], [62]

6.1.2.4. Reinforcement Learning with Sequences of Motion Primitives for Robust Manipulation

Participant: Freek Stulp.

Physical contact events often allow a natural decomposition of manipulation tasks into action phases and subgoals. Within the motion primitive paradigm, each action phase corresponds to a motion primitive, and the subgoals correspond to the goal parameters of these primitives. Current state-of-the-art reinforcement learning algorithms are able to efficiently and robustly optimize the parameters of motion primitives in very high-dimensional problems. These algorithms often consider only shape parameters, which determine the trajectory between the start- and end-point of the movement. In manipulation, however, it is also crucial to optimize the goal parameters, which represent the subgoals between the motion primitives. We therefore extend the policy improvement with path integrals (PI^2) algorithm to simultaneously optimize shape and goal parameters. Applying simultaneous shape and goal learning to sequences of motion primitives leads to the novel algorithm PI^2-Seq. We use our methods to address a fundamental challenge in manipulation: improving the robustness of everyday pick-and-place tasks. This work was published in IEEE Transactions on Robotics [31] and Robotics and Autonomous Systems [26].

6.1.2.5. Model-free Reinforcement Learning of Impedance Control in Stochastic Environments

Participant: Freek Stulp.

For humans and robots, variable impedance control is an essential component for ensuring robust and safe physical interaction with the environment. Humans learn to adapt their impedance to specific tasks and environments; a capability which we continually develop and improve until we are well into our twenties. We have reproduced functionally interesting aspects of learning impedance control in humans on a simulated robot platform.

As demonstrated in numerous force field tasks, humans combine two strategies to adapt their impedance to perturbations, thereby minimizing position error and energy consumption: 1) if perturbations are unpredictable, subjects increase their impedance through co-contraction; 2) if perturbations are predictable, subjects learn a feed-forward command to offset the perturbation. We show how a 7-DOF simulated robot demonstrates similar behavior with our model-free reinforcement learning algorithm , by applying deterministic and stochastic force fields to the robot's end-effector. We show the qualitative similarity between the robot and human movements.

Our results provide a biologically plausible approach to learning appropriate impedances purely from experience, without requiring a model of either body or environment dynamics. Not requiring models also facilitates autonomous development for robots, as pre-specified models cannot be provided for each environment a robot might encounter. This work was published in IEEE Transactions on Autonomous Mental Development [29].

6.1.2.6. Probabilistic optimal control: a quasimetric approach

Participants: Clément Moulin-Frier, Jacques Droulez, Steve Nguyen.

During his previous post-doc at the Laboratoire de Physiologie de la Perception et de l'Action (Collège de France, Paris), Clément Moulin-Frier joined Jacques Droulez and Steve N'Guyen to work on an alternative and original approach of probabilistic optimal control called the quasimetric. A journal paper (soon to be submitted) was written in 2012, where the authors propose a new approach for dealing with control under uncertainty.

6.1.3. Social learning and intrinsic motivation

6.1.3.1. Optimal Teaching on Sequential Decision Tasks

Participants: Manuel Lopes, Maya Cakmak.

A helpful teacher can significantly improve the learning rate of an autonomous learning agent. Teaching algorithms have been formally studied within the field of Algorithmic Teaching. These give important insights into how a teacher can select the most informative examples while teaching a new concept. However the field has so far focused purely on classification tasks. We introduced a novel method for optimally teaching sequential decision tasks. We present an algorithm that automatically selects the set of most informative demonstrations and evaluate it on several navigation tasks. Next, we present a set of human subject studies that investigate the optimality of human teaching in these tasks. We evaluate examples naturally chosen by human teachers and found that humans are generally sub-optimal. Then based on our proposed optimal teaching algorithm we try to elicit better teaching from humans. We do this by explaining the intuition of the teaching algorithm in an informal language prior to the teaching task. We found that this improves the examples elicited from human teachers on all considered tasks. This shows that a simple modification the instructions given to human teachers, has the potential of greatly improving the performance of the agent trained by the human [32].

6.1.3.2. Socially Guided Intrinsic Motivation for Skill Learning

Participants: Sao Mai Nguyen, Pierre-Yves Oudeyer.

We have explored how social interaction can bootstrap the learning of a robot for motor learning. We first studied how simple demonstrations by teachers could have a bootstrapping effect on autonomous exploration with intrinsic motivation by building a learner who uses both imitation learning and SAGG-RIAC algorithm [22], and thus designed the SGIM-D (Socially Guided Intrinsic Motivation by Demonstration) algorithm [105]. We then investigated on the reasons of this bootstrapping effect [55], to show that demonstrations by teachers can both enhance more tasks to be explored, as well as favor more easily generalized actions to be used. This analysis is generalizable for all algorithms using social guidance and goal-oriented exploration. We then proposed to build a strategic learner who can learn multiple tasks and with multiple strategies. An overview and theoretical study of multi-task, multi-strategy Strategic Learning is presented in [47]. We also forsook to build a learning algorithm for more natural interaction with the human users. We first designed the SGIM-IM algorithm so that it can determine itself when it should ask for help from the teacher while trying to explore autonomously as long as possible so as to use as little of the teacher's time as possible [54]. After tackling with the problem of how and when to learn, we also investigated an active learner who can determine who to ask for help: in the case of two teachers available, SGIM-IM can determine which strategy to adopt between autonomous exploration and learning by demonstration, and which teacher enhances most learning progress for the learner [56], and ask him for help.

While the above results have been shown in simulation environments: of a simple deterministic air hockey game (fig. 20), and a stochastic fishing experiment with a real-time physical simulator (fig. 21), we are now building the experimental setup of the fishing experiment in order to carry out the experiments with naive users.

6.1.3.3. Adaptive task execution for implicit human-robot coordination

Participants: Ievgen Perederieiev, Manuel Lopes, Freek Stulp.

We began a project which goal is to study how computational models of multi-agent systems can be applied in situations where one agent is a human. We aim at applications where robots collaborate with humans for achieving complex tasks..

A very important capability for efficient collaborative work is the mutual agreement of a task and the ability to predict the behavior of others. We address such aspect by studying methods that increase the predictability of the robot actions. An efficient motor execution becomes the one that not just optimize speed and minimizes energy but also the one that improves the reliability of the team behavior. We are studying policy gradient methods and working on policy improvement algorithms (PI^2, CEM and $CMAES$). A feasibility study will consider a simple task between a robot and a person where the goal is to coordinate the way a set of three colored buttons is pressed.

6.1.3.4. Formalizing Imitation Learning

Participants: Thomas Cederborg, Pierre-Yves Oudeyer.

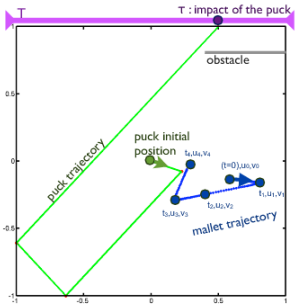


Figure 20. Illustration of SGIM-D and SGIM-IM algorithms

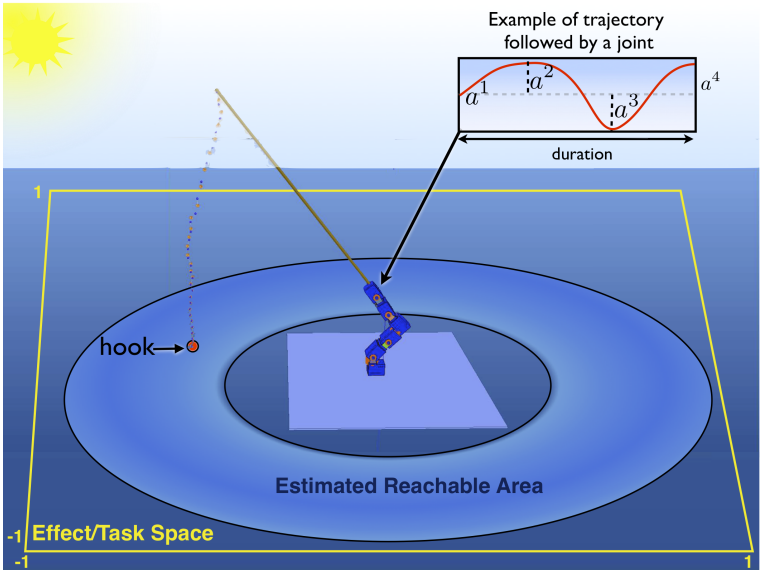


Figure 21. Illustration of SGIM-D and SGIM-IM algorithms

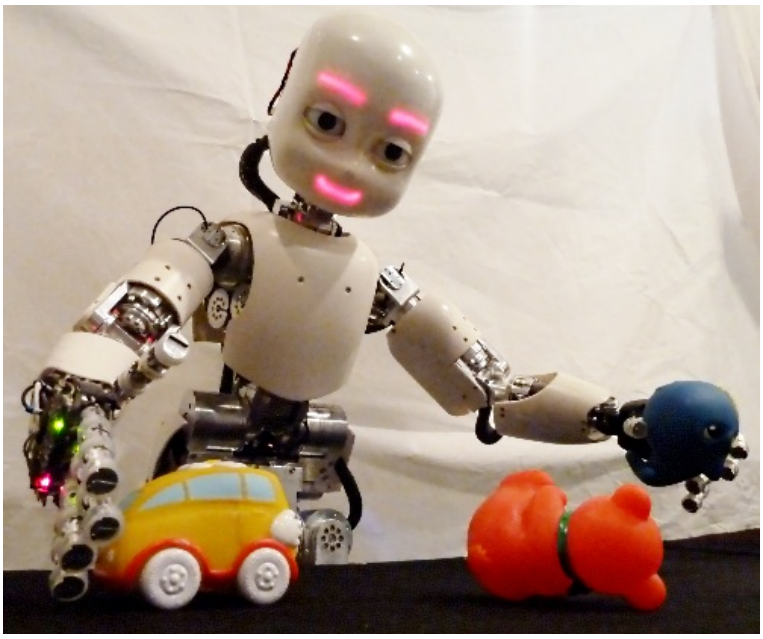


Figure 22. Illustration of SGIM-D and SGIM-IM algorithms

An original formalization of imitation learning was elaborated. Previous attempts to systematize imitation learning has been limited to categorizing different types of demonstrator goals (for example defining success in terms of the sequential joint positions of a dance, or in terms of environmental end states), and/or been limited to a smaller subset of imitation (such as learning from tele-operated demonstrations). The formalism proposed attempts to describe a large number of different types of learning algorithms using the same notation. Any type of algorithm that modifies a policy based on observations of a human, is treated as an interpretation hypothesis of this behavior. One example would be an update algorithm that updates a policy, partially based on the hypothesis that the demonstrator succeeds at demonstrations with probability 0.8, or an update algorithm that assumes that a scalar value is an accurate evaluation of an action compared to the latest seven actions. The formalism aims to give a principled way of updating these hypotheses, either rejecting some of a set of hypotheses regarding the same type of behavior, or set of parameters of an hypothesis. Any learning algorithm that modifies policy based on observations of a human that wants an agent to do something or act in some way, is describable as an interpretation hypothesis. If the learning algorithm is static, this simply corresponds to an hypothesis that is not updated based on observations. A journal article is currently being written.

6.1.4. Unsupervised learning of motor primitives

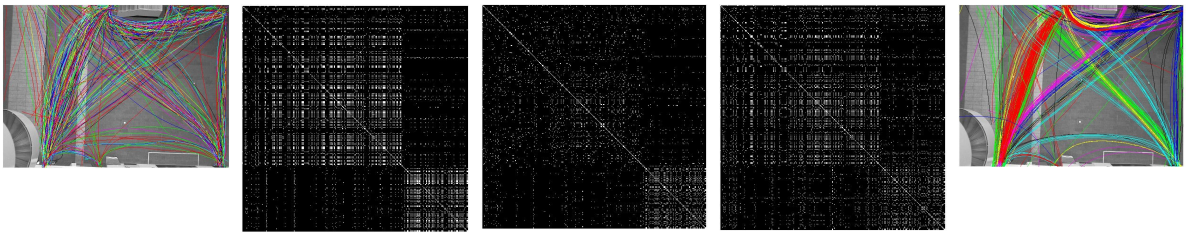
6.1.4.1. Clustering activities

Participants: Manuel Lopes, Luis Montesano.

Learning behaviors from data has applications in surveillance and monitoring systems, virtual agents and robotics among others. In our approach, ww assume that in a given unlabeled dataset of multiple behaviors, it is possible to find a latent representation in a controller space that allows to generate the different behaviors. Therefore, a natural way to group these behaviors is to search a common control system that generate them accurately.

Clustering behaviors in a latent controller space has two major challenges. First, it is necessary to select the control space that generate behaviors. This space will be parameterized by a set of features that will change for different behaviors. Usually, each controller will minimize a cost function with respect to several task features. The latent representation is in turn defined by the selected features and their corresponding weight. Second, an unknown number of such controllers is required to generate different behaviors and the grouping must be based on the ability of the controller to generate the demonstrations using a compact set of controllers.

We propose a Dirichlet Process based algorithm to cluster behaviors in a latent controller space which encodes the dynamical system generating the observed trajectories. The controller uses a potential function generated as a linear combination of features. To enforce sparsity and automatically select features for each cluster independently, we impose a conditional Laplace prior over the controller parameters. Based on this models, we derive a sparse Dirichlet Process Mixture Model (DPMM) algorithm that estimates the number of behaviors and a sparse latent controller for each of them based on a large set of features.



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

Figure 23. EIFPD dataset. (a) Trajectories of the EIFPD to be clustered (color is non-informative). (b-d) correspondence matrix for the 474 trajectories for the labeled ground truth, the KMeans in measurement space and the DPMM, respectively. (e) Reconstructed trajectories from the initial point using the estimated parameters of the DPMM algorithm. Due to the large number of clusters (37), colors are repeated for different clusters.

6.1.4.2. Learning the Combinatorial Structure of Demonstrated Behaviors with Inverse Feedback Control

Participants: Olivier Mangin, Pierre-Yves Oudeyer.

We have elaborated and illustrated a novel approach to learning motor skills from demonstration. This approach combines ideas from inverse feedback learning, in which actions are assumed to solve a task, and dictionary learning. In this work we introduced a new algorithm that is able to learn behaviors by assuming that the observed complex motions can be represented in a smaller dictionary of concurrent tasks. We developed an optimization formalism and show how we can learn simultaneously the dictionary and the mixture coefficients that represent each demonstration. We presented results on a idealized model where a set of potential functions represents human objectives or preferences for achieving a task in [51].

6.1.5. Maturational learning

6.1.5.1. Emergent Proximo-Distal Maturation through Adaptive Exploration

Participants: Freek Stulp, Pierre-Yves Oudeyer.

Life-long robot learning in the high-dimensional real world requires guided and structured exploration mechanisms. In this developmental context, we have investigated the use of the PI^2-CMAES episodic reinforcement learning algorithm, which is able to learn high-dimensional motor tasks through adaptive control of exploration. By studying PI^2-CMAES in a reaching task on a simulated arm, we observe two developmental properties. First, we show how PI^2-CMAES autonomously and continuously tunes the global exploration/exploitation trade-off, allowing it to re-adapt to changing tasks. Second, we show how PI^2-CMAES spontaneously self-organizes a maturational structure whilst exploring the degrees-of-freedom

(DOFs) of the motor space. In particular, it automatically demonstrates the so-called *proximo-distal maturation* observed in humans: after first freezing distal DOFs while exploring predominantly the most proximal DOF, it progressively frees exploration in DOFs along the proximo-distal body axis. These emergent properties suggest the use of PI^2-CMAES as a general tool for studying reinforcement learning of skills in life-long developmental learning contexts. This work was published in the IEEE International Conference on Developement and Learning [60].

6.1.5.2. Interaction of Maturation and Intrinsic Motivation for Developmental Learning of Motor Skills in Robots

Participants: Adrien Baranes, Pierre-Yves Oudeyer.

We have introduced an algorithmic architecture that couples adaptively models of intrinsic motivation and physiological maturation for autonomous robot learning of new motor skills. Intrinsic motivation, also called curiosity-driven learning, is a mechanism for driving exploration in active learning. Maturation denotes here mechanisms that control the evolution of certain properties of the body during development, such as the number and the spatio-temporal resolution of available sensorimotor channels. We argue that it is useful to introduce and conceptualize complex bidirectional interactions among these two mechanisms, allowing to actively control the growth of complexity in motor development in order to guide efficiently exploration and learning. We introduced a model of maturational processes, taking some functional inspiration from the myelination process in humans, and show how it can be coupled in an original and adaptive manner with the intrinsic motivation architecture SAGG-RIAC (Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity algorithm), creating a new system, called McSAGG-RIAC. We then conducted experiments to evaluate both qualitative and quantitative properties of these systems when applied to learning to control a high-dimensional robotic arm, as well as to learning omnidirectional locomotion in a quadruped robot equipped with motor synergies. We showed that the combination of active and maturational learning can allow to gain orders of magnitude in learning speed as well as reach better generalization performances. A journal article is currently being written.

6.1.6. Morphological computation and body intelligence

6.1.6.1. Comparative Study of the Role of Trunk in Human and Robot Balance Control

Participants: Matthieu Lapeyre [correspondant], Christophe Halgand, Jean-René Cazalet, Etienne Guillaud, Pierre-Yves Oudeyer.

Numerous studies in the field of functional motor rehabilitation were devoted to understanding the functioning of members but few are interested in the coordination of the trunk muscles and the relationship between axial and appendicular motricity which is essential in maintaining balance during travel. Acquiring new knowledge on this subject is a prerequisite in the development of new therapeutic strategies to restore motor function to the overall development of robotic orthosis that would assist the movement. Many robotic orthosis using EMG signals were unfortunately using few joints [85] and a system for controlling a multi articulated spine has not yet been developed. We propose here to use a multidisciplinary approach to define the neuro-mechanical principles where an axial system is operating in synergy with human and robot limbs.

To bring us a theoretical framework, we chose to study the reactions of the Acroban humanoid robot. Including 5 joints in the trunk, Acroban can reproduce in part the fluid movements of the human body [98] and especially to test its behavior when its trunk is held fixed or his arms are no longer used for rebalance. To disrupt postural balance in humans and robots, we have developed a low cost mobile platform (see Figure 24). This platform is made up of a broad stable support $(0.8x5m)$ mounted on a skateboard having a power of 800W. The substitution of the initial order of skate by an embedded microcontroller allows us to generate mono-axial perturbations precise intensity and duration to ensure repeatability of the disturbance. We capture movements (Optitrack 250Hz) and record the acceleration of the platform (accelerometer embedded 2kHz), the center of pressure (WiiBalanceBoard 60Hz), and electromyography (EMG).

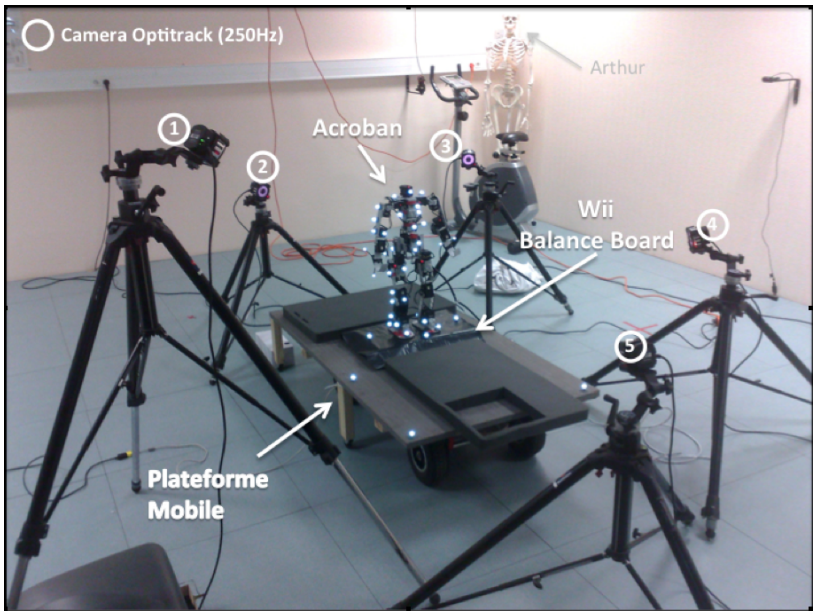


Figure 24. Experimental setup for comparative study of the role of the trunk in human and robot balance control

The experimental device (mobile platform and synchronized recordings) is operational. Preliminary experiments have allowed us to refine the profiles of disturbance on the robot Acroban. The analysis of preliminary results is in progress. Following this study, we hope to improve the modeling of the motor system in humans and robotic simulation as a basis for the development of robotic orthosis axial system. Second, the results provide a basis for improved balancing of Acroban primitives but also the development of future humanoid robots.

6.2. Autonomous and Social Perceptual Learning

6.2.1. The Impact of Human-Robot Interfaces on the Learning of Visual Objects

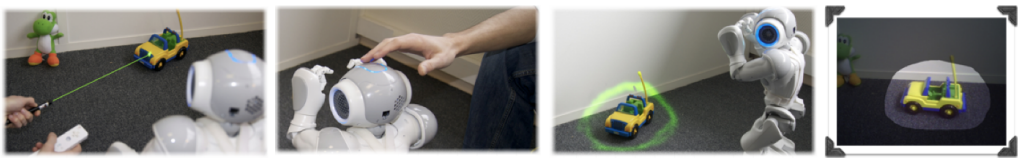
Participants: Pierre Rouanet, Pierre-Yves Oudeyer, Fabien Danieau, David Filliat.

We have continued and finalized a large-scale study of the impact of interfaces allowing non-expert users to efficiently and intuitively teach a robot to recognize new visual objects. We identified challenges that need to be addressed for real-world deployment of robots capable of learning new visual objects in interaction with everyday users. We argue that in addition to robust machine learning and computer vision methods, well-designed interfaces are crucial for learning efficiency. In particular, we argue that interfaces can be key in helping non-expert users to collect good learning examples and thus improve the performance of the overall learning system. Then, we have designed four alternative human-robot interfaces: three are based on the use of a mediating artifact (smartphone, wiimote, wiimote and laser), and one is based on natural human gestures (with a Wizard-of-Oz recognition system). These interfaces mainly vary in the kind of feedback provided to the user, allowing him to understand more or less easily what the robot is perceiving, and thus guide his way of providing training examples differently. We then evaluated the impact of these interfaces, in terms of learning efficiency, usability and user's experience, through a real world and large scale user study. In this experiment, we asked participants to teach a robot twelve different new visual objects in the context of a robotic game. This game happens in a home-like environment and was designed to motivate and engage users in an interaction where using the system was meaningful. We then analyzed results that show significant differences among

interfaces. In particular, we showed that interfaces such as the smartphone interface allows non-expert users to intuitively provide much better training examples to the robot, almost as good as expert users who are trained for this task and aware of the different visual perception and machine learning issues. We also showed that artifact-mediated teaching is significantly more efficient for robot learning, and equally good in terms of usability and user's experience, than teaching thanks to a gesture-based human-like interaction.
This work was accepted for publication in the IEEE Transactions on Robotics [28].

(a) draw the attention toward an object (b) trigger the capture (c) encircle the area of the object (d) the new learning example

Figure 25. Smartphone Interface. To make the robot collect a new learning example, users have to first draw the robot's attention toward the object they want to teach through simple gestures. Once the robot sees the object, they touch the head of the robot to trigger the capture. Then, they directly encircle the area of the image that represents the object on the screen. The selected area is then used as the new learning example. The combination of the video stream and the gestures facilitate the achievement of joint attention.



(a) draw the attention toward an object (b) trigger the capture (c) encircle the area of the object (d) the new learning example

Figure 26. Wiimote + laser pointer interface. With this interface users can draw the robot's attention with a laser pointer toward an object. The laser spot is automatically tracked by the robot. They can ensure that the robot detects the spot thanks to haptic feedback on the Wiimote. Then, they can touch the head of the robot to trigger the capture of a new learning example. Finally, they encircle the object with the laser pointer to delimit its area which will be defined as the new learning example.

6.2.2. Curiosity-driven exploration and interactive learning of visual objects with the ICub robot

Participants: Mai Nguyen, Serena Ivaldi, Natalia Lyubova, Alain Droniou, Damien Gerardeaux-Viret, David Filliat, Vincent Padois, Olivier Sigaud, Pierre-Yves Oudeyer.

We studied how various mechanisms for cognition and learning, such as curiosity, action selection, imitation, visual learning and interaction monitoring, can be integrated in a single embodied cognitive architecture. We have conducted an experiment with the iCub robot for active recognition of objects in 3D through curiosity-driven exploration, in which the robot can manipulate the robot or ask a human user to manipulate objects to gain information and recognise better objects (fig. 22). For this experiment carried out within the MACSi project, we address the problem of learning to recognise objects in a developmental robotics scenario. In a life-long learning perspective, a humanoid robot should be capable of improving its knowledge of objects

Figure 27. The real world environment designed to reproduce a typical living room. Many objects were added in the scene in order to make the environment cluttered.

with active perception. Our approach stems from the cognitive development of infants, exploiting active curiosity-driven manipulation to improve perceptual learning of objects. These functionalities are implemented as perception, control and active exploration modules as part of the Cognitive Architecture of the MACSi project. We integrated a bottom-up vision system based on swift feature points and motor-primitive based robot control with the SGIM-ACTS algorithm (Socially Guided Intrinsic Motivation with Active Choice of Task and Strategy as the active exploration module. SGIM-ACTS is a strategic learner who actively chooses which task to concentrate on, and which strategy is better according to this task. It thus monitors the learning progress for each strategy on all kinds of tasks, and actively interacts with the human teacher. We obtained an active object recognition approach, which exploits curiosity to guide exploration and manipulation, such that the robot can improve its knowledge of objects in an autonomous and efficient way. Experimental results show the effectiveness of our approach: the humanoid iCub is now capable of deciding autonomously which actions must be performed on objects in order to improve its knowledge, requiring a minimal assistance from its caregiver. This work constitutes the base for forthcoming research in autonomous learning of affordances.

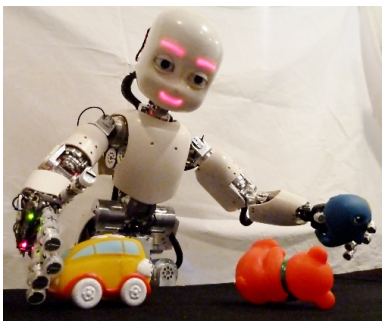


Figure 28. iCub performing curiosity-driven exploration and active recognition of visual objects in 3D

6.2.3. Discovering object concept through developmental learning

Participants: Natalia Lyubova, David Filliat.

The goal of this work is to design a visual system for a humanoid robot. Taking inspiration from child perception and following the principles of developmental robotics, the robot should detect and learn objects from interactions with people and from experiments it performs with objects, avoiding the use of image databases or of a separate training phase. In our model, all knowledge is therefore iteratively acquired from low-level features and builds up hierarchical object models, which are robust to changes in the environment, background and camera motion. In our scenario, people in front of the robot are supposed to interact with objects to encourage the robot to focus on them. We therefore assume that the robot is attracted by motion and we segment possible objects based on clustering of the optical flow. Additionally, the depth information from a Kinect is used to filter visual input, considering the constraints of the robot's working area and to refine the object contours obtained from motion segmentation.

The appearance of objects is encoded following the Bag of Visual Words approach with incremental dictionaries. We combine several complementary features to maximize the completeness of the encoded information (SURF descriptor and superpixels with associated colors) and construct pairs and triples of these features to integrate local geometry information. These features make it possible to decide if the current view has been already seen or not. A multi-view object model is then constructed by associating recognized views and views tracked during manipulations with an object.

This system is implemented on the iCub humanoid robot, which detects objects in the visual space and characterizes their appearance, their relative position and their occurrence statistics. The experiments were performed with up to ten objects; each of them was manipulated by a person during 1-2 minutes. Once the vocabulary reached a sufficient amount of knowledge, the robot was able to reliably recognize most of objects [48], [49], [43].

6.2.4. *Unsupervised object categorization*

Participants: Natalia Lyubova, David Filliat.

The developed unsupervised algorithm allows to identify segmented units of attention based on motion and depth information (proto-objects) into different categories such as robot hands, objects and humans.

The robot self-body category is discovered from the correlation between the proto-object positions and proprioception on the robot arms. This correlation it estimated by computing the mutual information between the changes in robot motor joints and the motion behavior of proto-objets in the visual field. The arm joints states are recorded from the robot and quantized to a vocabulary of possible arm configurations. The visual space is analyzed at the level of visual clusters that divide the perception field into regular regions. The mutual information is computed from the occurrence probabilities of the arm configurations and visual clusters.

In case of high correlation, the visual cluster is identified as a robot hand. Among the remaining proto-objects, objects are distinguished from human hands based on their quasi-static nature. Since most of objects don't move by themselves but rather are displaced by external forces, the object category is associated with regions of the visual space moving together mostly with recognized robot hands or human parts. This process make it possible to recognize the robot hands, even in case of changing appearance, and to learn to separate objects from parts of the caregivers bodies.

6.2.5. *Efficient online bootstrapping of sensory representations*

Participant: Alexander Gepperth.

This work [24] is a simulation-based investigation exploring a novel approach to the open-ended formation of multimodal representations in autonomous agents. In particular, we addressed here the issue of transferring (bootstrapping) features selectivities between two modalities, from a previously learned or innate reference representation to a new induced representation. We demonstrated the potential of this algorithm by several experiments with synthetic inputs modeled after a robotics scenario where multimodal object representations are bootstrapped from a (reference) representation of object affordances, focusing particularly on typical challenges in autonomous agents: absence of human supervision, changing environment statistics and limited computing power. We proposed an autonomous and local neural learning algorithm termed PROPRE (projection-prediction) that updates induced representations based on predictability: competitive advantages

are given to those feature-sensitive elements that are inferable from activities in the reference representation, the key ingredient being an efficient online measure of predictability controlling learning. We verified that the proposed method is computationally efficient and stable, and that the multimodal transfer of feature selectivity is successful and robust under resource constraints. Furthermore, we successfully demonstrated robustness to noisy reference representations, non-stationary input statistics and uninformative inputs.

6.2.6. Simultaneous concept formation driven by predictability

Participants: Alexander Gepperth, Louis-Charles Caron.

This work [40] was conducted in the context of developmental learning in embodied agents who have multiple data sources (sensors) at their disposal. We developed an online learning method that simultaneously discovers meaningful concepts in the associated processing streams, extending methods such as PCA, SOM or sparse coding to the multimodal case. In addition to the avoidance of redundancies in the concepts derived from single modalities, we claim that meaningful concepts are those who have statistical relations across modalities. This is a reasonable claim because measurements by different sensors often have common cause in the external world and therefore carry correlated information. To capture such cross-modal relations while avoiding redundancy of concepts, we propose a set of interacting self-organization processes which are modulated by local predictability. To validate the fundamental applicability of the method, we conducted a plausible simulation experiment with synthetic data and found that those concepts which are predictable from other modalities successively ”grow”, i.e., become overrepresented, whereas concepts that are not predictable become systematically under-represented. We additionally explored the applicability of the developed method to real-world robotics scenarios.

6.2.7. The contribution of context: a case study of object recognition in an intelligent car

Participants: Alexander Gepperth, Michael Garcia Ortiz.

In this work [23], we explored the potential contribution of multimodal context information to object detection in an ”intelligent car”. The used car platform incorporates subsystems for the detection of objects from local visual patterns, as well as for the estimation of global scene properties (sometimes denoted scene context or just context) such as the shape of the road area or the 3D position of the ground plane. Annotated data recorded on this platform is publicly available as the a ”HRI RoadTraffic” vehicle video dataset, which formed the basis for the investigation. In order to quantify the contribution of context information, we investigated whether it can be used to infer object identity with little or no reference to local patterns of visual appearance. Using a challenging vehicle detection task based on the ”HRI RoadTraffic” dataset, we trained selected algorithms (context models) to estimate object identity from context information alone. In the course of our performance evaluations, we also analyzed the effect of typical real-world conditions (noise, high input dimensionality, environmental variation) on context model performance. As a principal result, we showed that the learning of context models is feasible with all tested algorithms, and that object identity can be estimated from context information with similar accuracy as by relying on local pattern recognition methods. We also found that the use of basis function representations [1] (also known as ”population codes” allows the simplest (and therefore most efficient) learning methods to perform best in the benchmark, suggesting that the use of context is feasible even in systems operating under strong performance constraints.

6.2.8. Co-training of context models for real-time object detection

Participant: Alexander Gepperth.

In this work[41], we developed a simple way to reduce the amount of required training data in context-based models of real- time object detection and demonstrated the feasibility of our approach in a very challenging vehicle detection scenario comprising multiple weather, environment and light conditions such as rain, snow and darkness (night). The investigation is based on a real-time detection system effectively composed of two trainable components: an exhaustive multiscale object detector (signal-driven detection), as well as a module for generating object-specific visual attention (context models) controlling the signal-driven detection process. Both parts of the system require a significant amount of ground-truth data which need to be generated by

human annotation in a time-consuming and costly process. Assuming sufficient training examples for signal-based detection, we showed that a co-training step can eliminate the need for separate ground-truth data to train context models. This is achieved by directly training context models with the results of signal-driven detection. We demonstrated that this process is feasible for different qualities of signal-driven detection, and maintains the performance gains from context models. As it is by now widely accepted that signal-driven object detection can be significantly improved by context models, our method allows to train strongly improved detection systems without additional labor, and above all, cost.

6.3. Joint Learning and Development of Language and Action

6.3.1. Learning to recognize parallel motion primitives with linguistic descriptions using Non-Negative Matrix Factorization

Participants: Olivier Mangin, Pierre-Yves Oudeyer.

We have elaborated and experimented a novel approach to joint language and motor learning from demonstration. It enables discovery of a dictionary of gesture and linguistic primitives, that can be combined in parallel to represent training data as well as novel activities in the form of combinations of known gestures. These methods and the results of our experiments participate in addressing two main issues of developmental robotics: 1) symbol grounding for language learning; 2) achieving compositionality in motor-learning from demonstration, which enables re-using knowledge and thus scaling to complex tasks. In particular, we are interested in learning motor primitives active in parallel, a less explored way of combining such primitives. To address these challenges we have explored and studied the use of nonnegative matrix factorization to discover motor primitives from histogram representations of data acquired from real demonstrations of dancing movements. Initial results were presented in [99] and further results are presented in [52].

6.3.2. Curiosity-driven phonetic learning

Participants: Clément Moulin-Frier, Pierre-Yves Oudeyer.

We study how developmental phonetic learning can be guided by pure curiosity-driven exploration, also called intrinsically motivated exploration. Phonetic learning refers here to learning how to control a vocal tract to reach acoustic goals. We compare three different exploration strategies for learning the auditory-motor inverse model: random motor exploration, random goal selection with reaching, and curiosity-driven active goal selection with reaching. Using a realistic vocal tract model, we show how intrinsically motivated learning driven by competence progress can generate automatically developmental structure in both articulatory and auditory modalities, displaying patterns in line with some experimental data from infants. This work has been published in [53] and received the best paper award in computational models of development at the International Conference on Development and Learning, Epirob, San Diego, 2012.

We are now working on applying this approach to the control of a more complex articulatory synthesizer. We are interested in using the free software Praat, a powerful tool allowing to synthesize a speech signal from a trajectory in a 29-dimensional space of respiratory and oro-facial muscles. Numerous acoustic features can in turn be extracted from the synthesized sound, among which the Mel-frequency cepstral coefficients. Our hope is that a developmental robotics approach applied to a realistic articulatory model can appropriately manage the learning process of this complex mapping in high-dimensional spaces , and that observed developmental sequences can lead to interesting experimental data comparisons and predictions. In particular, using such a dynamic model controlled by muscle activity could hopefully allow to relate our results to more common speech acquisition data, in particular regarding infraphonological exploration and babbling.

6.3.3. Towards robots with teleological action and language understanding

Participants: Britta Wrede, Katharina Rohlfing, Jochen Steil, Sebastian Wrede, Jun Tani, Pierre-Yves Oudeyer.

It is generally agreed upon that in order to achieve generalizable learning capabilities of robots they need to be able to acquire compositional structures - whether in language or in action. However, in human development the capability to perceive compositional structure only evolves at a later stage. Before the capability to understand action and language in a structured, compositional way arises, infants learn in a holistic way which enables them to interact in a socially adequate way with their social and physical environment even with very limited understanding of the world, e.g. trying to take part in games without knowing the exact rules. This capability endows them with an action production advantage which elicits corrective feedback from a tutor, thus reducing the search space of possible action interpretations tremendously. In accordance with findings from developmental psychology we argue that this holistic way is in fact a teleological representation encoding a goal-directed perception of actions facilitated through communicational frames. This observation leads to a range of consequences which need to be verified and analysed in further research. We have written an article [64] where we discussed two hypotheses how this can be made accessible for action learning in robots: (1) We explored the idea that the teleological approach allows some kind of highly reduced one shot learning enabling the learner to perform a meaningful, although only partially correct action which can then be further refined through compositional approaches. (2) We discussed the possibility to transfer the concept of ”conversational frames” as recurring interaction patterns to the action domain, thus facilitating to understand the meaning of a new action. We conclude that these capabilities need to be combined with more analytical compositional learning methods in order to achieve human-like learning performance.

6.3.4. Imitation Learning and Language

Participants: Thomas Cederborg, Pierre-Yves Oudeyer.

We have studied how context-dependant imitation learning of new skills and language learning could be seen as special cases of the same mechanism. We argue that imitation learning of context-dependent skills implies complex inferences to solve what we call the ”motor Gavagai problem”, which can be viewed as a generalization of the so-called ”language Gavagai problem”. In a full symbolic framework where percepts and actions are continuous, this allows us to articulate that language may be acquired out of generic sensorimotor imitation learning mechanisms primarily dedicated at solving this motor Gavagai problem. Through the use of a computational model, we illustrate how non-linguistic and linguistic skills can be learnt concurrently, seamlessly, and without the need for symbols. We also show that there is no need to actually represent the distinction between linguistic and non-linguistic tasks, which rather appears to be in the eye of the observer of the system. This computational model leverages advanced statistical methods for imitation learning, where closed-loop motor policies are learnt from human demonstrations of behaviours that are dynamical responses to a multimodal context. A novelty here is that the multimodal context, which defines what motor policy to achieve, includes, in addition to physical objects, a human interactant which can produce acoustic waves (speech) or hand gestures (sign language). A book chapter was written and published [66] and a journal article was submitted.

6.3.5. COSMO (“Communicating about Objects using Sensory-Motor Operations”): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems

Participants: Clément Moulin-Frier, Jean-Luc Schwartz, Julien Diard, Pierre Bessière.

This work began with the PhD thesis of Clement Moulin-Frier at GIPSA-Lab, Grenoble, France, supervised by Jean-Luc Schwartz (GIPSA-Lab, CNRS), Julien Diard (LPNC, CNRS) and Pierre Bessière (College de France, CNRS). A few papers were finalized during his post-doc at FLOWERS in 2012. Firstly, an international journal paper based on the PhD thesis work of Raphael Laurent (GIPSA-Lab), extending Moulin-Frier’s model, was published [25], and a commentary in *Behavioral and Brain Sciences* was accepted but not yet published [68]. Both these papers provide computational arguments based on a sensory-motor cognitive model to feed the age-old debate of motor vs. auditory theories of speech perception. Secondly, in another journal paper under the submission process, we attempt to derive some properties of phonological systems (the sound systems of human languages) from the mere properties of speech communication. We introduce a model of the cognitive

architecture of a communicating agent, called COSMO (for "Communicating about Objects using Sensory-Motor Operations") that allows expressing in a probabilistic way the main theoretical trends found in the speech production and perception literature. This allows a computational comparison of these theoretical trends, helping to identify the conditions that favor the emergence of linguistic codes. We present realistic simulations of phonological system emergence showing that COSMO is able to predict the main regularities in vowel, stop consonant and syllable systems in human languages.

6.3.6. Recognizing speech in a novel accent: the Motor Theory of Speech Perception reframed

Participants: Clément Moulin-Frier, Michael Arbib.

Clément Moulin-Frier engaged this work with Michael Arbib during his 6-month visit in 2009 at the USC Brain Project, University of Southern California, Los Angeles, USA. An international journal paper is still under the revision process, in which we offer a novel computational model of foreign-accented speech adaptation, together with a thorough analysis of its implications with respect to the motor theory of speech perception.

6.3.7. Learning Simultaneously New Tasks and Feedback Models in Socially Guided Robot Learning

Participants: Manuel Lopes, Jonathan Grizou, Thomas Cederborg, Pierre-Yves Oudeyer.

We have developed a system that allows a robot to learn simultaneously new tasks and feedback models from ambiguous feedback in the context of robot learning by imitation. We have considered an inverse reinforcement learner that receives feedback from a user with an unknown and noisy protocol. The system needs to estimate simultaneously what the task is, and how the user is providing the feedback. We have further explored the problem of ambiguous protocols by considering that the words used by the teacher have an unknown relation with the action and meaning expected by the robot. This allows the system to start with a set of known symbols and learn the meaning of new ones. We have conducted human-robot interaction experiments where the user teaches a robot new tasks using natural speech with words unknown to the robot. The robot needs to estimate simultaneously what the task is and the associated meaning of words pronounced by the user. We have computational results showing that: a) it is possible to learn the task under unknown and noisy feedback, b) it is possible to reuse the acquired knowledge for learning new tasks and c) even in the presence of a known feedback, the use of extra unknown feedback signals while learning improves learning efficiency and robustness to mistakes. This algorithm has been applied on discrete and continuous problems and tested in a real world experiment using spoken words as feedback signals. A article to be submitted to a journal is currently being written.

6.3.8. Active Learning for Teaching a Robot Grounded Relational Symbols

Participants: Johannes Kulick, Tobias Lang, Marc Toussaint, Manuel Lopes.

The present work investigates an interactive teaching scenario, where a human aims to teach the robot symbols that abstract geometric (relational) features of objects. There are multiple motivations for this scenario: First, state-of-the-art methods for relational Reinforcement Learning demonstrated that we can successfully learn abstracting and well-generalizing probabilistic relational models and use them for goal-directed object manipulation. However, these methods rely on given grounded action and state symbols and raise the classical question Where do the symbols come from? Second, existing research on learning from human-robot interaction has focused mostly on the motion level (e.g., imitation learning). However, if the goal of teaching is to enable the robot to autonomously solve sequential manipulation tasks in a goal-directed manner, the human should have the possibility to teach the relevant abstractions to describe the task and let the robot eventually leverage powerful relational RL methods (see Figure 29). We formalize human-robot teaching of grounded symbols as an Active Learning problem, where the robot actively generates geometric situations that maximize his information gain about the symbol to be learnt. We demonstrate that the learned symbols can be used in a relational RL framework for the robot to learn probabilistic relational rules and use them to solve object manipulation tasks in a goal-directed manner. [44].

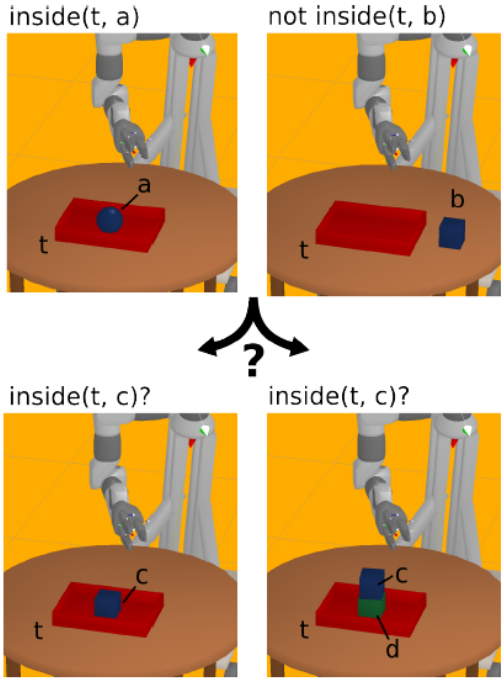


Figure 29. Active learning of symbol descriptions on a real world robot.

6.3.9. Multimodal Conversational Interaction with a Humanoid Robot

Participants: Adam Csapo, Emer Gilmartin, Jonathan Grizou, JingGuang Han, Raveesh Meena, Dimitra Anastasiou, Kristiina Jokinen, Graham Wilcock.

The paper presents a multimodal conversational interaction system for the Nao humanoid robot. The system was developed at the 8th International Summer Workshop on Multi-modal Interfaces, Metz, 2012. We implemented WikiTalk, an existing spoken dialog system for open-domain conversations, on Nao. This greatly extended the robot's interaction capabilities by enabling Nao to talk about an unlimited range of topics. In addition to speech interaction, we developed a wide range of multimodal interactive behaviours by the robot, including face- tracking, nodding, communicative gesturing, proximity detection and tactile interrupts. We made video recordings of user interactions and used questionnaires to evaluate the system. We further extended the robot's capabilities by linking Nao with Kinect. This work was presented in [34].

6.4. Other applications

6.4.1. Real-time Reaction-Diffusion Simulation: a Machine Learning Technique

Participants: Thomas Degris, Nejib Zemzemi.

Carmen is an Inria team working on modeling the electrical activity of the human heart. Their models are mainly based on reaction-diffusion equations. These methods are expansive in terms of computational costs which limits their use in practice. More specifically, some recent chirurgical intervention techniques on the heart (atrial ablation) requires to identify the source of the electrical wave. Finding such sources requires an optimization procedure. Using classical methods, this procedure is very heavy computationally.

In this project, our goal is to reduce the computational cost using supervised learning techniques. The idea is to replace the incremental resolution of partial differential equations by more suitable data structures for real-time running. Starting from data generated by simulating different excitations scenari on a human atria, this data is afterwords used as a training data set for machine learning algorithms. This approach will allow a faster optimization procedure.

This work is in collaboration with Nejib Zemzemi from the Inria Carmen team. This project is in preliminary steps.

IMARA Project-Team

6. New Results

6.1. Low speed automation

Participants: Paulo Lopes Resende, Fawzi Nashashibi, Hao Li, Evangeline Pollard.

The ABV project builds on the HAVEit philosophy (a previous IMARA project for high speed automation) of offering higher levels of automation on highways and organizing the cooperation between human and system along novel automation levels. It differs from HAVEit by focusing on congested traffic at speeds below 50 km/h and adding fully automated driving to the automation spectrum. By automatically following congested traffic, the ABV system relieves the human driver from monotonous tasks. During fully automated driving, the human driver is not required to monitor the system, but has to take over control at the end of the application zone.

6.2. Urban autonomous driving: dealing with intersections

Participants: Guillaume Tréhard, Evangeline Pollard, Fawzi Nashashibi.

The goal of this project, made in collaboration with Valeo is to develop a complete solution for autonomous driving on open roads. More specifically, IMARA's objectives are to provide the way to safely cross any kind of intersections for an autonomous vehicle in a urban context. Among the different relevant scenarios, we can notice:

- Intersection with different shapes: Roundabout, T junctions , X junctions;
- Intersection with different rules: With specific rules (traffic lights, main road...) or unspecified rules ("priority to the right");
- Different traffic: Busy or empty intersections;
- Deal with abnormal situations: road works, policemen, firemen,...

Possible steps for this work can be listed as follows:

- Model the intersection: define relevant information, find a generic model for every intersection;
- Detect the intersection (shape, drivable area, traffic flows);
- Understand the priorities that rules it;
- Locate the car in the intersection by crossing it;
- Plan a path to get out of the intersection.

6.3. Conception of a new communicative system for the protection of vulnerable people

Participants: Pierre Merdrignac, Evangeline Pollard, Oyunchimeg Shagdar, Fawzi Nashashibi.

A new research has been recently launched at IMARA team. The goal is to elaborate a new communicative system between vulnerable people (pedestrian, person with reduced mobility, bicyclist, etc.) and intelligent vehicles in order to improve safety and to limit collision risk. The main idea of this project is as follows. Intelligent vehicles are equipped with an obstacle detection/classification /tracking module in order to prevent injuries. On the other hand, to help the driver in this challenging task, vulnerable people use an application on their mobile phone to inform/share their status on location, type, and dynamics. The status information is transmitted to the driver utilizing wireless communications technology (e.g., 3G and Wi-Fi). In the vehicle, information coming from the communications device and obstacle detection module will be merged to improve the detection and classification tasks. In case of emergency, the vehicle can broadcast safety information to vulnerable people.

6.4. Visible Light V2V Communications for Platooning Control

Participants: Mohammad Abu Alhoul, Oyunchimeg Shagdar, Mohamed Marouf, Fawzi Nashashibi.

Fully automated vehicles have the potential to greatly improve the comfort of humans' life. For driving from one place to another, an automated vehicle must avoid collisions and be able to select non-congested roads for safe and efficient driving. In order to do that the vehicle needs to control its mobility in both macroscopic and microscopic levels by utilizing information exchange with other vehicles and roadside infrastructures based on wireless communications technology. While radio frequency channel is a convincing choose for vehicular communications due to its high data rate over relatively long coverage range (minimum several 100's meters), it is expected to experience channel congestion and low communication reliability especially for the scenario where there is high-density vehicles. In such scenarios vehicles still require to control the mobility on both the macroscopic and microscopic levels, we need to look for supportive and at the same time practical communication media with the ability to support sufficient connection between vehicles. According to the latest standard from IEEE, 802.15.7 for 2011, the communication coverage of Visible Light Communication (VLC) can reach up to 10's of meters, which match the information exchange requirements for mobility control in microscopic level. Motivated by this we started our research activity on modeling of visible light communications channel and design of microscopic mobility control, specifically platooning control, using VLC.

6.5. Augmented reality for the protection of vulnerable people

Participants: Hao Li, Fawzi Nashashibi.

A brand new idea of cooperative augmented reality is under development in IMARA team. It utilizes the results of cooperative local mapping to realize certain augmented reality effect. More specifically, the idea is to obtain an augmented effect of "seeing" through front vehicle, based on the intelligent vehicle sensor configurations.

Given a scenario of two vehicles: a front (first) vehicle and a following (second) vehicle. This front-following vehicles scenario is typical in traffic environment and is potentially dangerous, especially in some occasions such as during an overtaking, where the front vehicle occludes a part of the scene to the following vehicle. The idea of cooperative augmented reality is thus to project the visual perception of the front vehicle onto that of the following vehicle, abiding by perspective geometry. In other words, we patch the occluded part of the view of the following vehicle with corresponding part of the view of the front vehicle. This is not simply a process of partial view copying and pasting between the two vehicles; we have to transform the partial view of the front vehicle according to perspective geometry, in order to make a vivid and natural reproduction of this partial view for the following vehicle, just like if the following vehicle can directly see into the occluded area.

A prerequisite for performing the perspective transformation between the visual perceptions of the two vehicles is the knowledge of the visual perception depth. This knowledge can be estimated by stereo-vision, if correct correspondence is established (yet a challenging process) between the images pair in stereo-vision. However, approximate estimate of the visual perception depth was obtained with the help of 2D range perception in an innovative way and indirect vehicle-to-vehicle relative pose estimation method introduced in [36].

6.6. Step detection for Personal Mobility Vehicles

Participants: Evangeline Pollard, Joshué Pérez Rastelli, Fawzi Nashashibi.

Personal Mobility Vehicles (PMV) is an important part of the Intelligent Transportation System (ITS) domain. These new transport systems have been designed for urban traffic areas, pedestrian streets, green zones and private parks. In these areas, steps and curbs make the movement of disable or mobility reduced people with PMV, and with standard chair wheels difficult. In this work, a step and curb detection system based on laser sensors has been developed. This system is dedicated to vehicles able to cross over steps, for transportation systems, as well as for mobile robots. The system is composed of three laser range finders. Hokuyo UTM 30 LX devices were chosen for their large detection angle (270°) and their high angular resolution (0.25°) and range (30m).

Two laser sensors dedicated to the step detection have a vertical orientation in order to scan the altitude profile of the environment over two lines of sight and the third one, with a vertical orientation is dedicated to obstacle detection.

The step detection process is thus based on the study of the first derivative of the altitude and highlights the use of a new algebraic derivative method (Alien) adapted to laser sensor data. The system has been tested on several real scenarios. It provides the distance, altitude and orientation of the steps in front of the vehicle and offers a high level of precision, even with small steps.

6.7. PROSIVIC-CTS simulator

Participant: Joshué Pérez Rastelli.

The Architecture validation and experiments presented in this document have been implemented in a simulated environment, called ProSiVIC, which allows implementing a virtual Cybercars, among other vehicles. The algorithms are the same as in our Cybus platform, using the position from the SLAM and DGPS sensors. The ProSivic simulator offers a multi-sensorial environment, and takes into account several parameters of a real car such as the inertia, steering wheel response, lateral acceleration with yaw angles, damping suspension, simple weather conditions, friction parameters and more.

Moreover, synchronized time, acceleration (in wheel torque), steering, odometer information, lidar information and camera viewports are some of the components supporting the connection between the control architecture in RTMaps and the simulation.

The simulations show the behavior of the control architecture implemented for CTSs. Two urban scenarios were tested: roundabouts and intersections.

6.8. Autonomous docking based on infrared system for electric vehicle charging in urban areas

Participants: Benjamin Lefaudeux, Joshué Pérez Rastelli, Fawzi Nashashibi.

One of the recent aims of the Intelligent Transportation Systems (ITS) is the reduction of air pollution, reducing the fuel consumption in urban areas and improving road security. To this purpose, electric vehicles are a good and high demanded alternative. Nowadays, some big cities are launching the first electric car-sharing projects to clear its traffic jam, as an alternative to the classic public transportation systems. However, there are still some problems related to energy storage, charging and autonomy to be solved. To tackle this problem in the context of the French project AMARE, IMARA has developed an autonomous docking system, based on an infrared camera embarked in a vehicle equipped with dedicated ADAS, and some infrared diodes installed in the infrastructure, for recharging the vehicle batteries in a street parking area. The results obtained show a good behavior of the implemented system, which is working in a real scenario in the city of Paris.

Different experiments, departing from different points, show a good behavior of the proposed systems. Both lateral and longitudinal errors are lower that the limits of the charging station. The controller used is easy and intuitive for tuning, and the gains can be adapted according of the different vehicles characteristics. This technology permits to assist to human drivers in the charging process of electric vehicles in cities.

6.9. Reasoning for relaxing traffic regulations

Participants: Philippe Morignot, Fawzi Nashashibi.

This work [39] deals with relaxation of traffic rules in unusual but practical situations. For example, if a truck is unloading on a roadway, the automated vehicle should overtake it despite a continuous yellow line: traffic rules are indeed broken, which is illegal, but this might be tolerated due to the unusual aspect of the situation at hand.

An ontology has been developed in order to represent the road network (a directed graph, vertices being intersections and edges being lanes), the infrastructure (road signs, marks), the other road users and the intelligent vehicle. Reasoning on this representation is performed by inference rules (IF/THEN symbolic structures), encoding the deliberation on the encoded situation. Main rules conclude on the next discrete motion of the vehicle, e.g., "pass onto the adjacent lane" which involves crossing a continuous yellow line.

In practice, this ontology has been created using the PROTEGE ontology editor from Stanford University. IF/THEN rules are represented in SWRL (Semantic Web Rule Language), using the reasoner PELLET from the company Clark & Parsia (a plug-in of the tool PROTEGE).

Work over the next year involves porting this reasoning module on the vehicles: porting the generated JAVA source code as one component inside the RTMAPS architecture of CyberCars, and linking the ABoxes (assertional boxes) of the ontology to symbols extracted from signals by perception.

6.10. Communications and Management Control for Cooperative Vehicular Systems

Participants: Ines Ben Jemaa, Oyunchimeg Shagdar, Fawzi Nashashibi, Arnaud de La Fortelle.

One of the attractive applications of electric autonomous vehicles is electric automated Car Sharing service. In this application, a user requests a vehicle at a given geographical location triggering the car sharing system to allocate an autonomous vehicle for the user transport from the station to the user's desired destination. The application requires efficient cooperation among the autonomous vehicles and a service management centre for reliable and responsive car sharing service. Such cooperation is not possible unless vehicles exchange their information on e.g., position, motion, and coordination messages among themselves and with central management entities. While the existing wireless communications technologies can be applied for vehicle to vehicle and vehicle to the infrastructure communications, important research challenges remain including network partitioning problem caused by vehicles' mobility and inability of the convergence of geographically scoped V2V and Internet-based V2I communications. Targeting these issues, we study a topology control solution to tackle the network portioning problem and design of unicast/multicast/Geonetworking schemes for convergence of V2V and V2I communications systems for car-sharing applications [28].

6.11. New urban transportation platforms: Inria's Cybus

Participants: Laurent Bouraoui, François Charlot, Fawzi Nashashibi, Paulo Lopes Resende, Michel Parent, Armand Yvet.

Cybus is the newest prototyping and demonstration platform designed at Inria. Apart from the chassis and engines, the whole hardware and software systems were developed thanks to IMARA's researchers and engineers talents. These electric vehicles are based on a Yamaha chassis but the embedded intelligence is the result of two years of development. Much of the perception and control software has been improved. New guidance functionalities were developed this year, mainly with the introduction of stereovision-based SLAM.

The platforms developed here (Cybus) will be demonstrated in the context of the EU CityMobil-2 project. This time real operational mobility services demonstrations will be extended to 6-12 months in selected European cities! Other showcases are expected to take place in Asian cities in 2014.

6.12. Belief propagation inference for traffic prediction

Participants: Cyril Furtlehner, Yufei Han, Jean-Marc Lasgouttes, Victorin Martin.

This work [57] deals with real-time prediction of traffic conditions in a setting where the only available information is floating car data (FCD) sent by probe vehicles. The main focus is on finding a good way to encode some coarse information (typically whether traffic on a segment is fluid or congested), and to decode it in the form of real-time traffic reconstruction and prediction. Our approach relies in particular on the belief propagation algorithm.

Figure 1. The Cybus operated at La Rochelle City during 3 months as a free transport service.

These studies are done in particular in the framework of the projects Travesti and Pumas.

This year's highlights are

- A paper describing a new sufficient condition for local stability of the Belief Propagation algorithm has been published and presented in an international conference [38].
- The work about the theoretical aspects of encoding real valued variables into a binary Ising model has been summarized in a publication currently under reviewing process.
- Ideas about adding macroscopic variables within the Ising model are currently being tested using the software BPstruction developed last year.
- Victorin Martin has given a talk at the "Colloque Jeunes Probabilistes et Statisticiens" at CIRM, where he presented his work one the design of a latent Ising model for real valued inference.
- Cyril Furtlehner, Yufei Han and Victorin Martin presented the work done in the Travesti project at the workshop on inference organized by Inria and Mines ParisTech (see 9.1).

6.13. Non-negative Tensor factorization for spatio-temporal data analysis

Participant: Yufei Han.

This is a joint work with Fabien Moutarde from Mines ParisTech.

We investigate the use of non-negative tensor factorization for spatio-temporal data clustering and prediction. In general case, a spatio-temporal signal is represented as a set of multiple-variant temporal sequences. In the domain of intelligent traffic, the temporal records of traffic flow states (free-flowing/congestion) over a specific time duration with respect to hundreds of links in a transportation network can be considered as a simple but direct example of spatio-temporal signal. Both temporal causality between neighboring time sampling steps and spatial layout of the multiple-variant observation captured at each time sampling step are the focus of the spatio-temporal data analysis. Non-negative tensor factorization enables us to project the high dimensional spatio-temporal data into low-dimensional subspace and clustering/prediction can be then achieved on the derived subspace projection easily.

This year's highlights are

- A conference paper describing application of non-negative tensor factorization in traffic flow state prediction and clustering has been published and presented at ITS World Congress [30];
- The application of non-negative matrix factorization in clustering network-level traffic flow state in large-scale transportation network has been accepted for publication in a journal [11].

6.14. Sparse covariance inverse estimate for Gaussian Markov Random Field

Participants: Cyril Furtlehner, Yufei Han, Jean-Marc Lasgouttes, Victorin Martin.

We investigate in [53] different ways of generating approximate solutions to the inverse problem of pairwise Markov random field (MRF) model learning. We focus mainly on the inverse Ising problem, but discuss also the somewhat related inverse Gaussian problem. In both cases, the belief propagation algorithm can be used in closed form to perform inference tasks. We propose a novel and efficient iterative proportional scaling (IPS) based graph edit method to identify sparse graph linkage of GMRF model to fit underlined data distribution. We remark indeed that both the natural gradient and the best link to be added to a maximum spanning tree solution can be computed analytically. These observations open the way to many possible algorithms, able to find approximate sparse solutions compatible with belief propagation inference procedures and sufficiently flexible to incorporate various spectral constraints like e.g. walk summability. Experimental tests on various data sets with refined L_0 or L_1 regularization procedures indicate that this approach may be a competitive and useful alternative to existing ones.

The part of this work dedicated to Gaussian Markov Random Field has been submitted to the AISTATS 2013 conference.

6.15. Evaluation of dual mode transport system by event-driven simulation

Participants: Arnaud de La Fortelle, Jean-Marc Lasgouttes, Thomas Liennard.

The European project CATS — City Alternative Transport System — is developing and evaluating a new vehicle system using a single type of vehicle for two different usages: individual use or collective transport. Real experiments will necessarily take place with a limited number of vehicles and stations. Hence there is a need for evaluation using simulations. We have been developing a discrete events simulator for that purpose, based on a previous work done for collective taxis [58].

Our model relies on an adapted events/decision graph that extends previous graphs. The new feature of this model is the way we deal with two modes that can be extended to many other modes. This work therefore shows on a concrete example a method to efficiently merge multiple modes into one model.

This year has seen the overhaul of the simulator implementation, as well as the development of a result visualizer that can replay the simulations on a map and show various statistics.

6.16. Multi-speed exclusion processes

Participants: Cyril Furtlehner, Jean-Marc Lasgouttes.

The slow-to-start mechanism is known to play an important role in the particular shape of the fundamental diagram of traffic and to be associated to hysteresis effects of traffic flow. We study this question in the context of stochastic processes, namely exclusion and queueing processes, by including explicitly an asymmetry between deceleration and acceleration in their formulation. Spatial condensation phenomena and metastability are observed, depending on the level of the aforementioned asymmetry. The relationship between these 2 families of models is analyzed on the ring geometry, to yield a large deviation formulation of the fundamental diagram (FD)

This work has been published in the Journal of Statistical Physics [10].

6.17. Herding behavior in a social game

Participants: Guy Fayolle, Jean-Marc Lasgouttes.

The system *Ma Micro Planète* belongs to the so-called *Massively Multi-Player online Role Playing game* (MMORPG), its main goal being to incite users to have a sustainable mobility. Two objectives have been pursued.

- Construct an experimental platform to collect data in order to prompt actors of the mobility to share information (open data system).
- See how various mechanisms of a game having an additive effect could modify the transportation requests.

At the heart of the game are community-driven *points of interest* (POIs), or *sites*, which have a score that depends on the players activity. The aim of this work is to understand the dynamics of the underlying stochastic process. We analyze in detail the stationary regime of the system in the thermodynamic limit, when the number of players tends to infinity. In particular, for some classes of input sequences and selection policies, we provide necessary and sufficient conditions for the existence of a complete meanfield-like measure, showing off an interesting *condensation* phenomenon.

The work has been completed during this year [51] and has been submitted to a journal for publication.

6.18. Exact asymptotics of random walks in the quarter plane

Participant: Guy Fayolle.

In collaboration with K. Raschel (CNRS, Université F. Rabelais à Tours), we pursued the works initiated in 2011.

The enumeration of planar lattice walks, is a classical topic in combinatorics. For a given set $\mathcal{S}$ of allowed unit jumps (or steps), it is a matter of *counting the number of paths* starting from some point and ending at some arbitrary point in a given time, and possibly restricted to some regions of the plane.

Like in the probabilistic context, a common way of attacking these problems relies on the following analytic approach. Let $f(i,j,k)$ denote the number of paths in $\mathbb{Z}_+^2$ starting from $(0,0)$ and ending at (i,j) at time k. Then the corresponding CGF

$$F(x,y,z) = \sum_{i,j,k\geq 0} f(i,j,k)x^i y^j z^k$$

satisfies the functional equation

$$K(x,y)F(x,y,z) = c(x)F(x,0,z) + \widetilde{c}(y)F(0,y,z) + c_0(x,y),$$

where x, y, z are complex variables, although the time variable z plays somehow the role of a parameter. The question of the type of the associated counting generating functions, that is rational, algebraic, holonomic (solution of a linear differential equation with polynomial coefficients), was solved whenever the group is *finite* (see RA 2010). When the group is infinite, the problem is still largely.

It turns out that the nature of singularities play a deep important role in this classification. Making use of the general and powerful approach proposed in the book [2], the paper [9] has been presented at the 23rd International Conference *AofA 2012* on *Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, Montreal, June 17-22.

6.19. Statistical physics and hydrodynamic limits

Participant: Guy Fayolle.

These last years, having in mind a global project concerning the analysis of complex systems, we did focus on the interplay between discrete and continuous description: in some cases, this recurrent question can be addressed quite rigorously via probabilistic methods (see previous activity reports).

To describe the systems of interest, which are in touch with many application domains, we started from *paradigmatic* elements, namely discrete curves subject to stochastic deformations. Up to some convenient mappings, it appears that most models can be set in terms of interacting exclusion processes, the ultimate goal being to derive *hydrodynamic limits* after proper scalings.

The key ideas can be found in [56], where the basic ASEP system on the torus is the toy model. In this case, the usual sequence of empirical measures, converges in probability to a deterministic measure, which is the unique weak solution of a Cauchy problem.

The Gordian knot is indeed the analysis of a family of specific partial differential operators in infinite dimension. Indeed, the values of functions at given points play here the role of usual variables, their number becoming infinite. The method presents some new theoretical features, involving path integrals, promeasures (as introduced by Bourbaki), variational calculus, and the construction of *generalized measures*. In [56], we present a detailed analysis of the ASEP system on the torus $\mathbb{Z}/N\mathbb{Z}$. Then we claim that most of the arguments a priori for multi-type exclusion processes, and should lead to systems of coupled partial differential equations of Burgers' type. At the moment, this claim is being proved for the famous ABC model, reformulated in terms of the dynamics of a random walk on the triangular lattice.

LAGADIC Project-Team

6. New Results

6.1. Visual tracking

6.1.1. 3D model-based tracking

Participants: Antoine Petit, Eric Marchand.

Our 3D model-based tracking algorithm [2] was used in various contexts. We began a collaboration with Astrium EADS in 2010 in order to build a more versatile algorithm able to consider complex objects. The main principle is to align the projection of the 3D model of the object with observations made in the image for providing the relative pose between the camera and the object using a non-linear iterative optimization method. The approach proposed takes advantage of GPU acceleration and 3D rendering. From the rendered model, visible edges are extracted, from both depth and texture discontinuities. Potential applications would be the final phase of space rendezvous mission, in-orbit servicing, large debris removal using visual navigation, or airborne refuelling [41], [40], [32].

6.1.2. Omnidirectional vision system

Participant: Eric Marchand.

In this study performed in collaboration with Guillaume Caron and El Mustapha Mouaddib from Mis in Amiens, we have been interested by the redundancy brought by stereovision in omnidirectional vision sensors. This has been obtained by combining a single camera and multiple mirrors. Within this framework, we proposed to extend the 3D model-based tracking algorithm [2] for such system [15].

Thanks to a collaboration with Esiea in Laval, France, and the Inria and Irisa Hybrid team, we developed a system named Flyviz that has been patented. It is composed of a helmet mounted catadioptric camera coupled with an immersive display. The image acquired by the sensor is processed to give the user a full 360-degree panoramic view [27].

6.1.3. Pose estimation using mutual information

Participant: Eric Marchand.

Our work with Amaury Dame related to template tracking using mutual information [17] as registration criterion has been extended to 3D pose estimation using a 3D model. Since a homography was estimated, the tracking approach presented in [17] was usable for planar scenes. The new approach [45] can be considered for any scene or camera motion. Considering mutual information as similarity criterion, this approach is robust to noise, lighting variations and does not require a statistically robust estimation process. It has been used for visual odometry in large scale environment.

6.1.4. Pseudo-semantic segmentation

Participants: Rafik Sekkal, François Pasteau, Marie Babel.

To address the challenge of tracking initialization issues, we investigate joint segmentation and tracking approaches characterized by resolution and hierarchy scalability as well as a low computational complexity. Through an original scalable Region Adjacency Graph (RAG), regions can be adaptively processed at different scale representations according to the target application [42]. The results of this pseudo-semantic segmentation process are further used to initialize the object tracker (patch, visual objects, planes...) on several scales of resolutions.

6.1.5. Augmented reality using RGB-D camera

Participants: Hideaki Uchiyama, Eric Marchand.

We consider detection and pose estimation methods of texture-less planar objects using RGB-D cameras. It consists in transforming features extracted from the color image to a canonical view using depth data in order to obtain a representation invariant to rotation, scale, and perspective deformations. The approach does not require to generate warped versions of the templates, which is commonly needed by existing object detection techniques [35].

We also investigate the use of RGB-D sensors for object detection and pose estimation from natural features. The proposed method exploits depth information to improve keypoint matching of perspectively distorted images. This is achieved by generating a projective rectification of a patch around the keypoint, which is normalized with respect to perspective distortions and scale [34].

6.2. Visual servoing

6.2.1. Visual servoing using the sum of conditional variance

Participants: Bertrand Delabarre, Eric Marchand.

Within our study of direct visual servoing, we propose a new similarity function: the use of the sum of conditional variance [31] that replace SSD or mutual information [3]. It has been shown to be invariant to non-linear illumination variations and inexpensive to compute. Compared to other direct approaches of visual servoing, it is a good trade off between techniques using the pixels luminanc,e which are computationally inexpensive but non robust to illumination variations, and other approaches using the mutual information, which are more complicated to compute but offer more robustness towards the variations of the scene.

6.2.2. Photometric moment-based visual servoing

Participants: Manikandan Bakthavatchalam, Eric Marchand, François Chaumette.

The direct visual servoing approaches that have been developed in the group in the recent years, either using the luminance of each pixel, or the mutual information [3], or the sum of conditional variance described just above, allows reaching an excellent positioning accuracy. This good property is however counterbalanced by a small convergence domain due to the strong non linearities involved in the control scheme. To remedy to these problems, we started a study on using photometric moments as inputs of visual servoing. We expect to find again the nice decoupling and large convergence domain that we obtained for binary moments, without the need of any object segmentation.

6.2.3. Visual servoing using RGB-D sensors

Participants: Céline Teulière, Eric Marchand.

We propose a novel 3D servoing approach [43] that uses dense depth maps to perform robotic tasks. With respect to pose-based approaches, our method does not require the estimation of the 3D pose, nor the extraction and matching of 3D features. It only requires dense depth maps provided by 3D sensors. Our approach has been validated in servoing experiments using the depth information from a low cost RGB-D sensor. Thanks to the introduction of M-estimator in the control law, positioning tasks are properly achieved despite the noisy measurements, even when partial occlusions or scene modifications occur.

6.2.4. Visual servoing of cable-driven parallel robot

Participant: François Chaumette.

This study is realized in collaboration with Rémy Ramadour and Jean-Pierre Merlet from EPI Coprin at Inria Sophia Antipolis. Its goal is to adapt visual servoing techniques for cable-driven parallel robot in order to achieve acurate manipulation tasks. This study is in the scope of the Inria large-scale initiative action Pal (see Section 8.2.7).

6.2.5. Micro-Nanomanipulation

Participants: Eric Marchand, Le Cui.

In collaboration with Femto-ST in Besançon, we developed an accurate nanopositioning system based on direct visual servoing [20]. This technique relies only on the pure image signal to design the control law, by using the pixel intensity of each pixel as visual features. The proposed approach has been tested in terms of accuracy and robustness in several experimental conditions. The obtained results have demonstrated a good behavior of the control law and very good positioning accuracy: 89 nm, 14 nm, and 0.001 degrees in the x, y and θ_z axes of a positioning platform, respectively.

We begin a work, within the ANR P2N Nanorobust project (see Section 8.2.4), on the development of micro- and nano-manipulation within SEM (Scanning Electron Microscope). Our goal is to provide visual servoing techniques for positioning and manipulation tasks with a nanometer precision.

6.2.6. Autonomous landing by visual servoing

Participants: Laurent Coutard, François Chaumette.

This study was realized in collaboration with Dassault Aviation with the financial support of DGA. It was concerned with the autonomous landing of fixed wing aircrafts on carrier by visual servoing. A complete system has been developed [12]. The vision part consists in detecting the carrier in the image sequence and then tracking it using either dense template tracking or our 3D model-based tracker [2]. The visual servoing part consists in computing particular visual features able to correctly handle the aircraft degrees of freedom. Perturbations due to the wind and carrier motions have also been considered. The complete system has been validated in simulation using synthetic images provided by Xplane simulator and a dynamic model of the aircraft provided by Dassault Aviation.

6.3. Visual navigation of mobile robots

6.3.1. Visual navigation using mutual information

Participants: Eric Marchand, Bertrand Delabarre.

We have developed a visual navigation scheme based on the mutual information between the images acquired by an onboard camera and a visual memory to control the orientation of a vehicle during its navigation [18].

We also proposed to extend this approach to visual servoing with vision systems that consider the unified sphere model for central cameras using a normalized version of the mutual information. This permitted to apply the technique to large fields of view with a more reliable similarity function [30].

6.3.2. 3D Mapping and real time navigation

Participants: Maxime Meilland, Patrick Rives.

This study was realized in collaboration with Andrew Comport from I3S in Sophia Antipolis. Our approach relies on a monocular camera on board the vehicle and the use of a database of spherical images of the scene acquired during an offline step [14]. This geo-referenced database allows us to obtain a robust **drift free** localization. Basically, the database is composed of spherical images augmented by depth that are positioned in a GIS (Geographic information system). This spherical robot centered representation accurately represents all necessary information for vision-based navigation and mapping [37]. During the online navigation, the vehicle pose is computed by aligning the current image acquired by the camera with the closest reference sphere extracted from the database [26].

6.3.3. Indoors Slam

Participants: Cyril Joly, Patrick Rives, Pierre Martin, Eric Marchand.

We developed in Sophia Antipolis a new Slam method fusing laser scan data with the spherical images provided by an omnidirectional camera. Thanks to the trace of the laser scan projected onto the spherical view, we are able to compute a RGB-D model of the environment by using a dense visual Slam approach.

In Rennes and in collaboration with Orange Labs, we considered the development of a visual Slam algorithm. Since the targeted platforms in this this study are Android Smartphone, sequential Slam approaches have been studied.

6.3.4. Topological navigation

Participants: Alexandre Chapoulie, Patrick Rives.

This study is realized in collaboration with David Filliat from Ensta in Paris. Navigation algorithms are often sensitive to the robot orientation involving an impossibility to detect a place already visited from a different point of view. In order to alleviate this drawback, panoramic or omnidirectional cameras are often used. We have developed a loop closure detection algorithm based on an ego-centric spherical view that satisfies, in addition to other properties, a robot orientation independence [11].

A topological model captures the accessibility of the different places in the environment and allows a coarse localization. From a sequence of spherical views, we have developed a context-based segmentation algorithm. We hence define a topological place as having a structure which does not change, variation leading to a place change. The structure variations are detected with an efficient change-point detection algorithm [28].

6.3.5. Development of an autonomous shopping cart

Participants: Luca Marchetti, Patrick Rives.

This work is realized in collaboration with Pascal Morin from Isir in Paris. It consists in developing a shopping cart with autonomy capabilities (automatic user following, obstacle avoidance, etc), as part of the Inria Large-scale initiative action Pal, which aims at developing robotic tools for disabled persons or elderlies (see Section 8.2.7). Experiments have been successfully conducted both on the mobile robot Hannibal and on the wheeled walking aid ANG (Assistive Navigation Guide) developed by the EPI Coprin in Sophia Antipolis [36].

6.3.6. Automous navigation of wheelchairs

Participants: Rafik Sekkal, François Pasteau, Marie Babel.

This study is aimed at designing a robotic vision-based system dedicated to assisted navigation of electrical wheelchair in an unkown environment. In particular, going through doors, taking the elevator in a secure way without risking collision because of hazardous wheelchair motions remain a relevant issue. The idea is here to provide an embedded and flexible system able to ensure the immediate compatibility of the proposed system with existing electrical wheelchairs. From the platform described in Section 5.5 , we first addressed the door detection issue for automatically initializing the tracking process that is required for localisation and navigation purposes. We then defined a low complex solution of automatic door recognition that can be decomposed into three successive steps: line extraction (LSD-based algorithm), vanishing point estimation and door recognition itself by using geometrical cues. As soon as a door is detected and tracked through model-based trackers, the idea is to take into account the position of the wheelchair joystick in order to interpret the intention of the user. First experiments have shown the validity of the proposed approach. This study is conducted in conjunction with the scope of the Inria large-scale initiative action Pal (see Section 8.2.7).

6.3.7. Obstacle avoidance

Participants: Fabien Spindler, François Chaumette.

This study was realized in collaboration with Andrea Cherubini who is now Assistant Prof. at Université de Montpellier. It is concerned with our long term researches about visual navigation from a visual memory without any accurate 3D localization [9]. In order to deal with obstacle avoidance while preserving the visibility in the visual memory, we have proposed a control scheme based on tentacles for fusing the data provided by a pan-tilt camera and a laser range sensor [16].

6.4. Medical robotics

6.4.1. Visual servoing based on dense ultrasound information

Participants: Caroline Nadeau, Alexandre Krupa.

In the context of the ANR USComp project (see Section 8.2.3), we pursued our works on the development of ultrasound image-based visual servoing methods that directly use pixel intensities of the ultrasound image as control inputs. In opposite with methods based on geometrical visual features, this new approach does not require any image segmentation step that is difficult to robustly perform on ultrasound images. By coupling our method with a predictive control law based on the periodicity of physiological motion, we propose a solution to stabilize the ultrasound image by actively compensating the physiological motions of the patient. The principle consists in automatically synchronizing the 6 DOF motion of a 2D or 3D probe with the rigid motion of a soft tissue target. First ex-vivo results obtained on animal tissues demonstrated the validity of the concept [39].

In collaboration with Prof. Pierre Dupont from Harvard University at Boston, we also addressed the motion tracking of a target that can consist of either the tip of a robot inserted on a beating heart or cardiac tissues. Unlike the previous work, where the motion compensation task was realized physically by moving the probe attached to a robotic arm, we propose here to track the motion of the target using a 3D region of interest (ROI) which is automatically moved within the whole volume observed by a 3D probe thanks to our intensity-based ultrasound visual servoing method. In vivo animal experiments were conducted in Children's Hospital at Boston and validated this tracking approach [38].

6.4.2. Autonomous control modes for ultrasound probe guidance

Participants: Tao Li, Alexandre Krupa.

In the context of the ANR Prosit project (see Section 8.2.2), we proposed several autonomous control modes in order to assist a doctor during a robotized and teleoperated ultrasound examination (tele-echography). This year we developed an assistance functionality that automatically maintains the visibility of an anatomic element of interest while the doctor teleoperates the 2D ultrasound probe held by the medical robot. The method is based on a multi-task controller that gradually activates an ultrasound visual servoing in case some geometrical features leave a pre-defined safe area of the image in order to bring them back inside the view [33]. With this approach the DOFs of the robotized probe are not exclusively constrained by the visibility task but also available for the tele-operation. This new assistance functionality was implemented on the ANR Prosit robotic platform and first in vivo results obtained on a human volunteer validated the concept.

6.4.3. Real-time soft-tissue deformation tracking in 3D ultrasound

Participant: Alexandre Krupa.

We proposed a dense ultrasound tracking algorithm that estimates in real time both rigid and non-rigid motions of a region of interest observed in a sequence of 3D ultrasound images. The deformation is modeled by 3D thin-plate splines (TPS) whose parameters are estimated online from intensity difference measured in successive volumes. To increase the robustness of this approach to image noise, we proposed two solutions to mechanically constrain the deformable model. The first is based on the addition of a regularization term in the TPS model and the second consists in coupling the TPS with a mass-spring system. These methods were validated on simulated sequences of deformed 3D ultrasound images.

6.4.4. Needle detection and tracking in 3D ultrasound

Participant: Alexandre Krupa.

We designed an algorithm able to detect a needle inserted manually in a 3D ultrasound volume from an arbitrary point, and able to robustly track this needle in real-time. We also experimentally demonstrated the possibility to guide the ultrasound probe to keep the needle visible and aligned, using visual servoing. Such a system could assist an operator during manual insertions, which are currently performed under free-hand ultrasound monitoring. In addition, we plan in future works to combine this method to a needle steering robotic system for guiding accurately the needle toward a target while optimizing its visibility.

AYIN Team

6. New Results

6.1. Markov random fields and Marked point processes

6.1.1. Stochastic modeling for very high resolution SAR image processing

Participants: Aurélie Voisin, Vladimir Krylov, Josiane Zerubia [contact].

This work was done in collaboration with DITEN, University of Genoa, with Dr. Gabriele Moser and Pr. Sebastiano B. Serpico with partial financial support of the French Defense Agency, DGA (http://www.defense.gouv.fr/dga/). The data were provided by the Italian Space Agency, ISA (http://www.asi.it/en) in the framework of the project "Development and validation of multitemporal image analysis methodologies for multirisk monitoring of critical structures and infrastructures (2010-2012)".

high resolution, synthetic aperture radar data, multi-sensor data, urban areas, supervised classification, hierarchical Markov random fields, statistical modeling, wavelets, textural features

The classification of remote sensing images including urban areas is relevant in the context of the management of natural disasters (earthquakes, floodings...), and allows to determine land-use and establish land cover maps, or to localise damaged areas. Given the huge amount and variety of data available nowadays, the main difficulty is to find a classifier that takes into account multi-band, multi-resolution, and possibly multi-sensor data. A minor part of our work was also dedicated to the change detection [14], still in the frame of the management of natural disasters.

We developed a supervised Bayesian classifier that combines a joint class-conditional statistical modeling and a hierarchical Markov random field. The first classification step deals with the statistical modeling for each target class (e.g. vegetation, urban, etc.) by using a finite mixture model, estimated by resorting to a modified stochastic expectation maximization (SEM) algorithm. Such a model is well-adapted to deal with heterogeneous classes, and each mixture component reflects the contribution of the different materials contained in a given class. When considering optical images, the statistics are modeled by using finite mixtures of Gaussian distributions. In the case of SAR amplitude imagery, we favor a finite mixture of generalized Gamma distributions. Then, at each considered resolution, the different input bands are statistically combined by using multivariate copulas. The second classification step relies on the integration of this statistical modeling in a hierarchical Markov random field integrated in a quad-tree structure. Such contextual classifier helps improving the robustness of the method with respect to noise, or to SAR speckle. A variety of algorithms were proposed to estimate the labels on hierarchical graphs. The consideration of a specific graph, here a quad-tree, allows to benefit from its good properties (e.g. causality) and to apply non iterative algorithms. Among the different algorithms employed in the literature, we chose to take into account an exact estimator of the marginal posterior mode (MPM). The cost function associated to this estimator offers the possibility to penalize the errors according to their number and the scale at which they occur: an error at the coarsest scale is stronger penalized than an error at the finest scale. Moreover, we introduce a prior estimation update that experimentally leads to improved results.

The first experiments were run on single-polarized, mono-resolution synthetic aperture radar (SAR) amplitude images. The challenge of the problem considered here is that the given input is at a single resolution and should be integrated in a multi-scale tree. Thus, we extract an extra information in the form of a multi-scale wavelet decomposition from the initial image. Then, at each level, a textural feature map (e.g. Haralick's variance) is obtained from each image in the decomposition stack, and integrated as an additional information that aims at discriminating the urban areas. Finally, at each level, the wavelet image is combined with the textural image by using copulas, as described previously. These results were presented in [10], [23].

Such a classifier is sufficiently flexible to take into account different types of data [21], [22]. Thus, we also tested coregistered data of a given area acquired at different resolutions (e.g., multiresolution SAR images), directly integrated at the different levels of the hierarchical tree. The classification of multisensor (optical/SAR) data is illustrated in Fig. 1 . In this specific example, we consider a GeoEye acquisition (resolution: 65 centimeters) and a coregistered COSMO-SkyMed SAR acquisition (resolution: 2.5 meters) of the Port-au-Prince quay (Haiti). Spatially disjoint training and test areas were manually annotated. The classification is done following 5 classes: urban areas, natural landscape, sand, containers and wet areas.

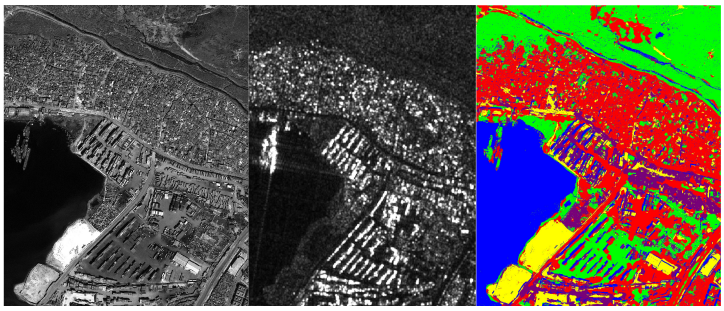


Figure 1. Left: Initial optical image of Port-au-Prince (Haiti) (©GeoEye, 2010). Middle: Initial SAR image of Port-au-Prince (Haiti) (©ISA, 2010). Right: Classification map obtained with the proposed hierarchical method for the 5 classes (Blue: wet areas; Green: vegetation; Red: urban areas; Yellow: sand; Purple: containers).

We have also run experiments on other types of acquisitions, such as histological images [22], to prove the robustness of the proposed algorithm with respect to different image sources.

6.1.2. Satellite image classification using Bootstrap EM

Participants: Siddharth Buddhiraju, Josiane Zerubia [contact].

This program has been partially funded by the Direction of International Relations of Inria (DRI).

Bootstrapping, Expectation-Maximization Algorithm, Iterated Conditional Expectation, Markov Random Fields, Simulated Annealing.

We implemented both Bootstrap EM and Iterated Conditional Expectation algorithms for parameter estimation of first order Markov Random Field models followed by Simulated Annealing, for optimal segmentation of gray-scale images. The objective was to perform a quantitative comparison of the two methods. Apart from successful implementation of these algorithms, an extension of these to multispectral images was performed, and the obtained results were found to be of superior quality compared against the original gray-scale ones (see Fig. 2).

First, Bootstrapped EM or Iterated Conditional Expectation were performed based on the user's requirement. The estimated parameters were then used to obtain the optimal segmentation of the image via simulated annealing. The algorithm was extended using multivariate Gaussian models to perform the same for multispectral images.

6.1.3. Boat detection and counting in Mediterranean Harbors using Marked Point Processes

Participants: Paula Craciun, Josiane Zerubia [contact].

This work has been conducted in collaboration with ACRI-ST (http://www.acri-st.fr/) and the French Space Agency (http://www.cnes.fr/), which provided the high resolution satellite images.

stochastic modeling, marked point process, object extraction, SEM, simulated annealing

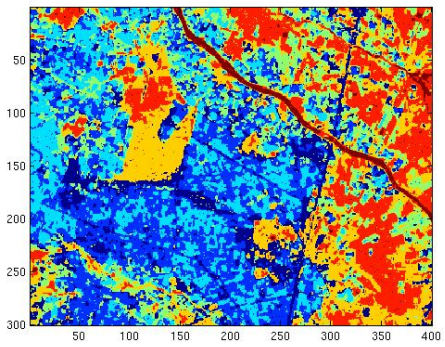


Figure 2. A sample result for a 4-band IRS-1A satellite image of a village in India. 7 clusters were considered to segment the image using Bootstrapped EM. The image was provided by Prof. Krishna Mohan Buddhiraju of CSRE, Indian Institute of Technology Bombay.

Marked point processes have been successfully applied in image processing analysis, when dealing with high resolution images in the purpose of feature extraction. The objective of this research was to improve the already existing marked point process model of ellipses to better fit the detection of boats in a harbor. The model involved two types of energy terms: a data term, used to determine the fidelity to the existing data (i.e. image) and a prior energy term, used to describe relationships between the objects. We proposed new energy components to model boat detection. The proposed model relied on a high number of parameters. While most of these parameters had an intuitive meaning and could be, thus, set manually, others were difficult to determine. We therefore used a parameter estimation method, based on the Stochastic Expectation - Maximization (SEM) algorithm, which proved to provide good results when combined with marked point processes. Furthermore, we proposed additional automatic procedures based on mathematical morphology to determine critical parameters of this model. Experimental results of boat detection are shown on Fig. 3 .

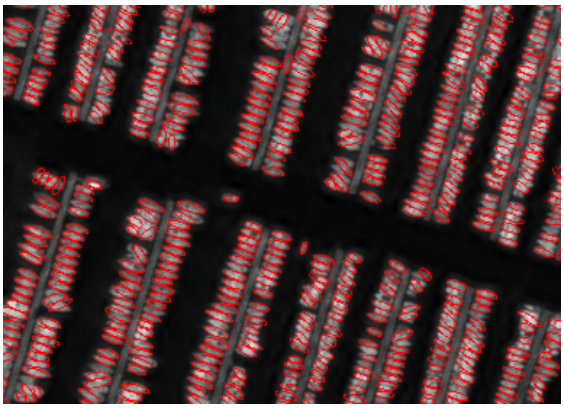


Figure 3. Result of boat detection using Marked Point Processes.

6.1.4. Contribution of object recognition on forest canopy images to the building of an allometric theory for trees and natural, heterogeneous forests

Participants: Jia Zhou, Josiane Zerubia [contact].

This work was done in collaboration with Xavier Descombes (Morpheme team, Inria-SAM), Dr. Pierre Coureron and Dr. Christophe Proisy at IRD, UMR AMAP (http://amap.cirad.fr/), Montpellier.

object detection, marked point processes, tree crowns, forest structure

This work aimed at providing information on the forest structure through the analysis of canopy properties as described by the spatial distribution and the crown size of dominant trees. Our approach was based on the Marked Point Processes (MPP) theory, which allows modeling tree crowns observed in remote sensing images by discs belonging to a two dimensional space. The potential of MPP to detect the trees crowns automatically was evaluated by using very high spatial resolution optical satellite images of both *Eucalyptus* plantations in Brazil and mangrove forest in French Guyana. LIDAR and simulated reflectance images were also analyzed for the mangrove application. Different adaptations (parameter settings, energy models) of the MPP method were tested and compared through the development of quantitative indices that allowed comparison between detection results and tree references derived from the field, photo-interpretation or the forest mockups.

In the case of mangroves, the estimated crown sizes from detections were consistent with the outputs from the available allometric models (Fig. 4 (Left and Middle)). Other results indicated that tree detection by MPP allowed mapping the local density of trees of young *Eucalyptus* plantations (Fig. 4 (Right), [11]) even if crown size was close to the image spatial resolution (0.5 m). However, the quality of detection by MPP decreased with canopy closeness. To improve the results, further work may involve MPP detection using objects with finer shapes and forest data measurements collected at the tree plant scale.

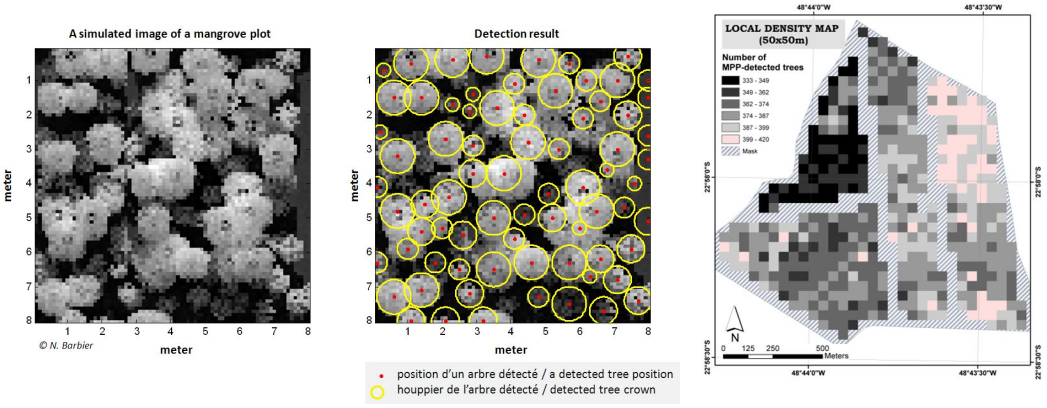


Figure 4. (Left and Middle) Result of detection on an image of mangrove. (Right) Example of a local tree density map computed over the entire plantation at 50 m resolution of the marked point process modeling. Masked areas are centered along the road network.

6.1.5. Detection of the hyperpigmentation of the skin on color images using Marked Point Process and Mathematical Morphology

Participants: Adrien Lacage, Josiane Zerubia [contact].

The source images were provided by the AYIN team itself for the study of folliculitis, and by an industrial leader in skin care for acne.

skin care, hyperpigmentation, acne, follicultis, marked point process, mathematical morphology

Automatic detection of the skin hyperpigmentation helps in estimating the severity of some skin diseases like *acne vulgaris* and *follicultis*. We compared two methods for studying acne and follicultis lesions and hyperpigmentation of the skin. We adapted a model based on Marked Point Processes and initially developed for flamingo's population counting to dermatological images of acne and follicultis. Then, we developed an algorithm which uses mathematical morphology together with volume and shadows compensation. Finally, we compared results in term of detection accuracy.

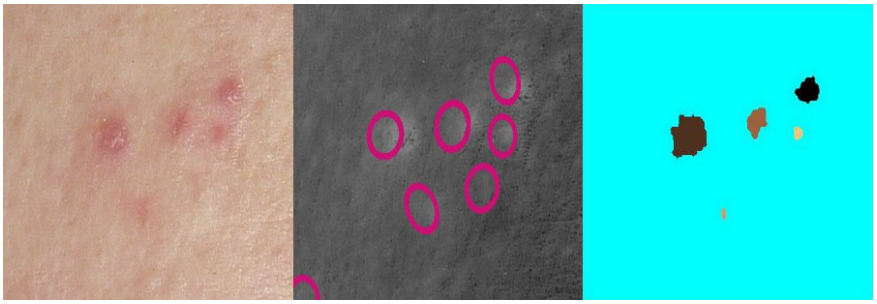


Figure 5. Source image of follicultis (Left) and results obtained with Marked Points Processes (Middle), then with Mathematical Morphology (Right)

6.1.6. Efficient Monte Carlo sampler for detecting parametric objects in large scenes

Participants: Yannick Verdie, Florent Lafarge [contact].

This work is supervised by Florent Lafarge (Geometrica team, Inria-SAM) in collaboration with the AYIN team.

Point processes have demonstrated both efficiency and relevance when addressing object recognition problems in vision. However, simulating these mathematical models is a difficult task, especially on large scenes. Existing samplers suffer from average performances in terms of computation time and stability. We propose a new sampling procedure based on a Monte Carlo formalism. Our algorithm exploits Markovian properties of point processes to perform the sampling in parallel. This procedure is embedded into a data-driven mechanism such that the points are non-uniformly distributed in the scene. The performances of the sampler are analyzed through a set of experiments on various object recognition problems from large scenes, and through comparisons to the existing algorithms.

6.2. Statistical methods

6.2.1. Change detection on synthetic aperture radar images based on hypothesis testing

Participants: Vladimir Krylov, Josiane Zerubia [contact].

This work was conducted in collaboration with DITEN, University of Genoa with Dr. Gabriele Moser and Prof. Sebastiano Serpico (http://spt.dibe.unige.it/) with the support of the Italian Space Agency, ASI (http://www.asi.it/en).

Change detection, synthetic aperture radar, hypothesis test, likelihood ratio test, high resolution

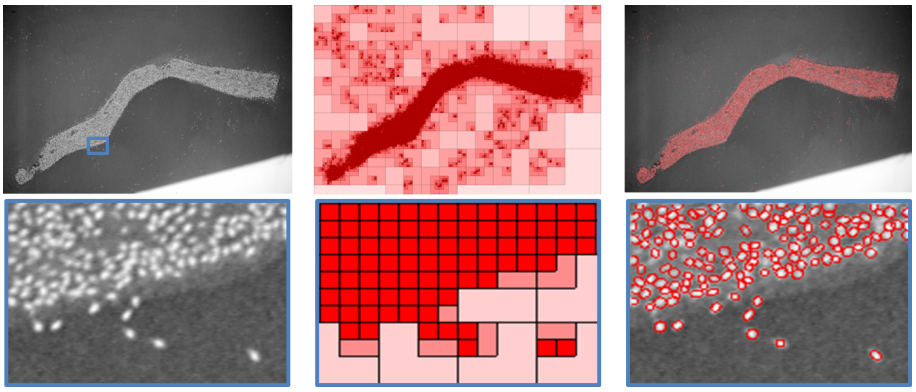


Figure 6. Bird counting by a point process of ellipses. (right) More than ten thousand birds are extracted by our algorithm in a few minutes from (left) a large scale aerial image. (middle) A quad-tree structure is used to create a non-uniform point distribution. Note, on the cropped parts, how the birds are accurately captured by ellipses in spite of the low quality of the image and the partial overlapping of birds.

Modern synthetic aperture radar (SAR) sensors represent an essential source of all-weather and 24-hour imagery with a fixed re-visit cycle at competitive high resolution. Two-date change detection from SAR images is a process that employs two SAR images acquired over the same geographical area with possibly the same (or close) acquisition characteristics at two different times to map the areas where changes occur between the two acquisition dates. The central disadvantage of the SAR imagery is given by an inherent multiplicative speckle noise, which restricts the direct application of optical-based change detection methods to SAR imagery.

We have developed a non-parametric statistical change detection approach. We avoided the ambiguity of choosing a restrictive clutter model by assuming no specific probability distribution function model [25] for the statistics of SAR. We developed a modified hypothesis test which is based on the classical Wilcoxon two-sample test that verifies whether one of two samples of independent observations tends to have larger values than the other. The choice of the Wilcoxon statistic as compared to the other available goodness-of-fit test statistics, such as, e.g., that of Cramér-von Mises' test, is a compromise solution to have simultaneously an analytically tractable asymptotic distribution (which is needed to formulate the likelihood ratio test) and a non-parametric testing procedure. Furthermore, the experimental validation demonstrated the adequacy of this statistic to the considered problem. To be able to take a decision at each pixel of the coregistered image pair we considered samples originating from the local windows centered in each pixel. Finally, we constructed a likelihood ratio test on the image with Wilcoxon statistic values. This formulation allowed to overcome the limitation of a classical independency assumption for the Wilcoxon test which is violated (at least, locally) with the local window samples. The resulting technique is related to the statistical false discovery rate approach developed for "large-scale simultaneous hypothesis testing" problems, however the derivation and interpretation are different.

Encouraging detection results were obtained on XSAR and very high resolution COSMO-SkyMed images [14].

6.2.2. Statistical analysis of skin pigmentation under treatment

Participants: Sylvain Prigent, Xavier Descombes, Josiane Zerubia [contact].

This work was partially funded by a contract with Galderma R&D (http://www.galderma.com/RampD.aspx). It was a collaboration between AYIN (Josiane Zerubia) and Morpheme (Xavier Descombes) teams.

multispectral imaging, skin, hyperpigmentation, hypothesis tests, statistical inferences

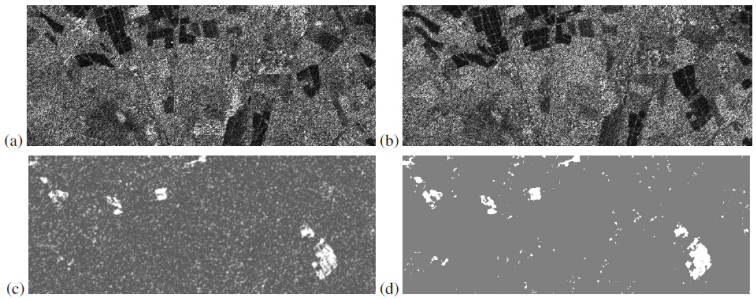


Figure 7. Coregistered XSAR images (©Univ. of Pavia) acquired on (a) April 16, 1994 and (b) April 18, 1994 and the detection results: (c) 5×5 window-based image ratio, (d) proposed method with 5×5 window.

One of the steps to evaluate the efficacy of a therapeutic solution is to test it on a clinical trial involving several populations of patients. Each population receives a studied treatment and a reference treatment for the disease. For facial hyper-pigmentation, a group of N_e patients receives the treatment on one cheek and a comparator on the other. The comparator can be a reference treatment or a placebo. To this end patients are selected to have the same hyper-pigmentation severity on the two cheeks. Then multi-spectral images are taken at different time t along the treatment period.
We proposed a methodology to estimate the efficacy a treatment by calculating three differential criteria: the darkness, the area and the homogeneity. The darkness measures the average intensity of the disease on a gray scaled image I obtained by a linear combination of the spectral bands of the original multi-spectral image. A differential darkness is then obtained by measuring the deviation between the initial measurement at time t_0, and the measurement at time t_k. The differential area criterion is calculated by analyzing the histogram of $I_{diff} = I_{t_0} - I_{t_k}$ a difference gray scale image between two measurements in a time series. The differential homogeneity criterion is obtained with a multi-scale analysis of I_{diff} adapted from the Statistical Parametric Mapping (SPM) methodology. Indeed, statistical inferences allow to assign a probability of change to each region of I_{diff} above a set of thresholds. These probabilities are calculated with respect to the maximum intensity and the spatial extend of each region. An integration of the obtained statistical map denoted SM, allows to get a homogeneity criterion.
The Fig. 8 illustrates the differential score calculated on a patient whose pathology decreases during the clinical trial. The proposed differential score has been tested in a full clinical study and provided results that agreed with the clinical analysis. This work have been patented, submitted to ISBI'13 conference and to the IEEE TMI journal, and published in Inria research reports [26], [27].

6.3. Hierarchical models

6.3.1. Hierarchical and graph cut-based models for multiyear sea ice floe segmentation

Participant: Yuliya Tarabalka [contact].

This work has been done in collaboration with Dr. Guillaume Charpiat (STARS team, Inria-SAM), Dr. Ludovic Brucker (NASA GSFC, USA) and Dr. James Tilton (NASA GSFC, USA).

hierarchical model, graph cut, segmentation, multiyear sea ice floes, shape analysis

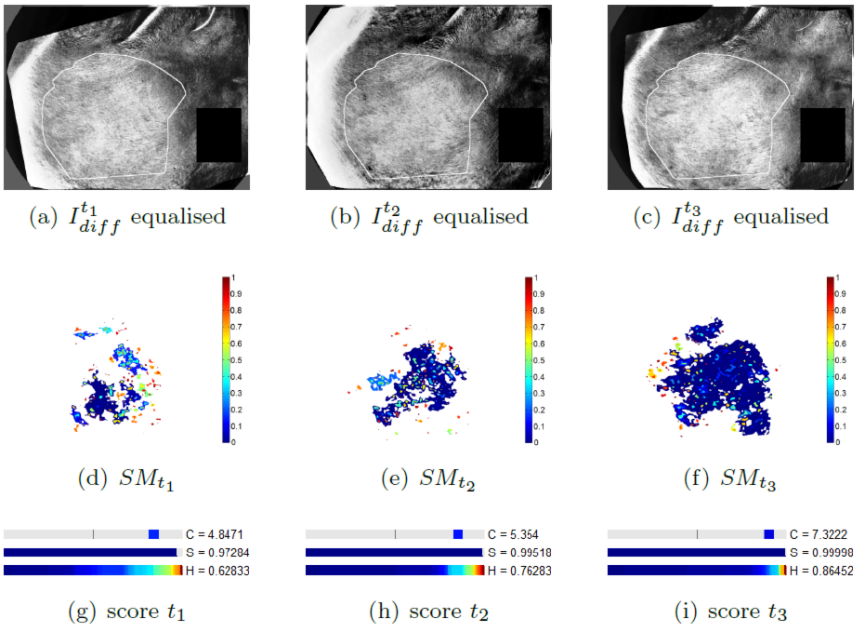


Figure 8. I_{diff}, SM and differential score for the three measurements t_1, t_2, t_3 calculated for a patient whose disease decrease.

The melting of sea ice is correlated to increases in sea surface temperature and associated climatic changes. Therefore, it is important to investigate how rapidly sea ice floes melt. We proposed two methods for segmentation of a time series of a melting sea ice floe. The first method employs hierarchical model for ice floe segmentation. Image features are extracted using morphological operators, and the floe of interest is marked based on AMSR-E satellite measurements. Then, hierarchical step-wise optimization segmentation is performed, by iteratively merging adjacent regions with the smallest dissimilarity criterion. We proposed to use area and shape parameters of the floe at two previous time moments as priors for computing a segmentation map at the next time moment.

Fig. 9 (a) depicts a graph of the multiyear ice floe area as a function of time, computed by applying the proposed hierarchical model to the summer series of Moderate-Resolution Imaging Spectroradiometer (MODIS) images. While a multiyear ice floe can only melt in the summer period, peaks on the graph correspond to segmentation errors, which are a consequence of either a cloud cover or weakness of contrast between the multiyear ice and the surrounding young ice floes. These segmentation imprecisions can be avoided by simultaneously optimizing all segmentation maps in a time series. For this purpose, we developed a new method based on graph cuts for joint segmentation of monotonously shrinking (or growing) shapes. We impose shape shrinkage (or growth, respectively) constraint in graph cuts, and minimization of energy computed on the resulting graph of the image sequence yields globally optimal segmentation. Fig. 9 (c-d) show examples of floe segmentations using the new approach. Fig. 9 (b) presents a graph of the floe area as a function of time computed by performing the proposed graph cut-based method. The results are compared to those obtained by applying graph cut segmentation to each single image in the considered time series. It can be seen that the new approach yields results with continuous shrinkage of the shape size.

6.3.2. Hierarchical model for spectral-spatial classification of hyperspectral images

Participant: Yuliya Tarabalka [contact].

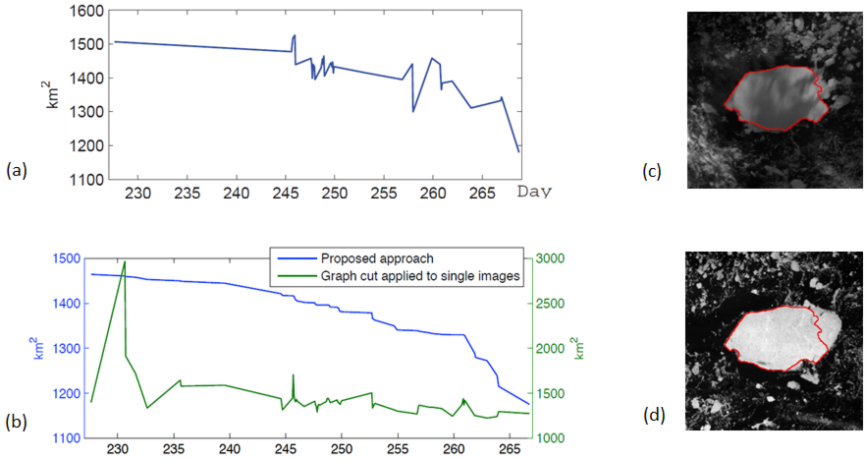


Figure 9. (a) Area of a multiyear ice floe as a function of time, computed by applying the proposed hierarchical model. (b) Area of the floe as a function of time (days), computed by using the proposed graph cut-based model (blue) and by applying graph cut segmentation to each single image in a time series. (c-d) Examples of floe segmentations using the proposed graph cut-based model.

This work has been done in collaboration with Dr. James Tilton (NASA GSFC, USA).

hyperspectral images, classification, segmentation, geometrical features, rectangularity.

The recent advances in hyperspectral remote sensor technology makes it possible to acquire data with a very high spectral (hundreds of spectral channels) and spatial (order of a meter) resolution. The rich spectral information of the hyperspectral data leads to the potential of a more accurate classification, but also presents challenges of high-dimensional data analysis.

We developed a new method for spectral-spatial classification of hyperspectral images. The method is based on the integration of probabilistic classification and shape analysis within the hierarchical step-wise optimization algorithm. First, probabilistic support vector machines classification is applied. Then, at each iteration two neighboring regions with the smallest dissimilarity criterion are merged, and classification probabilities are recomputed. We proposed to estimate a dissimilarity criterion between regions as a function of statistical, classification and geometrical (area and rectangularity) features. Fig. 10 shows the obtained classification results on a 102-band ROSIS image of the Center of Pavia, Italy, which are compared with Support Vector Machines (SVM) classification results. These results did show that the proposed method succeeded in taking advantage of both spatial and spectral information for accurate hyperspectral image classification.

6.3.3. Classification of combined hyperspectral and panchromatic data using spectral-spatial approaches

Participants: Yuliya Tarabalka [contact], Josiane Zerubia.

This work has been conducted in collaboration with the French Space Agency CNES (http://www.cnes.fr/), with Dr. Marie-José Lefèvre, Dr. Hélène DeBoissezon and Mr. Manuel Grizonnet.

hyperspectral data, HYPXIM, data fusion, panchromatic image, segmentation

Hyperspectral imaging records a detailed spectrum for each pixel, opening new perspectives in classification. Currently, several hyperspectral satellite missions such as EnMAP (210 bands, GSR 30m) are under development. The future hyperspectral satellite missions PRISMA and HYPXIM also include a panchromatic channel

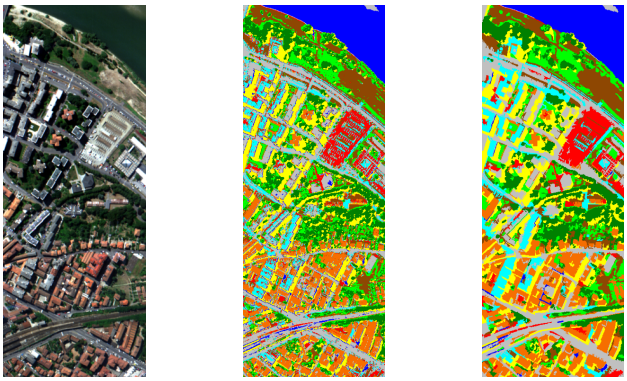


Figure 10. Center of Pavia image. (Left) Three-band color composite. (Middle) SVM classification map, overall classification accuracy = 94.9%. (Right) Classification map obtained using the proposed hierarchical approach, overall classification accuracy = 97.1%.

with better spatial resolution. We explored if a panchromatic channel at a higher spatial resolution (factor 4) contributes for more accurate classification of hyperspectral images in space conditions.

We adapted and compared several classification methods for combined hyperspectral and panchromatic images, and conducted experiments on the simulated HYPXIM data provided by CNES. We fused both data sources using principal component and Gram-Schmidt fusion methods, as well as the vector stacking approach. We then applied Support Vector Machines (SVM) classification on the resulting feature sets. Furthermore, we considered spatial information for more accurate classification by: (1) including Haralick's texture features in the feature set; (2) segmenting an image into homogeneous regions using a Hierarchical Step-Wise Optimization (HSWO) technique, and assigning each segmented region to the dominant class within this region.

Classification results are illustrated in Fig. 11 . We concluded that classification accuracies of the HYPXIM simulated data have been improved when including a panchromatic channel at a higher spatial resolution into a classification system. These results are close to hyperspectral aerial data classifications. For the image containing one-pixel regions and mixed pixels, standard spectral-spatial classification methods are not well adapted and thus do not improve accuracies when compared to pixelwise classification. In the future, we plan to develop methods which would use both spatial information and a spectral unmixing concept for efficient fusion of hyperspectral and panchromatic data.

6.4. Other detection approaches

6.4.1. Multiple-instance object detection using a higher-order active shape prior

Participants: Ikhlef Bechar, Josiane Zerubia [contact].

This work is done in collaboration with Dr. Ian Jermyn of Durham University (United Kingdom, https://www.dur.ac.uk/mathematical.sciences/) and was funded by a contract with the EADS foundation (http://www.fondation.eads.com/).

object detection, shape prior, transformation invariance, higher-order active contours, energy minimization, non-convex energy, exact convex relaxation.

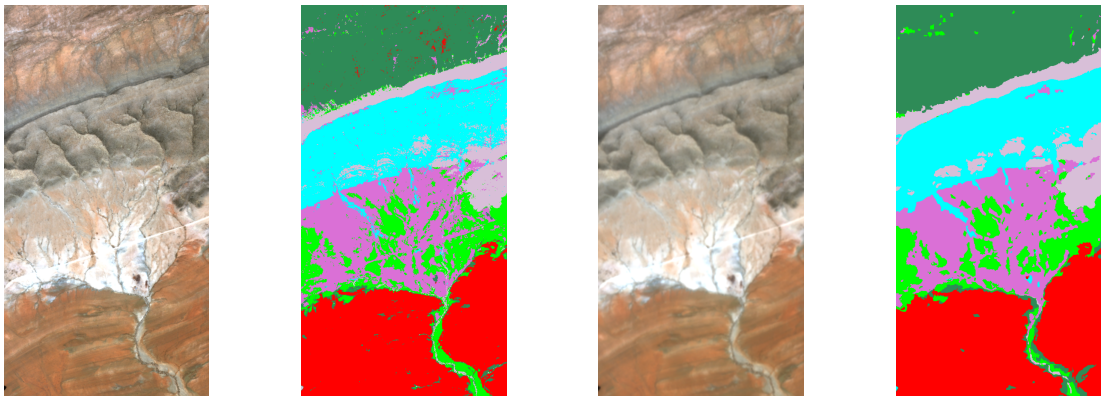


Figure 11. From left to right (a-d): (a) The false-color HYMAP aerial image (126 bands, GSR 4.8m). (b) SVM classification map for the HYMAP image, overall classification accuracy = 83%. (c) Simulated HYPXIM image (126 bands, GSR 14.4m). (d) Classification map of the fused by vector stacking hyperspectral image (c) and panchromatic image at GSR 4.8m, using HSWO-based spatial regularization, overall classification accuracy = 80.7%.

The problem under consideration is the multiple-instance object detection from imagery using prior shape knowledge. As mathematical and algorithmic framework, we have used the higher-order active contour (HOAC) model framework in order to incoporate prior shape knowledge about a class of objects of interest. On top of its robustness and its computational attractiveness (due to its parameter-estimation free method), the HOAC object-detection framework allows to incorporate shape knowledge about multiple occurrences of an object of interest in an image and to carry out object detection in a single algorithmic framework via the minimization of energy of the form:

$$\min_{\text{over all shapes}\gamma} E(\gamma) = E_{image}(\gamma) + E_{prior}(\gamma) \tag{2}$$

where γ stands for the contour an image object, $E_{image}(\gamma)$ stands for its image-based energy and $E_{prior}(\gamma)$ stands for a prior energy which is only a function of an objet's shape (and not of image data). The goal of this project is thus to model $E_{prior}(\gamma)$ using the HOAC methodology.

In this work, we have developed a fourth-order active contour (FOAC) framework for incorporating prior shape knowledge about target shapes. Typically, we express a FOAC energy model as

$$E_{foac}(\gamma) = \lambda_C L(\gamma) + \alpha_C A(\gamma) + \beta_C \int\int\int\int \langle \dot{\gamma}_p, \dot{\gamma}_q \rangle \, \mathbf{K}_C \left(|\gamma_p - \gamma_q|, |\gamma_s - \gamma_t| \right) \langle \dot{\gamma}_s, \dot{\gamma}_t \rangle \; dp\, dq\, ds\, dt \tag{3}$$

where $L(\gamma)$ and $A(\gamma)$ stand respectively for the length and the area of a contour γ and $\int\int\int\int \langle \dot{\gamma}_p, \dot{\gamma}_q \rangle \, \mathbf{K}_C \left(|\gamma_p - \gamma_q|, |\gamma_s - \gamma_t| \right) \langle \dot{\gamma}_s, \dot{\gamma}_t \rangle \; dp\, dq\, ds\, dt$ models fourth-order interactions between quadruples of contour points, and λ_C, α_C and β_C stand for some tradeoff parameters that control the contribution of each term of the FOAC energy. Note that the parameters of the method include both the real coefficients λ_C, α_C and β_C and the bivariate kernel $\mathbf{K}_C(u, v)$; $\forall\, u, v \in \mathbf{R}^+$. These parameters need to be tuned optimally for a given target shape γ^*. Thus we have developed a direct method for the optimal estimation of the FOAC parameters.

We have then shown that shapes with arbitrary geometric complexity can be modeled using such the FOAC framework 2 , and we have developed a direct method for the estimation of the parameters for a given class of shapes. In order to be able to detect multiple occurrences of a target object in an image, one needs to re-express such an originally contour-based energy 2 by replacing appropriately in formula 2 the one-dimensional contour quantity γ with an equivalent two-dimensional quantity (ie. with respect to the image domain) such as the characteristic function of γ and to minimize with respect to it the resulting energy functional. This allows topological changes of an evolving contour and hence the detection of possible multiple instances of a target object in an image. We have shown that such a new formalism is a third-order Markov Random Field (MRF) which practical optimization was a challenging question. Therefore, we have also developed a method for the exact minimization of the energy of the resulting MRF model (using a equivalent convex-relaxation approach, see Fig. 12).

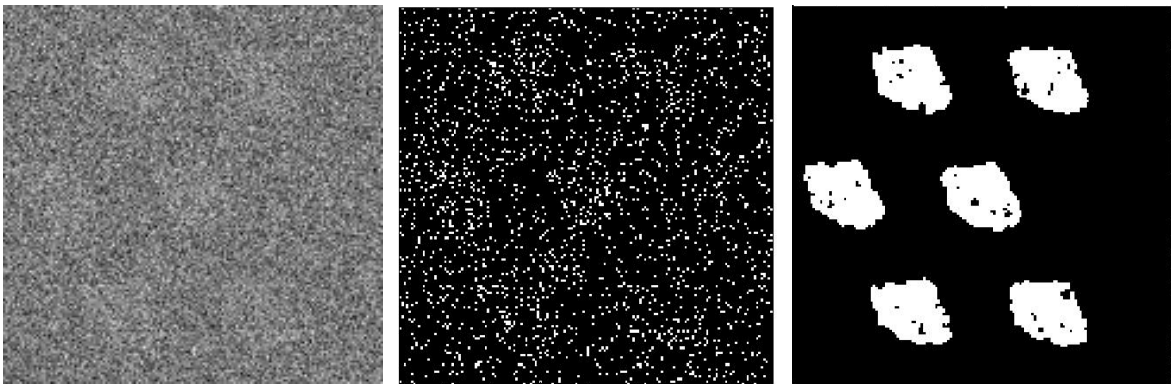


Figure 12. (Left) A very noisy input image; Multiple-instance object detection using: (Middle) a traditional segmentation method (Mumford-Shah model with a TV-based regularization) alone; (Right) with a FOAC shape prior.

6.4.2. Image analysis for automatic facial acne detection and evaluation

Participants: Zhao Liu, Josiane Zerubia [contact].

This work is part of LIRA Skin Care Project, which includes four key partners: Philips R&D (Netherlands, http://www.philips.nl), CWI (Netherlands, http://www.cwi.nl/), Fraunhofer Institutes (Germany, http://www.fraunhofer.de/en.html) and Inria (France).

image processing, feature extraction, pigmentation distributions, acne, cosmetology

Acne vulgaris is a highly prevalent skin disease, which has a significant life quality impact on sufferers. Although acne severity is readily observed by the human eye, it is an extremely challenging task to relate this visual inspection to measurable quantities of various skin tones and types. So far there is no golden standard for acne diagnosis in clinics, and it entirely depends on dermatologists' experience for evaluation of acne severity. But significant inter-rater variability among individual assessment may lead to less trustworthy diagnosis when several clinicians get involved in the study. In addition, less reproducibility of human evaluation makes comparison of acne changes over time difficult. Therefore, the long-term objective of this study is to construct an automatic acne grading system through applying spectroscopy imaging techniques and image processing methods, to objectively evaluate severity of skin disorder. Such a computer-based tool would also significantly benefit the development of better skin care products, if it can reliably characterize treatment effects of products in individual skin layers in agreement with physiological understanding.

Acne segmentation is normally considered as the first significant step in an automatic acne grading system, because segmentation accuracy directly influences the definition of acne pigmentation level, what has an impact on the goodness of acne severity evaluation. An initial unsupervised segmentation method is proposed for conventional RGB images, whose process is demonstrated in Fig. 13 (a). After several pre-processing steps (background and skin hair removal, illumination corrections), nine pigmentation descriptors were extracted from three RGB channels based on colorimetric transformations and absorption spectroscopy of major chromophores. It has been proved that the derived hemoglobin, normalized red, and normalized green descriptors can properly characterize pigmentation distributions of acne, and they are used as segmentation features. Finally, an iterative unsupervised segmentation was performed to maximize pigmentation distributions between acne and normal skin. Fig. 13 (b) shows an example of acne image on human face captured by a conventional RGB camera, while experimental result in Fig. 13 (f) illustrates that suspicious acne areas and healthy human skin can be automatically discriminated by applying the proposed method. Moreover, it only takes 90.8 seconds to segment the example image with the size of 640×428 pixels, which demonstrates the computation efficiency of the algorithm.

It should be noted that the segmentation method stated above is an initial approach. Shadows around non-flatten areas on human face (e.g. areas around nose) have a large influence on accuracy of automatic acne detection. However, based on the initial experimental results, it is difficult to entirely get rid of these effects using RGB channels only. Our finding is actually consistent with the existing studies, where researchers divided human face into several sub-regions and worked on these sub-regions individually to avoid shadow influence. Therefore, the next step study will compare acne segmentation results derived from RGB images and multi- or hyperspectral images, to investigate the most effective bands for describing acne pigmentation, as well as whether the introduction of multi- or hyperspectral analysis to the automatic acne detection and evaluation is necessary.

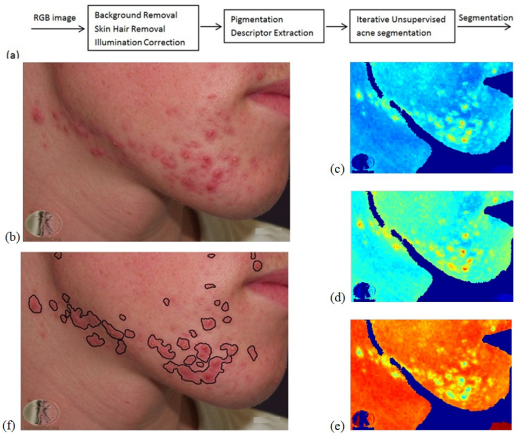


Figure 13. An initial method and result of automatic acne detection. (a) Specific steps in the acne segmentation method. (b) An example of acne disorder on human face (640×428 pixels) from DermnetNZ (http://www.dermnetnz.org/). (c)-(e) pigmentation descriptors: hemoglobin, normalized red, and normalized green, respectively. (f) Segmentation result outlined with black borders on original image.

IMEDIA2 Team

6. New Results

6.1. Feature space modeling

Participants: Vera Bakic, Nozha Boujemaa, Esma Elghoul, Hervé Goëau, Sofiene Mouine, Olfa Mzoughi, Anne Verroust-Blondet, Itheri Yahiaoui.

6.1.1. Spatial relations between salient points on a leaf

Participants: Sofiene Mouine, Itheri Yahiaoui, Anne Verroust-Blondet.

We have introduced a novel method for leaf species identification combining local and shape-based features. Our approach extends the shape context model in two ways:
- First of all, two different sets of points are distinguished when computing the shape contexts: the voting set, i.e. the points used to describe the coarse arrangement of the shape and the computing set containing the points where the shape contexts are computed.
Three shape descriptors are proposed, as illustrated in Figure 1 : SC0 (spatial relations between margin points), which corresponds to the original shape context; SC1 (spatial relations between salient points) where the voting set and the computing set are composed of the salient points of the image and SC2 (spatial relations between salient and margin points) where the voting set contains the margin points and the computing set consists of the salient points (see [11] for more details).

Figure 1. From left to right: points used in SC0, SC1 and SC2. The small circles represent the sample points on the leaf margin. The cross points represent the salient points computed with Harris detector.

-This representation is enriched by introducing local features computed in the neighborhood of the computing points.
We obtained excellent identification scores in the ImageCLEF 2012 plant identification task for scan and scan-like images of leaves (RUN2 in [20]).

6.1.2. Detection and extraction of leaf parts for plant identification

Participants: Olfa Mzoughi, Itheri Yahiaoui, Nozha Boujemaa.

Automatic plant identification is a relatively new research area in computer vision that has increasingly attracted high interest as a promising solution for the development of many botanical industries and for the success of biodiversity conservation. Most of the approaches proposed are based on the analysis of morphological properties of leaves. They have applied several well-known generic shape descriptors. Nevertheless, faced with the large amount of leaf species, botanical knowledge, especially about leaf parts (petiole, blade, insertion point, base and apex, rachis) is important to enhance their precision.
First of all, in order to extract them from leaf images, we introduced two types of symmetry in [12]: (i) the local translational symmetry, which is useful for petiole and rachis detection and (ii) the local symmetry of depth indentations, which is suited for base and apex detection.
Then, we studied the usefulness of parts detection (mainly petiole and insertion point) as a pre-processing stage for classic leaf shape retrieval schemes [13]. We showed that the removal of the petiole and the use of the insertion point as a starting point for the descriptors sensitive to the starting point improve retrieval results.

6.1.3. Multi-organ plant identification

Participants: Hervé Goëau, Vera Bakic, Souheil Selmi.

Inspired by citizen sciences, the main goal of this work is to speed up the collection and integration of raw botanical observation data, while providing to potential users an easy and efficient access to this botanical knowledge. We therefore designed and developed an original crowd-sourcing web application dedicated to the access of botanical knowledge through automated identification of plant species by visual content with multi-organ queries. Technically, the first side of the application deals with multi-organ content-based identification of plant. Indeed, most methods proposed for such automatic identification are actually based on leaf images, mostly based on leaf segmentation and boundary shape. However, leaves are far from being the only discriminant visual key between species and they are not visible all over the year for a large fraction of plant species. We propose to make the use of five different organs and plant's views including habit, flowers, fruits, leaves and bark. Thanks to an interactive and visual query widget, the tagging process of the different organs and views is as simple as drag-and-drop operations and does not require any expertise in botany.

Technically, as suggested by the results of ImageCLEF2011 for leaves [24], it is based on local features and large-scale matching. Interest points are detected with a modified color Harris detector, in order to favor points with a central position in the image and to reduce the impact of background features. Each interest point is then described with a SURF local feature and an HSV histogram. Automatic system-oriented and human-centered evaluations of the application show that the results are already satisfactory and therefore very promising in the long term to identify a richer flora. The second side of the application deals with interactive tagging and allows any user to validate or correct the automatic determinations returned by the system. Overall, this collaborative system enables the automatic and continuous enrichment of the visual botanical knowledge and therefore it increases progressively the accuracy of the automated identification. This application called 'Identify' (cf. Figure 2 and http://identify.plantnet-project.org) has been presented at the first ACM International workshop on Multimedia Analysis and Ecological Data [8]. This work has been done in collaboration with Inria team ZENITH and with the botanists of the AMAP UMR team (CIRAD). It is also closely related to a citizen science project around plant's identification that we developed with the support of the Tela Botanica social network inside the Pl@ntNet project.

6.1.4. Segmentation transfer method for articulated models

Participants: Esma Elghoul, Anne Verroust-Blondet.

Mesh segmentation consists in partitioning the surface into a set of patches that are uniform with respect to a given property. We are interested in retaining the semantic information during the segmentation. A particularly challenging task is then the automatic identification of semantically meaningful parts of a 3D model, which can be hard to achieve when only the shape geometry is considered. We have introduced a method using a pre-segmented example model to perform semantic-oriented segmentation of non-rigid 3D models of the same class (human, octopus, quadrupeds, etc.). Using the fact that the same type of non-rigid models share the same global topological structure, we exploit coarse topological shape attributes in conjunction with a seed-based segmentation approach to transfer a meaningful and consistent segmentation from the example mesh to the

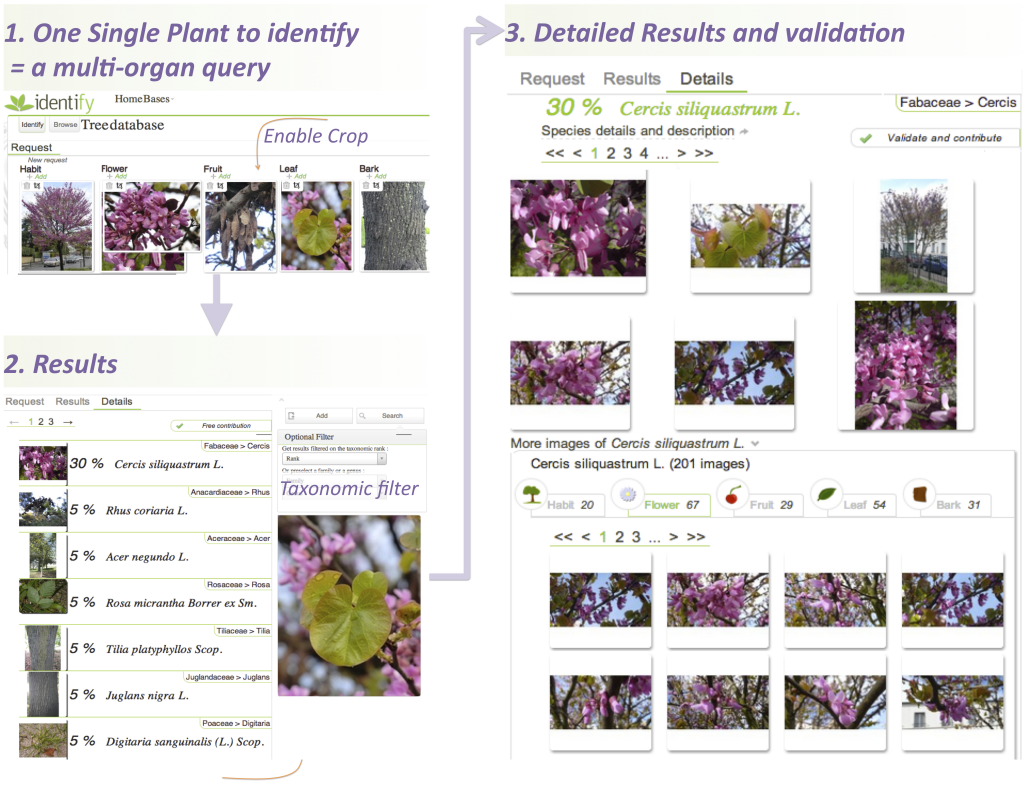


Figure 2. Example of a multi-organ query on one single plant submitted in the application.

target models. Promising results have been obtained on classes of articulated models (cf. Figure 3). This work has been submitted for publication.

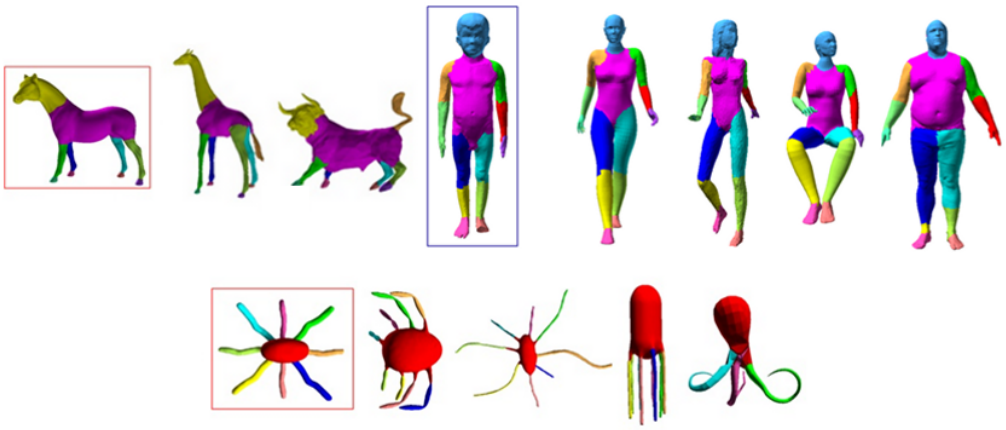


Figure 3. Segmentation transfer results on the quadrupeds, on the humans and on the octopus class. The exemplar segmentations are framed in red or in blue.

6.2. Feature space structuring

Participants: Nozha Boujemaa, Hervé Goëau, Amel Hamzaoui, Saloua Ouertani-Litayem, Mohamed Riadh Trad.

6.2.1. Plant Leaves Morphological Categorization with Shared Nearest Neighbors Clustering

Participants: Amel Hamzaoui, Hervé Goëau, Nozha Boujemaa.

In [9] we present an original experiment aimed at evaluating if state-of-the-art visual clustering techniques are able to automatically recover morphological classifications built by the botanists themselves. The clustering phase is based on a recent Shared-Nearest Neighbors (SNN) clustering algorithm, which allows combining effectively heterogeneous visual information sources at the category level. Each resulting cluster is associated with an optimal selection of visual similarities, allowing discovering diverse and meaningful morphological categories even if we use a blind set of visual sources as input. Experiments have been performed on ImageCLEF 2011 plant identification dataset [23], specifically enriched in this work with morphological attributes tags (annotated by expert botanists). The results presented in Figure 4 are very promising, since all clusters discovered automatically can be easily matched to one node of the morphological tree built by the botanists. This work is also described in details in Amel Hamzaoui's thesis [4].

6.2.2. Distributed KNN-Graph approximation via Hashing

Participants: Mohamed Riadh Trad, Nozha Boujemaa.

High dimensional data hashing is essential for scaling up and distributing data analysis applications involving feature-rich objects, such as text documents, images or multi-modal entities (scientific observations, events, etc.). In this first research track, we first investigated the use of high dimensional hashing methods for efficiently approximating K-NN Graphs [16], [19], [17], particularly in distributed environments. We highlighted the importance of balancing issues on the performance of such approaches and show why the baseline approach

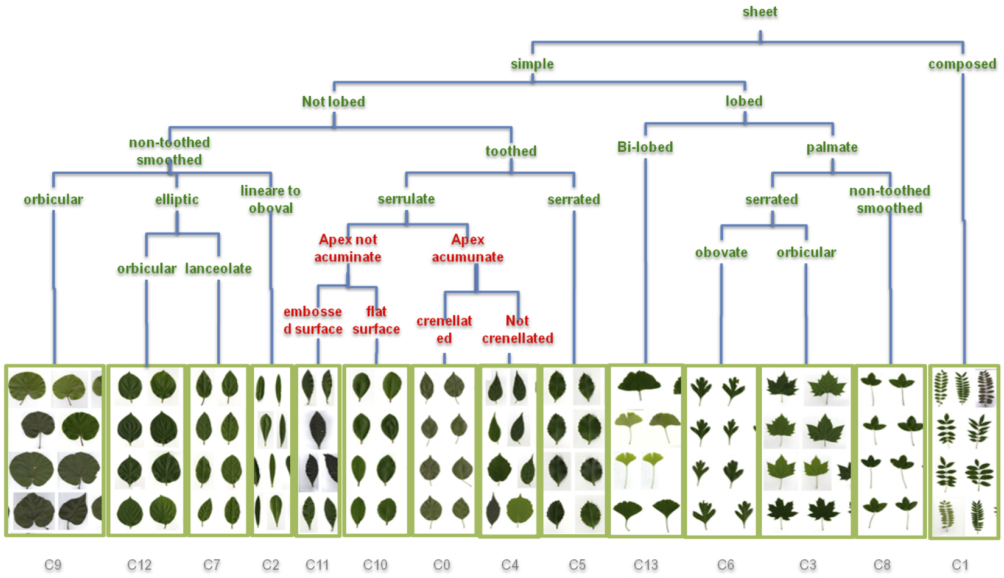


Figure 4. Hierarchical tree organization of the clusters produced by the SNN clustering method on the scan leaf dataset ImageCLEF2011.

using Locality Sensitive Hashing does not perform well. Our new KNN-join method is based on RMMH, a hash function family based on randomly trained classifiers that we introduced in 2011. We show that the resulting hash tables are much more balanced and that the number of resulting collisions can be greatly reduced without degrading quality. We further improve the load balancing of our distributed approach by designing a parallelized local join algorithm, implemented within the MapReduce framework.

6.2.3. Hash-Based Support Vector Machines Approximation for Large Scale Prediction

Participants: Saloua Ouertani-Litayem, Nozha Boujemaa.

How-to train effective classifiers on huge amount of multimedia data is clearly a major challenge that is attracting more and more research works across several communities. Less efforts however are spent on the counterpart scalability issue: how to apply big trained models efficiently on huge non annotated media collections ? In [10], we addressed the problem of speeding-up the prediction phase of linear Support Vector Machines via Locality Sensitive Hashing. We proposed building efficient hash-based classifiers that are applied in a first stage in order to approximate the exact results and filter the hypothesis space. Experiments performed with millions of one-against-one classifiers show that the proposed hash-based classifier can be more than two orders of magnitude faster than the exact classifier with minor losses in quality (cf. Figure 5).

6.3. Pattern recognition and statistical learning

Participants: Nozha Boujemaa, Michel Crucianu, Donald Geman, Wajih Ouertani, Asma Rejeb Sfar.

6.3.1. Machine identification of biological shapes

Participants: Asma Rejeb Sfar, Donald Geman, Nozha Boujemaa.

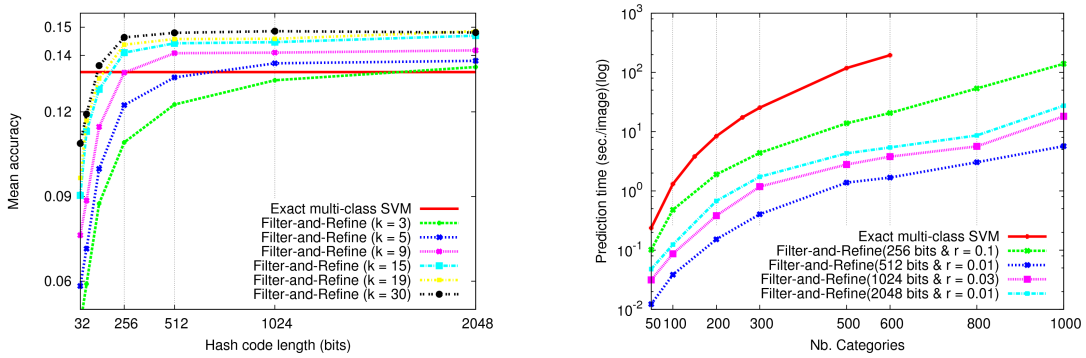


Figure 5. Exact Multi-class SVM vs HBMS based Filter-And-Refine method in terms of accuracy and prediction time

The increasing availability of digital images in the traditional sciences, the growing interest in biodiversity and the ongoing shortage of skilled taxonomists combine to make the automated categorization algorithms, increasingly important in many fields such as botany, agriculture and medicine. In this work, we propose a hierarchical coarse-to-fine approach to identify botanical species from a scanned sample of a plant organ, e.g., a leaf or a flower. To this end, we exploit domain-specific knowledge about taxonomy and landmarks. Promising recognition rates are achieved on several leaf datasets. Results have been submitted for publication.

6.3.2. Relevance feedback on partial image query

Participants: Wajih Ouertani, Michel Crucianu, Nozha Boujemaa.

scalability, hashing, SVM, prediction, approximation

Even if cropping an image to perform one-shot partial query filters a considerable amount of senseless regions for target definition, it does not yet clearly illustrates what the user is looking for. Indeed, the user target is either closer to the instance level or to the category level. Then we may have numerous suggested examples within the first response ranks while possibly some of them are totally irrelevant examples.

We claim that a localization interaction is still more appropriate than having a holistic decision about image relevance if it is performed on more examples. We go beyond the first partial query and investigate machine learning process to learn intention iteratively and interactively. Our learning process is based on what user delimit within additional images taken from the first response ranks. Our motivations include dealing with semantic gap revealed by local features hit falling into false regions within retrieved images. Those images might be either totally irrelevant, where all partial zones are out of the interest, or partially relevant, not because of the zones expected by the system (false-localization) but rather because of some missed zones. Through local annotations we expect the ability of redirecting the recognition session to those relevant regions and studying how much we can reduce the semantic gap within interactive localization.

This year, we studied several learning strategies based on several assumptions heuristically extracted on user interaction. The presented strategies have been also combined with features filtering within object representation. The filtering includes grouping contextualizing and varying features set representations.

6.4. IKONA/MAESTRO sofware

Participants: Vera Bakic, Laurent Joyeux, Sofiene Mouine, Souheil Selmi.

This year, IKONA has been extended in the context of Pl@ntNet, Glocal and I-SEARCH projects.

- For the Pl@ntNet project, along the continuing improvements and optimizations in the MAESTRO software, a number of new features were added:

new options for interest points distribution and filtering with the segmented image,
a new shape context descriptor (corresponding to [11]),
various combinations of descriptors in one vector per interest point or region,
for regions: extraction of sub-images, EOH and Fourier descriptors,
more options for update of calculated signatures, new score type (used in ImageCLEF2012) and decision rules (adaptive Knn calculation based on individual plant information) for statistical tools.
In addition, a number of new web services and functionality were developed/updated and deployed:
the addition of new databases (Vignes, Musa), while some with the organ annotation (Photoflora, Girod), and the update of the multi search views for the new datasets;
the development of new services allowing to return of botanical information in several formats (csv, xml, JSON...);
the update of the indexation system so that it can crawl images from different sources (internal or external sources like CEL web service, which uses html or identiplante web service (in JSON format);
the development of html pages to annotate a set of images by organ type;
the development of an applet demonstration of leaf architecture (this applet interfaces a library developed by a PhD student working on this project);
the development of an application "Pl@ntNet Identify" for android platform and its interfacing with the existing web services.

- For the Glocal project, functionalities such as fraud detection and similarity search were integrated in the mock up of user interface and in the final demonstration of the project. In addition, a dynamic indexation system of images from AFP (Agence France-Presse) was implemented as well as the similarity web services working on this dynamic dataset.
- For the I-SEARCH project, an integration of the video mining component in I-SEARCH platform was done. The component extracts visual objects that are the most recurrent from a set of images, or in a video.

LEAR Project-Team

6. New Results

6.1. Visual recognition in images

6.1.1. Correlation-Based Burstiness for Logo Retrieval

Participants: Matthijs Douze, Jerome Revaud, Cordelia Schmid.

Detecting logos in photos is challenging. A reason is that logos locally resemble patterns frequently seen in random images. In [21] we propose to learn a statistical model for the distribution of incorrect detections output by an image matching algorithm. It results in a novel scoring criterion in which the weight of correlated keypoint matches is reduced, penalizing irrelevant logo detections. In experiments on two very different logo retrieval benchmarks, our approach largely improves over the standard matching criterion as well as other state-of-the-art approaches.

Figure 1. Illustration of a logo detected by our method.

6.1.2. Towards Good Practice in Large-Scale Learning for Image Classification

Participants: Zeynep Akata, Zaid Harchaoui, Florent Perronnin [XRCE], Cordelia Schmid.

In [19] we propose a benchmark of several objective functions for large-scale image classification: we compare the one-vs-rest, multiclass, ranking and weighted average ranking SVMs. Using stochastic gradient descent optimization, we can scale the learning to millions of images and thousands of classes. Our experimental evaluation shows that ranking based algorithms do not outperform a one-vs-rest strategy and that the gap between the different algorithms reduces in case of high-dimensional data. We also show that for one-vs-rest, learning through cross-validation the optimal degree of imbalance between the positive and the negative samples can have a significant impact. Furthermore, early stopping can be used as an effective regularization strategy when training with stochastic gradient algorithms. Following these "good practices", we were able to improve the state-of-the-art on a large subset of 10K classes and 9M of images of ImageNet from 16.7% accuracy to 19.1%. Some qualitative results can be seen in Figure 2 .

6.1.3. Discriminative Spatial Saliency for Image Classification

Participants: Frédéric Jurie [Université de Caen], Cordelia Schmid, Gaurav Sharma.

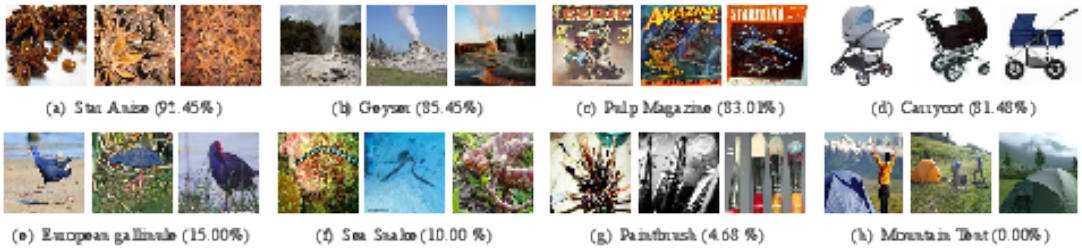


Figure 2. ImageNet10K results (top-1 accuracy in %) obtained with w-OVR and 130K-dim Fisher vectors. (a-d) Sample classes among the best performing ones. (e-h) Sample classes among the worst performing ones.

In many visual classification tasks the spatial distribution of discriminative information is (i) non uniform e.g. "person reading" can be distinguished from "taking a photo" based on the area around the arms i.e. ignoring the legs, and (ii) has intra class variations e.g. different readers may hold the books differently. Motivated by these observations, we propose in [22] to learn the discriminative spatial saliency of images while simultaneously learning a max-margin classifier for a given visual classification task. Using the saliency maps to weight the corresponding visual features improves the discriminative power of the image representation. We treat the saliency maps as latent variables and allow them to adapt to the image content to maximize the classification score, while regularizing the change in the saliency maps. See Figure 3 for an illustration. Our experimental results on three challenging datasets, for (i) human action classification, (ii) fine grained classification, and (iii) scene classification, demonstrate the effectiveness and wide applicability of the method.

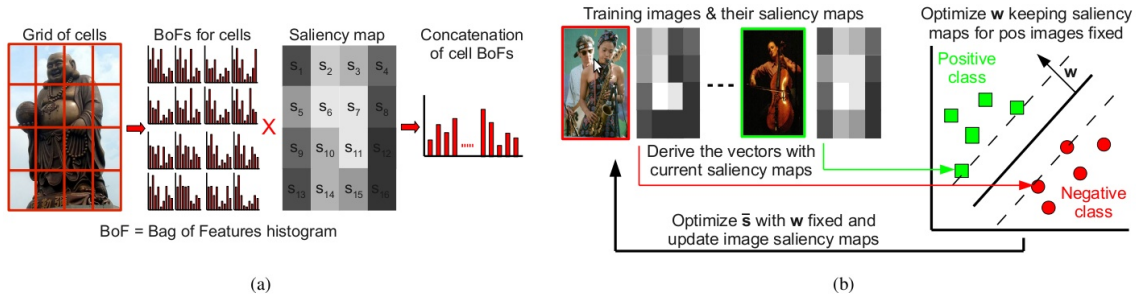


Figure 3. (a) The images are represented by concatenation of cell bag-of-features weighted by the image saliency maps. (b) We propose to use a block coordinate descent algorithm for learning our model. As in a latent SVM, we optimize in one step the weight vector w keeping the saliency maps of the positive images fixed, and in the other step we optimize the saliency keeping w fixed.

6.1.4. Tree-structured CRF Models for Interactive Image Labeling

Participants: Gabriela Csurka [XRCE], Thomas Mensink, Jakob Verbeek.

In [8] we propose structured prediction models for image labeling that explicitly take into account dependencies among image labels. In our tree structured models, image labels are nodes, and edges encode dependency relations. To allow for more complex dependencies, we combine labels in a single node, and use mixtures of

trees. Our models are more expressive than independent predictors, and lead to more accurate label predictions. The gain becomes more significant in an interactive scenario where a user provides the value of some of the image labels at test time. Such an interactive scenario offers an interesting trade-off between label accuracy and manual labeling effort. The structured models are used to decide which labels should be set by the user, and transfer the user input to more accurate predictions on other image labels. We also apply our models to attribute-based image classification, where attribute predictions of a test image are mapped to class probabilities by means of a given attribute-class mapping. Experimental results on three publicly available benchmark data sets show that in all scenarios our structured models lead to more accurate predictions, and leverage user input much more effectively than state-of-the-art independent models.

6.1.5. Metric Learning for Large Scale Image Classification: Generalizing to new classes at near-zero cost

Participants: Gabriela Csurka [XRCE], Thomas Mensink, Florent Perronnin [XRCE], Jakob Verbeek.

In [18], [27] we consider the task of large scale image classification in open ended datasets. Many real-life datasets are open-ended and dynamic: new images are continuously added to existing classes, new classes appear over time and the semantics of existing classes might evolve too. In order to be able to handle new images and new classes at near-zero cost we consider two distance based classifiers, the k-nearest neighbor (k-NN) and nearest class mean (NCM) classifiers. For the NCM classifier we introduce a new metric learning approach, which has advantageous properties over the classical Fisher Discriminant Analysis. We also introduce an extension of the NCM classifier to allow for richer class representations, using multiple centroids per class. Experiments on the ImageNet 2010 challenge dataset, which contains over one million training images of thousand classes, show that, surprisingly, the NCM classifier compares favorably to the more flexible k-NN classifier. Moreover, the NCM performance is comparable to that of linear SVMs which obtain current state-of-the-art performance. Experimentally we study the generalization performance to classes that were not used to learn the metrics. Using a metric learned on 1,000 classes, we show results for the ImageNet-10K dataset which contains 10,000 classes, and obtain performance that is competitive with the current state-of-the-art, while being orders of magnitude faster. Furthermore, we show how a zero-shot class prior based on the ImageNet hierarchy can improve performance when few training images are available. See Figure 4 for an illustration.

6.2. Learning and statistical models

6.2.1. Image categorization using Fisher kernels of non-iid image models

Participants: Ramazan Cinbis, Cordelia Schmid, Jakob Verbeek.

Bag of visual words treat images as an orderless sets of local regions and represent them by visual word frequency histograms. Implicitly, regions are assumed to be identically and independently distributed (iid), which is a very poor assumption from a modeling perspective; see Figure 5 for an illustration. In [13], we introduce non-iid models by treating the parameters of bag-of-word models as latent variables which are integrated out, rendering all local regions dependent. Using the Fisher kernel we encode an image by the gradient of the data log-likelihood with respect to hyper-parameters that control priors on the model parameters. In fact, our models naturally generate transformations similar to taking square-roots, providing an explanation of why such non-linear transformations have proven successful. Using variational inference we extend the basic model to include Gaussian mixtures over local descriptors, and latent topic models to capture the co-occurrence structure of visual words, both improving performance. Our models yields state-of-the-art image categorization performance using linear classifiers, without using non-linear kernels, or (approximate) explicit embeddings thereof, e.g. by taking the square-root of the features.

6.2.2. Conditional gradient algorithms for machine learning

Participants: Zaid Harchaoui, Anatoli Juditsky [UJF], Arkadi Nemirovski [Georgia Tech].

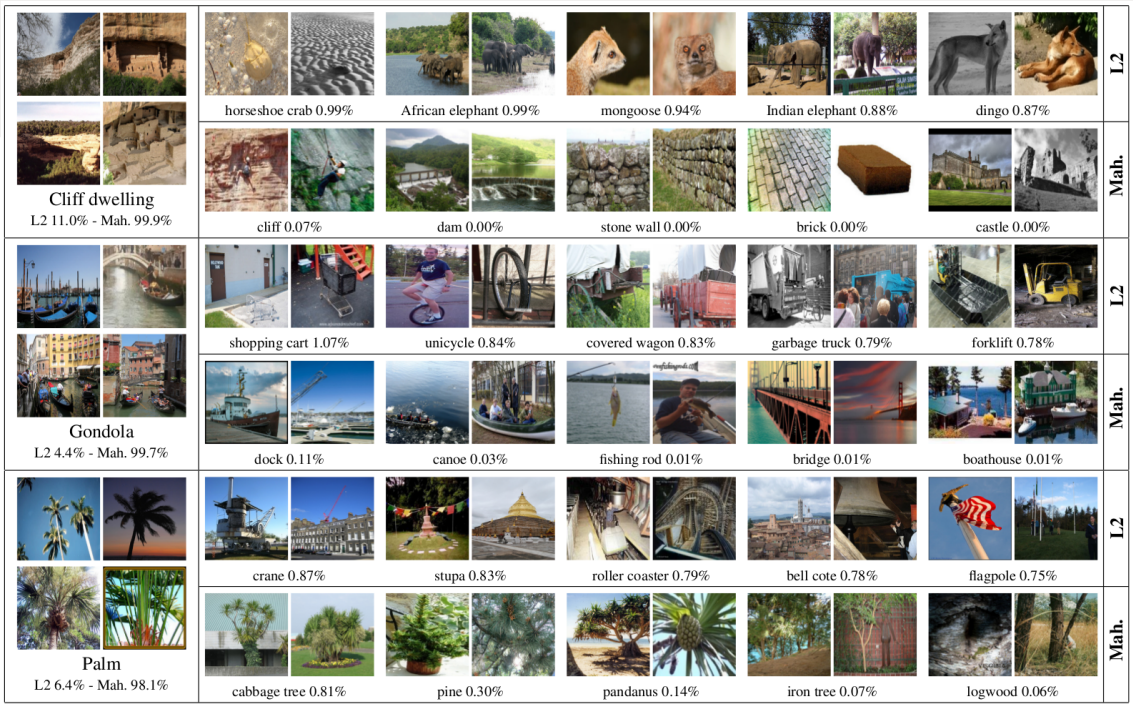


Figure 4. Examples of three classes, and the five most similar classes for each according to the standard ℓ_2 metric and our learned Mahalanobis metric.

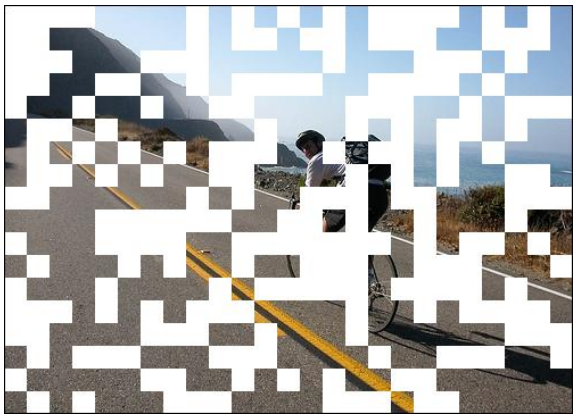


Figure 5. Illustration of why local image patches are not independent: we can easily guess the image content in the masked areas.

In [17] we consider convex optimization problems arising in machine learning in high-dimensional settings. For several important learning problems, such as e.g. noisy matrix completion, state-of-the-art optimization approaches such as composite minimization algorithms are difficult to apply and do not scale up to large datasets. We study three conditional gradient-type algorithms, suitable for large-scale problems, and derive their finite-time convergence guarantees. Promising experimental results are presented on two large-scale real-world datasets.

6.2.3. Large-scale classification with trace-norm regularization

Participants: Matthijs Douze, Miro Dudik [Microsoft Research], Zaid Harchaoui, Jérôme Malick [BiPoP Team Inria Grenoble], Mattis Paulin [ETHZ].

In [16] we introduce a new scalable learning algorithm for large-scale multi-class image classification, based on the multinomial logistic loss and the trace-norm regularization penalty. Reframing the challenging non-smooth optimization problem into a surrogate infinite-dimensional optimization problem with a regular ℓ_1-regularization penalty, we propose a simple and provably efficient accelerated coordinate descent algorithm. Furthermore, we show how to perform efficient matrix computations in the compressed domain for quantized dense visual features, scaling up to 100,000s examples, 1,000s-dimensional features, and 100s of categories. Promising experimental results on the "Fungus", "Ungulate", and "Vehicles" subsets of ImageNet are presented, where we show that our approach performs significantly better than state-of-the-art approaches for Fisher vectors with 16 Gaussians.

6.2.4. Tree-walk kernels for computer vision

Participants: Francis Bach [Inria SIERRA team], Zaid Harchaoui.

In [25] we propose a family of positive-definite kernels between images, allowing to compute image similarity measures respectively in terms of color and of shape. The kernels consists in matching subtree-patterns called "tree-walks" of graphs extracted from the images, e.g. the segmentation graphs for color similarity and graphs of the discretized shapes or the point clouds in general for shape similarity. In both cases, we are able to design computationally efficient kernels which can be computed in polynomial-time in the size of the graphs, by leveraging specific properties of the graphs at hand such as planarity for segmentation graphs or factorizability of the associated graphical model for point clouds. Our kernels can be used by any kernel-based learning method, and hence we present experimental results for supervised and semi-supervised classification as well as clustering of natural images and supervised classification of handwritten digits and Chinese characters from few training examples.

6.2.5. Lifted coordinate descent for learning with trace-norm regularization

Participants: Miro Dudik [Microsoft Research], Zaid Harchaoui, Jérôme Malick [BiPoP Team Inria Grenoble].

In [14] we consider the minimization of a smooth loss with trace-norm regularization, which is a natural objective in multi-class and multi-task learning. Even though the problem is convex, existing approaches rely on optimizing a non-convex variational bound, which is not guaranteed to converge, or repeatedly perform singular-value decomposition, which prevents scaling beyond moderate matrix sizes. We lift the non-smooth convex problem into an infinitely dimensional smooth problem and apply coordinate descent to solve it. We prove that our approach converges to the optimum, and is competitive or outperforms the state of the art.

6.3. Recognition in video

6.3.1. Large-scale multi-media event detection in video

Participants: Matthijs Douze, Zaid Harchaoui, Dan Oneata, Danila Potapov, Jerome Revaud, Cordelia Schmid, Jochen Schwenninger [Fraunhofer Institute, Bonn], Jakob Verbeek, Heng Wang.

This year we participated in the TrecVid Multimedia Event Detection (MED) task. The goal is to detect events categories (such as "birthday party", or "changing a vehicle tire") in a large collection of around 100,000 videos with a total duration of around 4,000 hours. To this end we implemented an efficient system based on our recently developed MBH video descriptor (see Section 5.4) , SIFT descriptors and, MFCC audio descriptors (contributed by Fraunhofer Institute). All these low-level descriptors are encoded using the Fisher vector representation (see Section 5.3). In addition we implemented an optical character recognition (OCR) system to extract textual features from the video. The system is described in a forthcoming paper [31], and ranked first and second in two evaluations among the 17 systems submitted by different international teams participating to the task. See Figure 6 for an illustration.

Figure 6. Illustration of videos retrieved for two event categories. From left to right, we show for each a frame from (i) the top ranked video, (ii,iii) the first negative video, and the postive just before, and (iv) the last positive video.

6.3.2. Learning Object Class Detectors from Weakly Annotated Video

Participants: Javier Civera, Vittorio Ferrari, Christian Leistner, Alessandro Prest, Cordelia Schmid.

Object detectors are typically trained on a large set of still images annotated by bounding-boxes. In [20] we introduce an approach for learning object detectors from real-world web videos known only to contain objects of a target class. We propose a fully automatic pipeline that localizes objects in a set of videos of the class and learns a detector for it. The approach extracts candidate spatio-temporal tubes based on motion segmentation and then selects one tube per video jointly over all videos. See Figure 7 for an illustration. To compare to the state of the art, we test our detector on still images, i.e., Pascal VOC 2007. We observe that frames extracted from web videos can differ significantly in terms of quality to still images taken by a good camera. Thus, we formulate the learning from videos as a domain adaptation task. We show that training from a combination of weakly annotated videos and fully annotated still images using domain adaptation improves the performance of a detector trained from still images alone.

6.3.3. Recognizing activities with cluster-trees of tracklets

Participants: Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

In [15] we address the problem of recognizing complex activities, such as pole vaulting, which are characterized by the composition of a large and variable number of different spatio-temporal parts. We represent a video as a hierarchy of mid-level motion components. This hierarchy is a data-driven decomposition specific to each video. We introduce a divisive clustering algorithm that can efficiently extract a hierarchy over a large

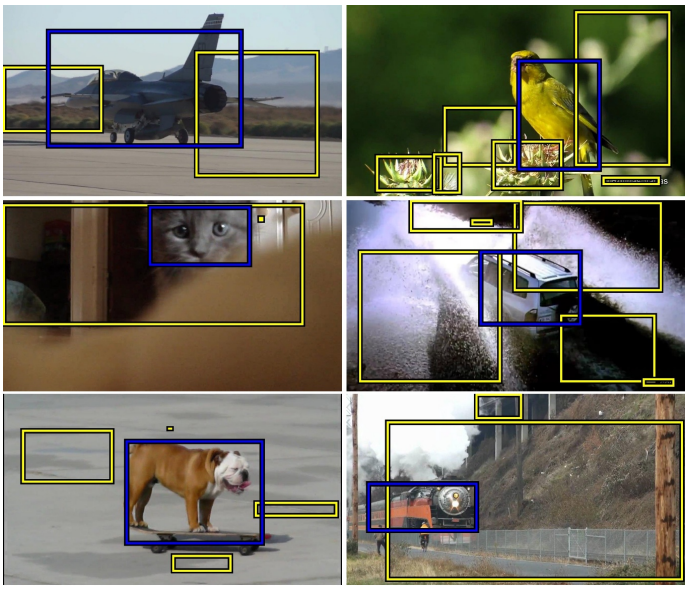


Figure 7. Yellow boxes represent tubes extracted by our method on the YouTube-Objects dataset. Blue boxes indicate the automatically selected tubes.

number of local trajectories. We use this structure to represent a video as an unordered binary tree. This tree is modeled by nested histograms of local motion features, see Figure 8 . We provide an efficient positive definite kernel that computes the structural and visual similarity of two tree decompositions by relying on models of their edges. Contrary to most approaches based on action decompositions, we propose to use the full hierarchical action structure instead of selecting a small fixed number of parts. We present experimental results on two recent challenging benchmarks that focus on complex activities and show that our kernel on per-video hierarchies allows to efficiently discriminate between complex activities sharing common action parts. Our approach improves over the state of the art, including unstructured activity models, baselines using other motion decomposition algorithms, graph matching, and latent models explicitly selecting a fixed number of parts.

6.3.4. Action Detection with Actom Sequence Models

Participants: Adrien Gaidon, Zaid Harchaoui, Cordelia Schmid.

We address the problem of detecting actions, such as drinking or opening a door, in hours of challenging video data. In [26] we propose a model based on a sequence of atomic action units, termed "actoms", that are semantically meaningful and characteristic for the action. Our Actom Sequence Model (ASM) represents the temporal structure of actions as a sequence of histograms of actom-anchored visual features, see Figure 9 for an illutration. Our representation, which can be seen as a temporally structured extension of the bag-of-features, is flexible, sparse, and discriminative. Training requires the annotation of actoms for action examples. At test time, actoms are detected automatically based on a non-parametric model of the distribution of actoms, which also acts as a prior on an action's temporal structure. We present experimental results on two recent benchmarks for temporal action detection: "Coffee and Cigarettes" and the "DLSB" dataset. We also adapt our approach to a classification by detection set-up and demonstrate its applicability on the challenging "Hollywood 2" dataset. We show that our ASM method outperforms the current state of the art in temporal action detection, as well as baselines that detect actions with a sliding window method combined with bag-of-features.

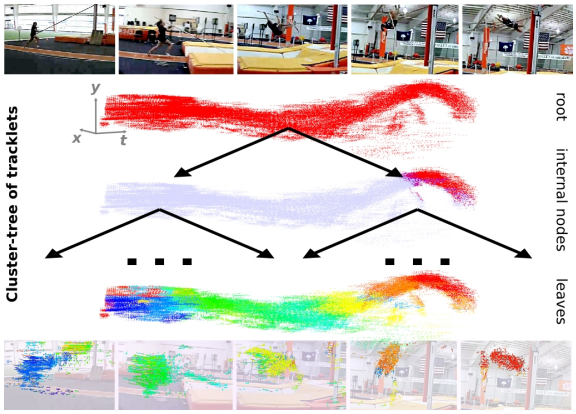


Figure 8. Illustration of tracklets found in a video and their hierarchical decomposition.

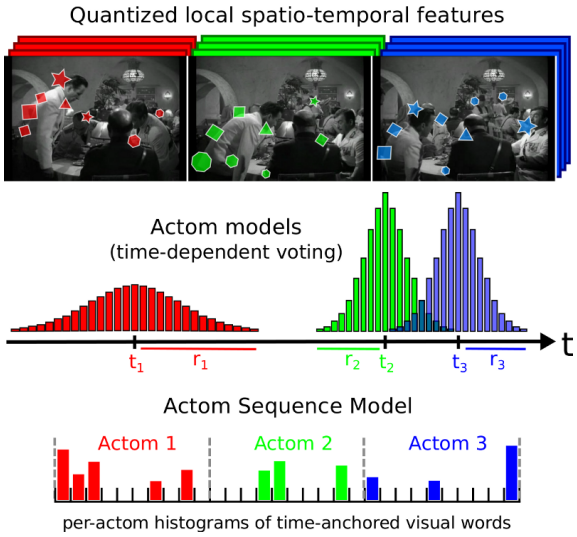


Figure 9. Illustration of the "Actom" video representation, see text for details.

6.3.5. Action recognition by dense trajectories

Participants: Alexander Kläser, Cheng-Lin Liu [Chinese Academy of Sciences], Cordelia Schmid, Heng Wang [Chinese Academy of Sciences].

In [28] we introduce a video representation based on dense trajectories and motion boundary descriptors. Trajectories capture the local motion information of the video. A state-of-the-art optical flow algorithm enables a robust and efficient extraction of the dense trajectories. As descriptors we extract features aligned with the trajectories to characterize shape (point coordinates), appearance (histograms of oriented gradients) and motion (histograms of optical flow). Additionally, we introduce a descriptor based on motion boundary histograms (MBH) (see the visualization in Figure 10), which is shown to consistently outperform other state-of-the-art descriptors, in particular on real-world videos that contain a significant amount of camera motion. We evaluate our video representation in the context of action classification on nine datasets, namely KTH, YouTube,Hollywood2, UCF sports, IXMAS, UIUC, Olympic Sports, UCF50 and HMDB51. On all datasets our approach outperforms current state-of-the-art results.

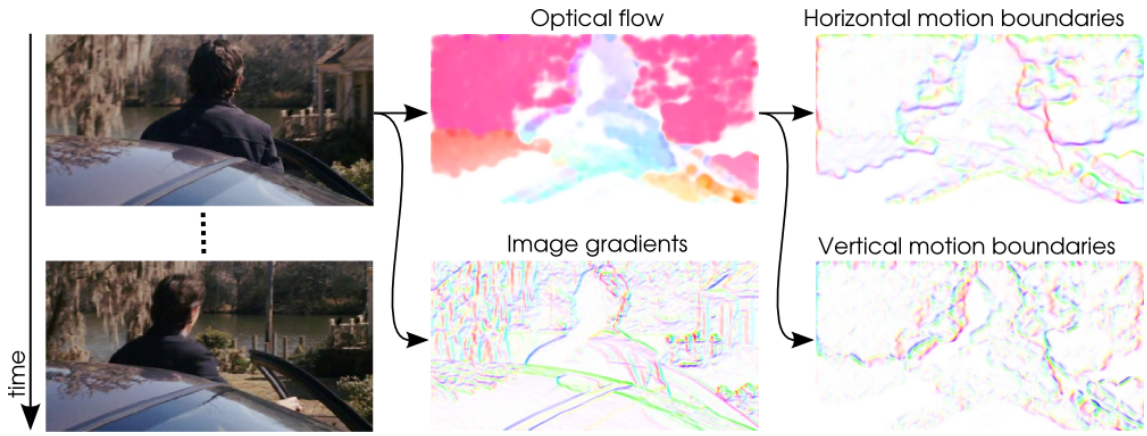


Figure 10. Illustration of the information captured by HOG, HOF, and MBH descriptors. Gradient/flow orientation is indicated by color (hue) and magnitude by saturation. The optical flow (top, middle) shows constant motion in the background, which is due to the camera movements. The motion boundaries (right) encode the relative motion between the person and the background.

MAGRIT Project-Team

6. New Results

6.1. Motion, scene and camera reconstruction

Participants: Marie-Odile Berger, Srikrishna Bhat, Christel Leonet, Gilles Simon, Frédéric Sur.

- **Enhancing the grid method for in-plane strain measurements**
  This work is motivated by a problem from experimental solid mechanics. The grid method permits to measure the heterogeneous strains on the surface of specimens subjected to mechanical tests. Among full-field measurement techniques, the grid method consists in transferring a regular grid on the surface of the specimen and in taking images of the grid before and after deformation. Windowed Fourier analysis then gives an estimate of the surface displacement and strain components. In a collaboration with Institut Pascal (Université Blaise Pascal, Clermont Ferrand), we have shown that the estimations obtained by this technique are approximately the convolution of the actual values with the analysis window. We have also characterized how the noise in the grid image impairs the displacement and strain maps [18]. This study has allowed us to improve the metrological performance of the grid method with deconvolution algorithms. A numerical and experimental study can be found in [17].
- **Visual words for pose computation**
  Visual vocabularies are standard tools in the object/image classification literature, and are emerging as a new tool for building point correspondences for pose estimation. Within S. Bhat's PhD thesis, we have proposed several methods for visual word construction dedicated to point matching, with structure from motion and pose estimation applications in view. The three dimensional geometry of a scene is first extracted with bundle adjustment techniques based on keypoint correspondences. These correspondences are obtained by grouping the set of all SIFT descriptors from the training images into visual words using transitive closure (TC) techniques. We obtain a more accurate 3D geometry than with classical image-to-image point matching. In a second on-line step, these visual words serve as 3D point descriptors that are robust to viewpoint change, and are used for building 2D-3D correspondences on-line during application, yielding the pose of the camera by solving the PnP problem. Several visual word formation techniques have been compared with respect to robustness to viewpoint change between the learning and the test images. Our experiments showed that the adaptive TC visual words are better in many ways when compared to other classical techniques such as K-means.

  More specifically, the work of this year has focused on improving pose estimation from visual words with respect to strong viewpoint changes. 2D-3D correspondences are actually difficult to establish if there are too large viewpoint changes between the image whose pose is sought and the images that yielded the visual words attached to 3D points. We assessed several viewpoint simulation techniques in order to enrich the visual word description of the 3D points.
- **Acquisition of 3D calibrated data**
  Christel Leonet joined the team in October 2010 as an Inria assistant engineer with the aim of building an integrated 3D acquisition system. More specifically, the objective of her work is to combine an IMU (Inertial Measurement Unit), a GPS receiver, a laser rangefinder and a video camera for ground truth data acquisitions of camera movements and scene structures. These data will be useful to validate several algorithms developed in our team. This year, a new visual pan tracking method has been designed and implemented. We considered spherical environments made of sparse video images instead of fully-covered environment maps which often suffer from geometric and photometric misalignments. The scanning process has been improved in order to increase the accuracy of the recovered polygons and allow for visual assessments of this accuracy. The 3D laser pointer has been validated in several indoor environments. Finally, the GPS has been integrated to the system and preliminary results have been obtained in outdoor environments.

6.2. Medical imaging

Participants: René Anxionnat, Marie-Odile Berger, Nazim Haouchine, Erwan Kerrien, Pierre-Frédéric Villard, Brigitte Wrobel-Dautcourt, Ahmed Yureidini.

- **Vessel reconstruction with implicit surfaces**
  This research activity is led in collaboration with Shacra project-team from Inria Lille-Nord Europe and the Department of Interventional Neuroradiology from Nancy University Hospital. It was pursued this year in the context of the SOFA-InterMedS Inria Large-Scale Initiative (http://www.sofa-framework.org/).

  Our objective is to offer the interventional radiologists with a patient-based interactive simulator [16]. The medical applications are training to endovascular procedures, planning the intervention, and augmenting the intra-operative images with 3D simulated data. Our contributions address vasculature modeling from patient data, namely 3D rotational angiography (3DRA) volumes. The segmentation should be both user friendly and generate a vascular surface model that is compliant with the computing constraints set in interactive simulation. Within A. Yureidini's PhD thesis, a new model was developed consisting of a tree of local implicit blobby models. The algorithm consists of two steps: first, a vessel tracking step to extract the vessel topology and, second, fitting local surface data points with implicit blobby models at each node point on the vessel centerline.

  An extensive validation of our RANSAC-based vessel tracking algorithm was performed [14], by comparison with state of the art Multiple Hypothesis Testing [19] on 10 patient data. Fitting the implicit model to patient data relies on the minimization of a multi-termed energy. A closed form solution was derived, and a blob selection and subdivision heuristic was described to implement an efficient energy minimization algorithm. Both the geometric accuracy and compactness of the resulting vascular models were shown to be excellent [15].

  Our current goals are: first, to further enhance model compactness by relying on the robustness and versatility of the modeling algorithm and using sparser vascular centerline trees; second, to mathematically ensure the continuity between neighboring local implicit models; and third, to reintroduce the raw image data for a more accurate energy computation, with the aim to design a blobby deformable model.

  This model was implemented in Sofa simulation platform, enabling interactive simulation time and thereby showing an impressive realism during tool navigation. On-going preliminary medical evaluation is being carried on by our fellow interventional radiologist in the framework of intervention planning.

- **Designing respiration models for patient based simulators**
  The work presented here has been done within a collaboration with Imperial College of London, Bangor University and Inria Aviz team.

  Respiratory models could be a key component in increasing realism in medical simulators. We have previously developed such kind of model. However finding the good parameters to tune the model so that it corresponds to a real patient behavior is not an easy task.

  This year, we have studied methods to automatically tune the elasticity of soft-tissues and the respiratory model parameters based on patient data. The estimation is based on two 3D Computed Tomography scans of the same patient at two different time steps. The parametrization of the model is considered as an inverse problem. Optimization techniques have then been deployed to solve the problem.

  In [13], we used a random search algorithm to generate a given number of sets of 15 random parameters. The set of parameters that provides the lowest fitness is extracted and corresponds to the solution of the optimization problem.

In [9], we have made use of an ad-hoc evolutionary algorithm that is able to explore a search space with 15 dimensions. Our method is fully automatic and auto-adaptive. A compound fitness function has been designed to account for various quantities that have to be minimized. The algorithm efficiency was experimentally analyzed on several real test-cases: i) three patient datasets have been acquired with the "breath hold" protocol, and ii) two datasets corresponds to 4D CT scans. The performance was compared with two traditional methods (downhill simplex and conjugate gradient descent), our random search method and a basic real-valued genetic algorithm. The results showed that our evolutionary scheme provides more significantly stable and accurate results.

- **Physics-based augmented reality**
  The development of AR systems for use in the medical field faces one major challenge: the correct superposition of pre-operative data onto intraoperative images. This task is especially difficult when laparospic surgery is considered since superposition must be achieved on deformable organs. Most existing AR systems only consider rigid registration between the pre and intraoperative data and the transformation is often computed interactively or from markers attached to the patient's body. In cooperation with the Shacra team, we have introduced an original method to perform augmented or mixed reality on deformable objects. Compared to state-of-the-art techniques, our method is able to track deformations of volumetric objects and not only surfacic objects. A flexible framework that relies on the combination of 3D motion estimation obtained from stereoscopic data and a physics-based deformable model used as a regularization and interpolation step allows us to perform non-rigid and robust registration between the pre and intraoperative images [10].

MORPHEO Team

6. New Results

6.1. A discrete 3D+t Laplacian framework for mesh animation processing

In this work we extend the discrete 3D Laplacian framework to mesh animations, represented as temporally coherent sequences of meshes (Figure 3). In order to let the user control the motion influence with respect to the geometry, we introduce a parameter for the time dimension. Our discrete 3D+t Laplace operator holds the same properties as the discrete 3D Laplacian, as soon as this parameter is non negative. We demonstrate the usefulness of this framework by extending Laplacian-based mesh editing and fairing techniques to mesh animations [15].

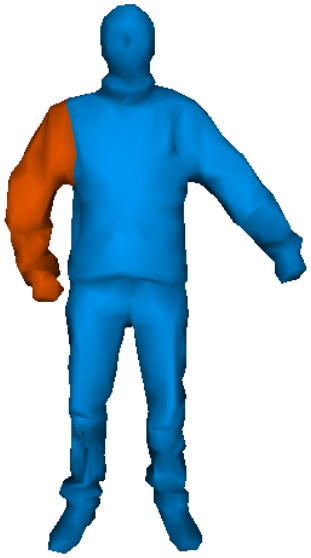


Figure 3. 3D+t Laplacian

6.2. Surface Flow

Recovering dense motion information is a fundamental intermediate step in the image processing chain upon which higher level applications can be built, such as tracking or segmentation. For that purpose, pixel observations in the image provide useful motion cues through temporal variations of the intensity function. We have studied the estimation of dense, instantaneous 3D motion fields over non-rigidly moving surface observed by multi-camera systems. The motivation arises from multi-camera applications that require motion information for arbitrary subjects, in order to perform tasks such as surface tracking or segmentation. To this aim, we have proposed a novel framework that allows to efficiently compute dense 3D displacement fields using low level visual cues and geometric constraints. The main contribution is a unified framework that combines flow constraints for small displacements with temporal feature constraints for large displacements

and fuses them over the surface using local rigidity constraints. The resulting linear optimization problem allows for variational solutions and fast implementations. Experiments conducted on synthetic and real data demonstrated the respective interests of flow and feature constraints as well as their efficiency to provide robust surface motion cues when combined.
As an extension of this work, we also studied the situation where a depth camera and one or more color cameras are available, a common situation with recent composite sensors such as the Kinect. In this case, geometric information from depth maps can be combined with intensity variations in color images in order to estimate smooth and dense 3D motion fields. We propose a unified framework for this purpose, that can handle both arbitrary large motions and sub-pixel displacements. The novelty with respect to existing scene flow approaches is that it takes advantage of the geometric information provided by the depth camera to define a surface domain over which photometric constraints can be consistently integrated in 3D. Experiments on real and synthetic data provide both qualitative and quantitative results that demonstrated the interest of the approach[12].

6.3. Progressive Shape Models

In this work we address the problem of recovering both the topology and the geometry of a deformable shape using temporal mesh sequences (Figure 4). The interest arises in multi-camera applications when unknown natural dynamic scenes are captured. While several approaches allow recovery of shape models from static scenes, few consider dynamic scenes with evolving topology and without prior knowledge. In this nonetheless generic situation, a single time observation is not necessarily enough to infer the correct topology of the observed shape and evidences must be accumulated over time in order to learn this topology and to enable temporally consistent modelling. This appears to be a new problem for which no formal solution exists. We have proposed a principled approach based on the assumption that the observed objects have a fixed topology. Under this assumption, the topology can be progressively learned during the capture of a dynamic scene evolutions. The approach has been successfully experimented on several standard 4D datasets and we believe that it paves the way to more general multi-view scene capture and analysis[8].

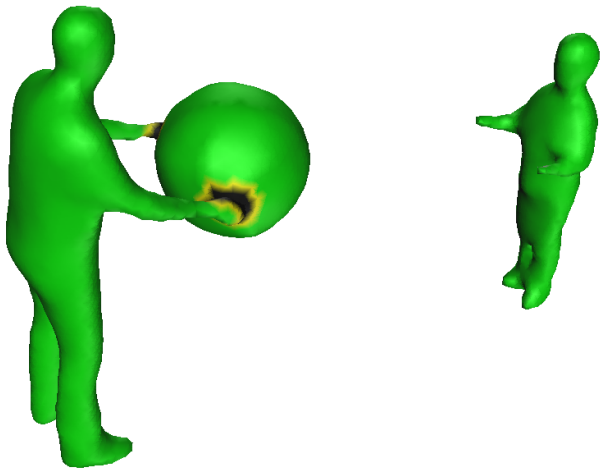


Figure 4. Progressive Shape Models : the balloon can be separated from humans

6.4. Principal Geodesic Dynamics

This work presents a new integration of a data-driven approach using dimension reduction and a physically-based simulation for real-time character animation (Figure 5). We exploit Lie group statistical analysis techniques (Principal Geodesic Analysis, PGA) to approximate the pose manifold of a motion capture sequence by a reduced set of pose geodesics. We integrate this kinematic parametrization into a physically-based animation approach of virtual characters, by using the PGA-reduced parametrization directly as generalized coordinates of a Lagrangian formulation of mechanics. In order to achieve real-time without sacrificing stability, we derive an explicit time integrator by approximating existing variational integrators. Finally, we test our approach in task-space motion control. By formulating both physical simulation and inverse kinematics time stepping schemes as two quadratic programs, we propose a features-based control algorithm that interpolates between the two metrics. This allows for an intuitive trade-off between realistic physical simulation and controllable kinematic manipulation[9].

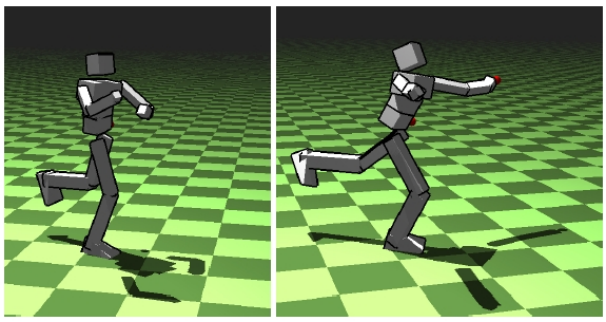


Figure 5. Principal Geodesic Dynamics : test of the balance controller

6.5. A Minimal Solution for Camera Calibration Using Independent Pairwise Correspondences

We have proposed a minimal algorithm for fully calibrating a camera from 11 independent pairwise point correspondences with two other calibrated cameras. Unlike previous approaches, our method neither requires triple correspondences, nor prior knowledge about the viewed scene. This algorithm can be used to insert or re-calibrate a new camera into an existing network, without having to interrupt operation. Its main strength comes from the fact that it is often difficult to find triple correspondences in a camera network. This makes the algorithm, for the specified use cases, probably the most suited calibration solution that does not require a calibration target, and hence can be performed without human interaction [10].

6.6. N-Tuple Color Segmentation for Multi-View Silhouette Extraction

We have presented a new method to extract multiple segmentations of an object viewed by multiple cameras, given only the camera calibration. This method relies on the n-tuple color model to express inter-view consistency when inferring in each view the foreground and background color models permitting the final segmentation. A color n-tuple is a set of pixel colors associated to the n projections of a 3D point. The first goal is set as finding the MAP estimate of background/foreground color models based on an arbitrary sample set of such n-tuples, such that samples are consistently classified, in a soft way, as "empty" if they project in the background of at least one view, or "occupied" if they project to foreground pixels in all views. An Expectation Maximization framework is then used to alternate between color models and soft classifications. In a final step, all views are segmented based on their attached color models. The approach is significantly

simpler and faster than previous multi-view segmentation methods, while providing results of equivalent or better quality. [6].

6.7. Cage-based Motion Recovery using Manifold Learning

We have proposed a flexible model-based approach for the recovery of parameterized motion from a sequence of 3D meshes without temporal coherence (Figure 6). Unlike previous model-based approaches using skeletons, we embed the deformation of a reference mesh template within a low polygonal representation of the mesh, namely the cage, using Green Coordinates. The advantage is a less constrained model that more robustly adapts to noisy observations while still providing structured motion information, as required by several applications. The cage is parameterized with a set of 3D features dedicated to the description of human morphology. This allows to formalize a novel representation of 3D meshed and articulated characters, the Oriented Quads Rigging (OQR). To regularize the tracking, the OQR space is subsequently constrained to plausible poses using manifold learning. Results are shown for sequences of meshes, with and without temporal coherence, obtained from multiple view videos preprocessed by visual hull. Motion recovery applications are illustrated with a motion transfer encoding and the extraction of trajectories of anatomical joints. Validation is performed on the HumanEva II database[7].

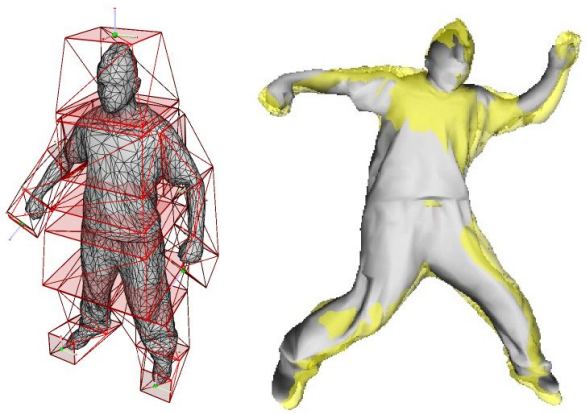


Figure 6. Cage-based Motion Recovery using Manifold Learning

6.8. Segmentation of temporal mesh sequences into rigidly moving components

This work considers the segmentation of meshes into rigid components given temporal sequences of deforming meshes (Figure 7). We have proposed a fully automatic approach that identifies model parts that consistently move rigidly over time. This approach can handle meshes independently reconstructed at each time instant. It allows therefore for sequences of meshes with varying connectivities as well as varying topology. It incrementally adapts, merges and splits segments along a sequence based on the coherence of motion information within each segment. In order to provide tools for the evaluation of the approach, we also introduce new criteria to quantify a mesh segmentation. Results on both synthetic and real data as well as comparisons are provided in the paper[3].

6.9. Keypoints and Local Descriptors of Scalar Functions on 2D Manifolds

This work addresses the problem of describing surfaces using local features and descriptors. While methods for the detection of interest points in images and their description based on local image features are very well

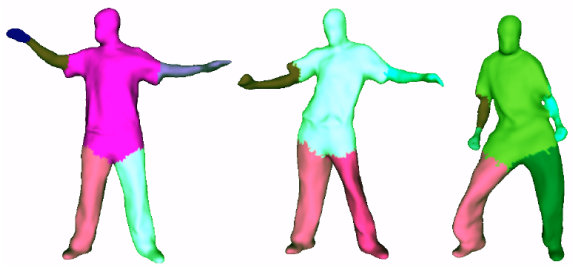


Figure 7. Segmentation of temporal mesh sequences into rigidly moving components

understood, their extension to discrete manifolds has not been well investigated. We provide a methodological framework for analyzing real-valued functions defined over a 2D manifold, embedded in the 3D Euclidean space, e.g., photometric information, local curvature, etc. Our work is motivated by recent advancements in multiple-camera reconstruction and image-based rendering of 3D objects: there is a growing need for describing object surfaces, matching two surfaces, or tracking them over time. Considering polygonal meshes, we propose a new methodological framework for the scale-space representations of scalar functions defined over such meshes. We propose a local feature detector (MeshDOG) and region descriptor (MeshHOG). Unlike the standard image features, the proposed surface features capture both the local geometry of the underlying manifold and the scale-space differential properties of the real-valued function itself. We provide a thorough experimental evaluation. The repeatability of the feature detector and the robustness of feature descriptor are tested, by applying a large number of deformations to the manifold or to the scalar function[4].

PERCEPTION Team

6. New Results

6.1. 3D shape analysis and registration

We address the problem of 3D shape registration and we propose a novel technique based on spectral graph theory and probabilistic matching. Recent advancement in shape acquisition technology has led to the capture of large amounts of 3D data. Existing real-time multi-camera 3D acquisition methods provide a frame-wise reliable visual-hull or mesh representations for real 3D animation sequences The task of 3D shape analysis involves tracking, recognition, registration, etc. Analyzing 3D data in a single framework is still a challenging task considering the large variability of the data gathered with different acquisition devices. 3D shape registration is one such challenging shape analysis task. The main contribution of this chapter is to extend the spectral graph matching methods to very large graphs by combining spectral graph matching with Laplacian embedding. Since the embedded representation of a graph is obtained by dimensionality reduction we claim that the existing spectral-based methods are not easily applicable. We discuss solutions for the exact and inexact graph isomorphism problems and recall the main spectral properties of the combinatorial graph Laplacian; We provide a novel analysis of the commute-time embedding that allows us to interpret the latter in terms of the PCA of a graph, and to select the appropriate dimension of the associated embedded metric space; We derive a unit hyper-sphere normalization for the commute-time embedding that allows us to register two shapes with different samplings; We propose a novel method to find the eigenvalue-eigenvector ordering and the eigenvector sign using the eigensignature (histogram) which is invariant to the isometric shape deformations and fits well in the spectral graph matching framework, and we present a probabilistic shape matching formulation using an expectation maximization point registration algorithm which alternates between aligning the eigenbases and finding a vertex-to-vertex assignment. See [22], [34], [19] for more details.

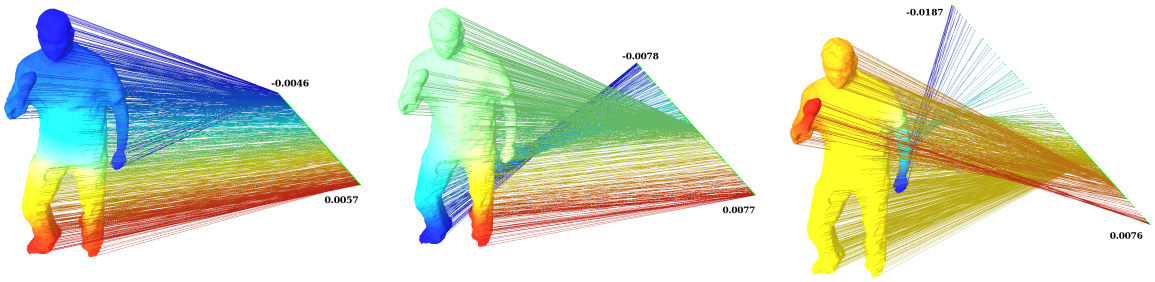


Figure 5. This is an illustration of the concept of the PCA of a shape embedding. The shapes's vertices are projected onto the second, third and fourth eigenvectors of the Laplacian matrix. These eigenvectors can be viewed as the principal directions of the shape (see [34] for more details).

6.2. High-resolution depth maps based on TOF-stereo fusion

The combination of range sensors with color cameras can be very useful for a wide range of applications, e.g., robot navigation, semantic perception, manipulation, and telepresence. Several methods of combining range- and color-data have been investigated and successfully used in various robotic applications. Most of these systems suffer from the problems of noise in the range-data and resolution mismatch between the range

sensor and the color cameras, since the resolution of current range sensors is much less than the resolution of color cameras. High-resolution depth maps can be obtained using stereo matching, but this often fails to construct accurate depth maps of weakly/repetitively textured scenes, or if the scene exhibits complex self-occlusions. Range sensors provide coarse depth information regardless of presence/absence of texture. The use of a calibrated system, composed of a time-of-flight (TOF) camera and of a stereoscopic camera pair, allows data fusion thus overcoming the weaknesses of both individual sensors. We propose a novel TOF-stereo fusion method based on an efficient seed-growing algorithm which uses the TOF data projected onto the stereo image pair as an initial set of correspondences. These initial "seeds" are then propagated based on a Bayesian model which combines an image similarity score with rough depth priors computed from the low-resolution range data. The overall result is a dense and accurate depth map at the resolution of the color cameras at hand. We show that the proposed algorithm outperforms 2D image-based stereo algorithms and that the results are of higher resolution than off-the-shelf color-range sensors, e.g., Kinect. Moreover, the algorithm potentially exhibits real-time performance on a single CPU. See [27], [33] for more details.

6.3. Simultaneous sound-source separation and localization

Human-robot communication is often faced with the difficult problem of interpreting ambiguous auditory data. For example, the acoustic signals perceived by a humanoid with its on-board microphones contain a mix of sounds such as speech, music, electronic devices, all in the presence of attenuation and reverberations. We proposed a novel method, based on a generative probabilistic model and on active binaural hearing, allowing a robot to robustly perform sound-source separation and localization. We show how interaural spectral cues can be used within a constrained mixture model specifically designed to capture the richness of the data gathered with two microphones mounted onto a human-like artificial head. We describe in detail a novel expectation-maximization (EM) algorithm that alternates between separation and localization, we analyse its initialization, speed of convergence and complexity, and we assess its performance with both simulated and real data. Subsequently, we studied the *binaural manifold*, i.e., the low-dimensional space of sound-source locations embedded in the high-dimensional space of perceived interaural spectral features, and we provided a method for mapping interaural cues onto source locations. See [25], [24], [26]

6.4. Sound localization and recognition with a humanoid robot

We addressed the problem of localizing recognizing everyday sound events in indoor environments with a consumer robot. For localization, we use the four microphones that are embedded into the robot's head. We developed a novel method that uses four non-coplanar microphones and that guarantees that for each set of pairwise TDOA (time difference of arrival) there is a unique 3D source location. For recognition, sounds are represented in the spectrotemporal domain using the stabilized auditory image (SAI) representation. The SAI is well suited for representing pulse-resonance sounds and has the interesting property of mapping a time-varying signal into a fixed-dimension feature vector space. This allows us to map the sound recognition problem into a supervised classification problem and to adopt a variety of classifications schemes. We developed a complete system that takes as input a continuous signal, splits it into significant isolated sounds and noise, and classifies the isolated sounds using a catalogue of learned sound-event classes. The method is validated with a large set of audio data recorded with a humanoid robot in a typical home environment. Extended experiments showed that the proposed method achieves state-of-the-art recognition scores with a twelve-class problem, while requiring extremely limited memory space and moderate computing power. A first real-time embedded implementation in a consumer robot show its ability to work in real conditions. See [23], [28] for more details.

6.5. Audiovisual fusion based on a mixture model

The problem of multimodal clustering arises whenever the data are gathered with several physically different sensors. Observations from different modalities are not necessarily aligned in the sense there there is no obvious way to associate or to compare them in some common space. A solution may consist in considering multiple clustering tasks independently for each modality. The main difficulty with such an approach is to guarantee that the unimodal clusterings are mutually consistent. In this paper we show that multimodal

clustering can be addressed within a novel framework, namely conjugate mixture models. These models exploit the explicit transformations that are often available between an unobserved parameter space (objects) and each one of the observation spaces (sensors). We formulate the problem as a likelihood maximization task and we derive the associated expectation-maximization algorithm. The algorithm and its variants are tested and evaluated within the task of 3D localization of several speakers using both auditory and visual data. See [36], [30], [29] for more details.

PRIMA Project-Team

6. New Results

6.1. Social behaviors recognition

Participants: Wafa Benkaouar, Claudine Combe, Dominique Vaufreydaz [correspondant].

Figure 5. On the left image, one can see the telemeter range in red, the foot detection (blue spot) and the angle view from the Kinect (in green). the middle and right image represent RGB camera en depth view from the Kinect.

Recognition of social behaviors is an unconscious innate cognitive process vital to human communication. This skill enables anticipation and increases interactive exchanges quality between humans. Among social behaviors, engagement is the expression of intention for interaction. During engagement phase, many non-verbal signals are used to communicate this intention to the partner, e.g. posture, gaze, spatial information, gestures, vocal cues. Within the context of frail or elderly people at home, companion robots must also be able to detect the engagement of humans in order to adapt their responses during interaction with humans to increase their acceptability.

Classical approaches in the domain are dealing with spatial information. Our hypothesis was that relative spatial information of people and robot are not discriminative in a home-like environment [15]. Our approach integrates multimodal features gathered using a robot companion equipped with a Kinect from Microsoft (see figure 5). Confronted to a robot centered dataset for multimodal social signal processing recorded in a home-like environment, the evaluation highlights its robustness and validates use of such technique in real environment (50% of error reduction). Our experimentations also confirm results from cognitive science domain [61].

6.2. Live monitoring and correction of 3DTV broadcasts

Participants: Pierre Arquier, Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora, Matthieu Volat.

6.2.1. 3D broadcast monitoring and correction:

One of the achievements of the 3DLive FUI project was the transfer of real-time 3D video monitoring and correction algorithms to the Binocle company, and their integration into the TaggerLive product, which was used during several 3DTV broadcasts between 2010 and 2012 for live monitoring and correction of stereoscopic video. The algorithms that were developed within the PRIMA team and transferred into the TaggerLive are:

Multiscale view-invariant feature detection and matching on the GPU.

Computation of a temporally smooth and robust correction (or rectification) to remove the vertical disparity in the stereoscopic video while keeping the image aspect.

Real-time monitoring of the “depth budget”, or the histogram of the horizontal disparity;

Live alerts when stereoscopic production rules are broken, such as when the disparities are too large, or when there is a stereoscopic window violation.

Real-time implementation of a state-of-the-art dense stereo matching method on the GPU.

6.2.2. 3D content adaptation:

3D shape perception in a stereoscopic movie depends on several depth cues, including stereopsis. For a given content, the depth perceived from stereopsis highly depends on the camera setup as well as on the display size and distance. This can lead to disturbing depth distortions such as the cardboard effect or the puppet theater effect. As more and more stereoscopic 3D content is produced in 3D (feature movies, documentaries, sports broadcasts), a key point is to get the same 3D experience on any display. For this purpose, perceived depth distortions can be resolved by performing view synthesis. We have proposed [19] a real time implementation of a stereoscopic player based on the open-source software Bino, which is able to adapt a stereoscopic movie to any display, based on user-provided camera and display parameters.

6.2.3. Focus mismatch detection:

Live-action stereoscopic content production requires a stereo rig with two cameras precisely matched and aligned. While most deviations from this perfect setup can be corrected either live or in post-production, a difference in the focus distance or focus range between the two cameras will lead to unrecoverable degradations of the stereoscopic footage. We have developed a method [18] to detect focus mismatch between views of a stereoscopic pair in four steps. First, we compute a dense disparity map. Then, we use a measure to compare focus in both images. After this, we use robust statistics to find which images zones have a different focus. Finally, to give useful feedback, we show the results on the original images and give hints on how to solve the focus mismatch.

6.3. Simultaneous localization and mapping (SLAM)

Participants: James Crowley, Marion Decrouez, Frédéric Devernay.

Localisation, place recognition, object recognition. Live processing of a video sequence taken from a single camera enables to model an a priori unknown 3D scene. Metrical SLAM (Simultaneous Localization and Mapping) algorithms track the camera pose while reconstructing a sparse map of the visual features of the 3D environment. Such approaches provide the geometrical foundation for many augmented reality applications in which informations and virtual objects are superimposed on live images captured by a camera. Improving such systems will enable in the future precise industrial applications such as guided-maintenance or guided-assembly in wide installations.

A problem with current methods is the assumption that the environment is static. Indoor environments such as supermarket ailes and factory floors may contain numerous objects that are likely to be moved, disrupting a localization and mapping system. We explore methods for automatic detection and modeling of such objects. We define the scene as a static structure that may contain moving objects and objects are defined as a set of visual features that share a common motion compared to the static structure [39]. Using several explorations of a camera in the same scene, we detect and model moved objects while reconstructing the environment. Experiments highlignt the performance of the method in a real case of localization in an unknown indoor environment.

6.4. Post-production tools for 3-D Cinema

Participants: Laurent Boiron, Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora.

Over the past 6 years, we have been developing 3D movie processing techniques which have been used for the production and post-production of 3D movies (mainly feature-length movies, documentaries and commercials). These include image alignment, view interpolation, depth map computation, etc. These algorithms were developed as C++ libraries, and can be executed using standalone tools. Since the movie post-production workflow relies mainly on standard tools for compositing, color grading, etc., and these tools can be extended by plugin mechanisms, we integrated our post-production algorithms into such a tool, namely Nuke by The Foundry.

We also developed a new method for stereoscopic video cut and paste. Video cut-and-paste consists in semi-interactively segmenting a video object from a video stream, and pasting the segmented video object in another video. The object segmentation is done using a small number of strokes made on a few frames of the video, and can be corrected interactively. Existing methods only worked on monoscopic videos, and extending it to stereoscopic videos required solving important challenges:

The video object must not only remain consistent over time, but also between the left and right views.

The video object may be partially occluded in one or both views.

The camera setup may be different between the first and the second video, causing depth distortion or different depth effects.

We solved the first two challenges by adding left-right stereo consistency based on dense stereo matching, as well as temporal consistency based on optical flow, in an optimization framework based on graph cuts. The user interface was also taken into consideration in the algorithm: any correction of the results (i.e. new strokes on an image) will only propagate forward in time.

6.5. Scene flow computation from RGBZ data

Participants: Frédéric Devernay [correspondant], Julian Quiroga.

The scene flow describes the motion of each 3D point between two times steps. With the arrival of new depth sensors, as the Microsoft Kinect, it is now possible to compute scene flow with a single camera, with promising repercussion in a wide range of computer vision scenarios. We proposed [22] a novel method to compute scene flow by tracking in a Lucas-Kanade framework. Scene flow is estimated using a pair of aligned intensity and depth images, but rather than computing a dense scene flow as in most previous methods, we get a set of 3D motion fields by tracking surface patches. Assuming a 3D local rigidity of the scene, we propose a rigid translation flow model that allows to solve directly for the scene flow by constraining the 3D motion field both in intensity and depth data. In our experimentation we achieve very encouraging results. Since this approach solves simultaneously for the 2D tracking and for the scene flow, it can be used for action recognition in existing 2D tracking based methods or to define scene flow descriptors.

6.6. Knit your Ideas Into Smart Spaces

Participants: Joelle Coutaz, Alexandre Demeure [correspondant], Emeric Fontaine.

We developped KISS (Knit your Ideas Into Smart Spaces), an end-user development system for the home. KISS enables users to program their home with sentences expressed in a pseudo-natural language. Programs can be tested either with the virtual home or in the real home. We led an evaluation that shows that users are able to program a real-life scenario. This work is described in the phd manuscript of Emeric Fontaine [46]. An experimental evaluation shows that KISS can be used to programm a real life scenario.

Particpants encountered some difficulties related to the restricted vocabulary used for the experiment. Some difficulties also occured relative to the understanding of "progressive verbs". To overcome these problems, we envision a system for co-constructing vocabulary with the system, which may lead to the definition of multiple language for communicating with the system.

6.7. Attention-Based Navigation

Participants: Thomas Fisher, Thierry Fraichard [correspondant], Patrick Reignier.

Assistant robots and robot companions are designed to share the human living space, to navigate among and interact with human beings. From the mobility point of view, roboticists have recently striven to develop navigation scheme geared towards achieving so-called "socially acceptable motions". To that end, various concepts borrowed from environmental psychology and anthropology have been used, the "personal space" concept from Proxemics being perhaps the most widely used.

The purpose of our work here is to further the research in this area by taking into account other factors such as human activities, interaction configurations and intentions. An attentional model derived from cognitive psychology is used to dynamically determine the "focus of attention" of the persons involved in a given task. Depending on the task at hand, the robot uses the attention information in order to decide its future course of action so as, for instance, to attract one person's attention or, on the contrary, to minimize the disturbance caused.

SIROCCO Project-Team

6. New Results

6.1. Analysis and modeling for compact representation and navigation

3D modelling, multi-view plus depth videos, Layered depth images (LDI), 2D and 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

6.1.1. Computational modelling of visual attention

Participants: Josselin Gautier, Olivier Le Meur, Zhi Liu.

6.1.1.1. Time-dependent saliency map

The study related to the deployment of visual attention in 2D and 3D has been completed in 2012. The purpose of this study was to investigate whether or not there is a difference between eye movements recorded while observers viewed natural images in 2D and 3D conditions. Results show that visual exploration in depth layer detection task is affected by the binocular disparity. In particular, participants tend to look first at closer areas just after the stimuli onset with the introduction of disparity, and then direct their gaze to more widespread locations. Based on these conclusions, a computational model of visual attention taking into account the temporal dimension has been designed. An Expectation-Maximisation (EM) algorithm has been used to infer the weight of different visual features (saliency, depth, center bias) over time. Results have been published in the journal Cognitive Computation.

A new study on a similar subject has started during the summer 2012. The purpose is again to investigate the influence of binocular disparity, scene complexity on visual scanpaths obtained in 2D and 3D viewing conditions. The main differences with the previous study are twofold. First, a new database of content has been designed. All parameters such as the amount of disparity are accurately mastered. Second is about the context of the study which deals with quality assessment of 3D video content.

6.1.1.2. Salient object detection

In 2012, Dr. Liu, who has joined the team in August for 2 years has started a study dealing with salient object detection. The goal is to extract automatically the most interesting object in an image or video sequence. The proposed approach is based on low-level visual features and extensively used a superpixel method. Starting from the superpixel representation of an image, the saliency measure of each superpixel is evaluated based on its global uniqueness and local contrasts with other superpixels. A saliency-directed region merging algorithm with a dynamic scale control scheme is then exploited to generate more meaningful regions. The region merging process is recorded using a Binary Partition Tree (BPT), in which each leaf node represents each superpixel and each non-leaf node represents each generated region during the region merging process. Finally, a node selection algorithm based on saliency density difference is used to select suitable nodes from BPT to form the salient object detection result. First experimental results on a public dataset (MSRA) are promising and demonstrate the effectiveness of the proposed approach.

6.1.2. Similarity metrics for image processing

Participants: Mounira Ebdelli, Christine Guillemot, Olivier Le Meur, Raul Martinez Noriega, Aline Roumy.

Several image processing problems addressed by the team (inpainting, loss concealment, super-resolution, denoising) require having patch objective similarity metrics as close as possible to ground truth visual similarity. The derivation of such metrics has been investigated along several directions. First, a performance analysis of the most used fidelity metrics (SSD, SSIM, two SSD-weighted Battacharya metrics) has been carried out to assess the perceptual similarities between patches. A statistical analysis of subjective tests has shown that some of these metrics (the SSD-weighted Battacharya) are more suitable than others to respect human decisions in terms of patch similarities. This conclusion has been confirmed with the results of Non Local means (NL-means) denoising algorithm which are highly sensitive to the used similarity metrics. The value of each pixel p in the blurred image is updated using a weighted average of the collocated pixels values in the most similar patches to the block centered on p. We show that SSD, which is the most used similarity metric, is not necessary the best correlated with the perceptual criteria.

Greedy algorithms for inpainting are based on the assumption of self-similarity within an image. A patch located on the boundary of the hole to be filled in, contains a known part and an unknown part. The known part is used to select other (completely known) patches and called exemplars. Then, these exemplars are used to reconstruct the unknown part of the patch being processed. Such an approach faces two main problems, decision of filling-in order and selection of good exemplars from which the missing region is synthesized. In [29], we proposed an algorithm that tackles these problems with improvements in the preservation of linear edges, and reduction of error propagation compared to well-known algorithms from the literature. Our improvement in the filling-in order is based on a combination of priority terms, previously defined, that better encourages the early synthesis of linear structures. The second contribution helps reducing the error propagation thanks to a better detection of outliers from the candidate patches carried. This is obtained with a new metric based on the Hellinger distance between the patches that incorporates the whole information of the candidate patches.

6.1.3. Epitome-based image representation

Participants: Safa Cherigui, Christine Guillemot.

This work is carried out in collaboration with Technicolor (D. Thoreau, Ph. Guillotel, P. Perez) and aims at designing a compresion algorithm based on the concept of epitomes. An epitome is a condensed representation of an image (or a video) signal containing the essence of the textural properties of this image. Different forms of epitomes have been proposed in the literature, such as a patch-based probability model learned either from still image patches or from space-time texture cubes taken from the input video. These probability models together with appropriate inference algorithms, are useful for content analysis inpainting or super-resolution. Another family of approaches makes use of computer vision techniques, like the KLT tracking algorithm, in order to recover self similarities within and across images. In parallel, another type of approach consists in extracting epitome-like signatures from images using sparse coding and dictionary learning.

The method developed aims at tracking self-similarities within an image using a block matching (BM) algorithm. The epitome is constructed from disjoint pieces of texture ("epitome charts") taken from the original image and a transform map which contains translational parameters. Those parameters keep track of the correspondences between each block of the input image and a block of the epitome. An Intra image compression scheme based on the epitome has been developed showing a rate saving of up to 12% on some images, including the rate cost of the epitome texture and of the transform map. The entire image can be reconstructed from the epitome texture with the help of the transform map.

6.2. Rendering, inpainting and super-resolution

image-based rendering, inpainting, view synthesis, super-resolution

6.2.1. Joint projection/filling method for virtual view synthesis

Participants: Christine Guillemot, Fabien Racapé.

This study is carried out in collaboration with INSA/IETR (Luce Morin). Associated with a view synthesis method, a multi-view plus depth video allows the generation of virtual views of the scene from any viewpoint. State-of-the-art synthesizers use Depth-Image-Based Rendering (DIBR) techniques based on warping equations, which project a reference view onto a virtual viewpoint. In classical DIBR schemes, the rendering proceeds in several distinct steps, each one designed to solve a specific problem. First, the depth map is warped onto the virtual viewpoint and filtered with a median filter. The filtered depth map is then used in a backward warping of the virtual view (as illustrated in FIg.1). The resulting depth map is inpainted, to fill in disocclusion areas. Finally, this complete depth map is used by a depth-aided inpainting algorithm to fill in disocclusions in the color map. However, all these steps are inter-dependent, and errors introduced by each one are amplified by the following one, producing annoying artifacts, as shown in Fig. 2 -(c).

The proposed Joint Projection Filling (JPF) method performs forward projection, using connectivity information to fill in disocclusions in a single step. Applied on the depth map warping, JPF enables a depth-aided inpainting of color disocclusions after backward projection, as shown in Fig. 1 . Fig. 2 -(e) presents a resulting synthesis which contains less artifacts.

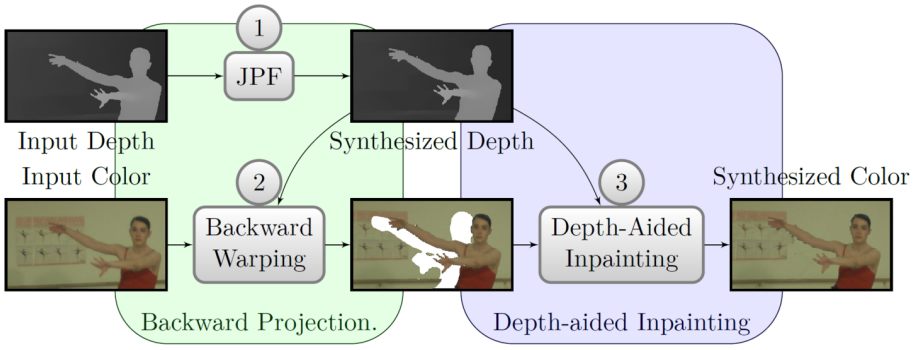


Figure 1. Virtual view generation chain, based on Joint Filling Projection. The depth map is jointly warped and inpainted. Depth-aided inpainting can be then used on disoccluded areas.

In the context of multi-view plus depth video coding (3D-HEVC standardization), inter view coding tools are added in the vein of temporal inter frame coding. We have tested our method as a projection tool for View Synthesis Prediction (VSP). However, the 3D-HEVC common test conditions, limited to rectified views as input, restrict the possible gains induced by efficient projection tools. Moreover, JPF outperforms other methods in synthesizing disoccluded areas with a good visual quality where VSP tools are not selected by MSE-based decision. JPF remains an efficient tool for extrapolating multi-view plus depth content with a minimum of artifacts on disoccluded areas.

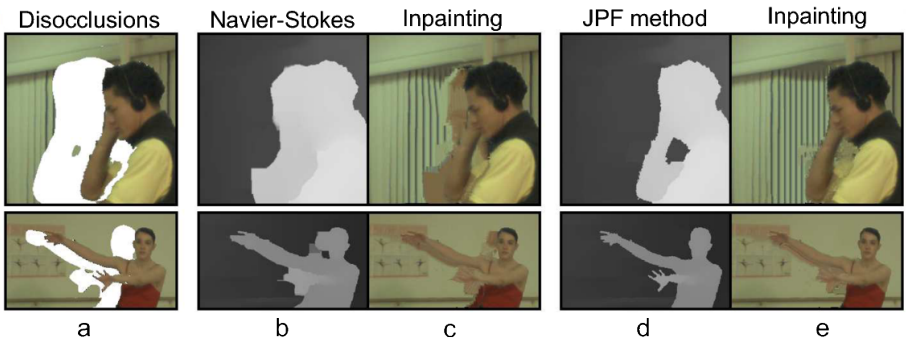


Figure 2. Disocclusion filling. (a) warped image before inpainting. Depth map inpainting: Navier-Stokes (b), JPF (d). Resulting depth-aided inpainting: Navier-Stokes (c), JPF (e).

6.2.2. Image inpainting using neighbor embedding and super-resolution

Participants: Mounira Ebdelli, Christine Guillemot, Olivier Le Meur.

Inpainting methods play an important role in a wide range of applications. Removing text and advertisements (such as logos), removing undesired objects, noise reduction and image reconstruction from incomplete data are the key applications of inpainting methods. Algorithms can be classified into two categories: PDE (Partial Derivative Equation)-based schemes and examplar-based schemes. The former uses diffusion schemes in order

to propagate structures in a given direction. Their drawback is the introduction of blur due to diffusion. The latter relies on the sampling and the copying of texture from the known parts of the picture.

Image inpainting is a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the patch to be filled in. Novel inpainting methods have been developed in the team along complementary directions: 1/- considering new priority functions exploiting the structure within the patch for defining the patch processing order; 2/- investigating various neighbor embedding techniques for estimating the unknown pixels; 3/- considering a coarse to fine multi-resolution approach where a low resolution version of the input image is first inpainted, this first step being followed by a a super-resolution based enhancement of the image.

Priority functions: Different priority functions using structure tensors and edge based information have been considered and their advantage over classical functions projecting isophote directions on the normal to the front line has been demonstrated.

Neighbor-embedding based inpainting: Examplar-based inpainting algorithms using various neighbor embedding techniques (LLE, LLE-LDNR, NMF with various solvers) have been developed. The methods have been shown to enhance the quality of inpainted images when compared to classical examplar-based solutions using simple template matching techniques to estimate the missing pixels, or similarity weights (NLM) (see Fig. 3).

Figure 3. Inpainting results: Original image; Mask of the image to be inpainted; Inpainting results with examplar-based inpainting using similarity weights; Inpainting results with neighbor embedding (LLE-LDNR).

Super-resolution aided inpainting: A novel super-resolution aided inpainting approach has been introduced which consists in first inpainting a coarse version of the input image and then in a second step, using a hierarchical super-resolution algorithm, to recover the native resolution [28]. However, to be less sensitive to the setting of the inpainting methods, the low-resolution input picture is inpainted several times with different settings. Results are efficiently combined with a loopy belief propagation. A super-resolution algorithm is then applied to recover the details. Experimental results in a context of image editing, texture synthesis and 3D view synthesis demonstrate the effectiveness of the proposed method. Fig.4 show texture synthesis results obtained with this approach.

Figure 4. Texture synthesis results obtained with super-resolution aided inpainting.

6.3. Representation and compression of large volumes of visual data

sparse representations, data dimensionality reduction, compression, scalability, perceptual coding, rate-distortion theory

6.3.1. Multi-view plus depth video compression

Participants: Christine Guillemot, Thomas Guionnet, Laurent Guillo, Fabien Racapé.

Multi-view plus depth video content represent very large volumes of input data wich need to be compressed for storage and tranmission to the rendering device. The huge amount of data contained in multi-view sequences indeed motivates the design of efficient representation and compression algorithms. In collaboration with INSA/IETR (Luce Morin), we have studied layered depth image (LDI) and layered depth video (LDV) representations as a possible compact representation format of multi-view video plus depth data. LDI give compact represention of 3D objects, which can be efficiently used for photo-realistic image-based rendering (IBR) of different scene viewpoints, even with complex scene geometry. The LDI extends the 2D+Z representation, but instead of representing the scene with an array of depth pixels (pixel color with associated depth values), each position in the array may store several depth pixels, organised into layers. A novel object-based LDI representation which is more tolerant to compression artifacts, as well as being compatible with fast mesh-based rendering techniques has been developped.

The team has also studied motion vector prediction in the context of HEVC-compatible Multi-view plus depth (MVD) video compression. The HEVC compatible MVD compression solution implements a 6 candidate vector list for merge and skip modes. As part of the 3D video encoding, an inter–view motion vector predictor is added at the first position of this list. Our works show that this new list can be improved in optimizing the order of the candidates and in adding two more relevant candidates. When a merge or a skip mode is selected, a merge index is written in the bitstream. This index is first binarized using a unary code, then encoded with the CABAC. A CABAC context is dedicated to the first bin of the unary coded index while the remaining bins are considered as equiprobable. This strategy is efficient as long as the candidate list is ordered

by decreasing index occurrence probability. However, this is not always the case when the inter-view motion vector predictor is added. To dynamically determine which candidate is the most probable, a merge index histogram is computed on the fly at the encoder and decoder side. Thus a conversion table can be calculated. It allows deriving the merge index to encode given the actual index in the list, and conversely, the actual index in the list given a decoded index. When using dynamic merge index, index re-allocation can happen at any time. Statistics of the first bin, which is encoded with CABAC, are modified. That is why a set of 6, one for each possible permutation of indexes, CABAC contexts dedicated to the first bin is defined. A bit rate gain of 0.1% for side views is obtained with no added complexity. These results are improved and reach 0.4% when additional CABAC contexts are used to take into account also the first three bins.

Candidates added by default in the merge list are not always the most relevant. As part of 3D video encoding using multiple rectified views, having a fine horizontal adjustment might be meaningful for efficient disparity compensated prediction. Therefore, we have proposed to replace some candidates in the merge list with candidates pointing to the base view and shifted by the horizontal offsets +4 and -4. To do so, the merge list is scanned to get among the first four candidates the first disparity compensated candidate. Once this vector found, the +4 and -4 offsets are added to its horizontal component and the two resulting vectors are inserted in the list two positions further if there is still room just after otherwise. With this improvement, a bit rate gain of 0.3% for side views is obtained with no added complexity.

6.3.2. Diffusion-based depth maps coding

Participants: Josselin Gautier, Olivier Le Meur.

A novel approach to compress depth map has been developed [26]. The proposed method exploits the intrinsic depth maps properties. Depth images indeed represent the scene surface and are characterized by areas of smoothly varying grey levels separated by sharp edges at the position of object boundaries. Preserving these characteristics is important to enable high quality view rendering at the receiver side. The proposed algorithm proceeds in three steps: the edges at object boundaries are first detected using a Sobel operator. The positions of the edges are encoded using the JBIG algorithm. The luminance values of the pixels along the edges are then encoded using an optimized path encoder. The decoder runs a fast diffusion-based inpainting algorithm which fills in the unknown pixels within the objects by starting from their boundaries.

6.3.3. Neighbor embedding for image prediction

Participants: Safa Cherigui, Christine Guillemot.

The problem of texture prediction can be regarded as a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the block to be predicted. We have in 2010 and 2011 developed texture prediction methods as well as inpainting algorithms based on neighbor embedding techniques which come from the area of data dimensionality reduction [18], [31], [27]. The methods which we have more particularly considered are Locally Linear Embedding (LLE), LLE with Low-dimensional neighborhood representation (LDNR), and Non-negative Matrix Factorization (NMF) using various solvers.

The first step in the developed methods consists in searching, within the known part of the image, for the K nearest (KNN) patches to the set of known samples in the neighborhood of the block to be predicted (or of samples to be estimated in the context of inpainting). In a prediction (compression) context, in order for the decoder to proceed similarly, the K nearest neighbors are found by computing distances between the known pixels in a causal neighborhood (called template) of the input block and the co-located pixels in candidate patches taken from a causal window. Similarly, the weights used for the linear approximation are computed in order to best approximate the template pixels. Although efficient, these methods suffer from limitations when the template and the block to be predicted are not correlated, e.g. in non homogenous texture areas. To cope with these limitations, we have developed new image prediction methods based on neighbor embedding techniques in which the K-NN search is done in two steps and aided, at the decoder, by a block correspondence map, hence the name Map-Aided Neighbor Embedding (MANE) method. Another optimized variant of this approach, called oMANE method, has also been introduced. The resulting prediction methods

are shown to bring significant Rate-Distortion (RD) performance improvements when compared to H.264 Intra prediction modes (up to 44.75%) [13]. Figure 5 illustrates the prediction quality obtained with different neighbor embedding methods, as well as the encoder selection rate of the oMANE-based prediction mode. This method has been presented at the IEEE International ICIP conference and the paper has been selected among the 11 finalists (out of 500 student papers) for a best student paper award.

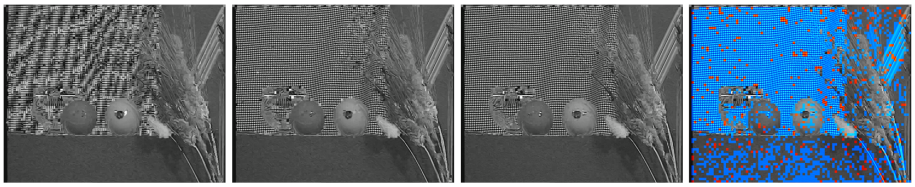


Figure 5. Spatial prediction for "Snook" with modes dynamically chosen according to a RD criterion with (a) H.264 Intra modes (High Profile), (b) LLE-based prediction, (c) Hybrid LLE-oMANE prediction and (d) selection rate of the two modes: LLE (red) and oMANE (blue).

6.3.4. Generalized lifting for video compression

Participants: Christine Guillemot, Bihong Huang.

This research activity is carried out in collaboration with Orange labs (Felix Henry) and UPC (Philippe Salembier) in Barcelona. The objective is to design new algorithmic tools for efficient loosless and lossy compression using generalized lifting concepts. The generalized lifting is a framework which permits the creation of nonlinear and signal probability density function (pdf) dependent and adaptive transforms. The use of such adaptive transforms for efficient coding of different HEVC syntax elements is under study.

6.3.5. Dictionary learning methods for sparse coding of satellite images

Participants: Jeremy Aghaei Mazaheri, Christine Guillemot, Claude Labit.

In the context of the national partnership Inria-Astrium, we explore novel methods to encode sequences of satellite images with a high degree of restitution quality and with respect to usual constraints in the satellite images on-board codecs. In this study, a geostationary satellite is used for surveillance and takes sequences of images. Then these pictures are stabilized and have to be compressed on-board before being sent to earth. Each picture has a high resolution and so the rate without compression is very high (about 70 Gbits/sec) and the goal is to achieve a rate after compression of 600 Mbits/sec, that is a compression ratio more than 100. On earth, the pictures are decompressed with a high necessity of reconstruction quality, especially for moving areas, and visualized by photo-interpreters. That is why the compression algorithm requires here a deeper study.The first stage of this study is to develop dictionary learning methods for sparse representations and coding of the images. These representations are commonly used for denoising and more rarely for image compression.

Sparse representation of a signal consists in representing a signal $y \in \Re^n$ as a linear combination of columns, known as atoms, from a dictionary matrix. The dictionary $D \in \Re^{n \times K}$ is generally overcomplete and contains K atoms. The approximation of the signal can thus be written $y \approx Dx$ and is sparse because a small number of atoms of D are used in the representation, meaning that the vector x has only a few non-zero coefficients. The choice of the dictionary is important for the representation. A predetermined transform matrix, as overcomplete wavelets or DCT, can be chosen. Another option is to learn the dictionary from training signals to get a well adapted dictionary to the given set of training data. Previous studies demonstrated that dictionaries have the potential to outperform the predetermined ones. Various advanced dictionary learning schemes have been

proposed in the literature, so that the dictionary used is well suited to the data at hand. The popular dictionary learning algorithms include the K-SVD, the Method of Optimal Directions (MOD), Sparse Orthonormal Transforms (SOT), and (Generalized) Principle Component Analysis (PCA).

Recently, the idea of giving relations between atoms of a dictionary appeared with tree-structured dictionaries. Hierarchical sparse coding uses this idea by organizing the atoms of the dictionary as a tree where each node corresponds to an atom. The atoms used for a signal representation are selected among a branch of the tree. The learning algorithm is an iteration of two steps: hierachical sparse coding using proximal methods and update of the entire dictionary. Even if it gives good results for denoising, the fact to consider the tree as a single dictionary makes it, in its current state, not well adapted to efficiently code the indices of the atoms to select when the dictionary becomes large. We introduce in this study a new method to learn a tree-structured dictionary offering good properties to code the indices of the selected atoms and to efficiently realize sparse coding. Besides, it is scalable in the sense that it can be used, once learned, for several sparsity constraints. We show experimentally that, for a high sparsity, this novel approach offers better rate-distortion performances than state-of-the-art "flat" dictionaries learned by K-SVD or Sparse K-SVD, or than the predetermined overcomplete DCT dictionary. We recently developped a new sparse coding method adapted to this tree-structure to improve the results. Our dictionary learning method associated with this sparse coding method is also compared to other methods previously introduced in the recent litterature such as TSITD (Tree-Structured Iteration-Tuned Dictionary) or BITD (Basic Iteration-Tuned Dictionary) algorithms.

6.4. Distributed processing and robust communication

information theory, stochastic modelling, robust detection, maximum likelihood estimation, generalized likelihood ratio test, error and erasure resilient coding and decoding, multiple description coding, Slepian-Wolf coding, Wyner-Ziv coding, information theory, MAC channels

6.4.1. Loss concealment based on video inpainting

Participants: Mounira Ebdelli, Christine Guillemot, Ronan Le Boulch, Olivier Le Meur.

In 2011, we have started developing a loss concealment scheme based on a new video examplar-based inpainting algorithm. The developed video inpainting approach relies on a motion confidence-aided neighbor embedding techniques. Neighbor embedding approaches aim at approximating input vectors (or data points) as a linear combination of their neighbors. We have considered two neighbor embedding approaches namely locally linear embedding (LLE) and non-negative matrix factorization (NMF), in a way that each patch of the target region is inpainted with the best estimation provided using template matching, LLE and NMF. The motion confidence introduced in the neighbor embedding improves the robustness of the algorithm with limiting the error propagation effects which may result from uncertainties on the motion information of the unknown pixels to be estimated. Evaluations of the algorithm in a context of video editing (object removal) show natural looking videos with less annoying artifacts [24].

This approach has then been adapted to the context of loss concealment that is to estimate unknown pixels after decoding when the corresponding transport packets have been lost on the transmission network. For this purpose, a preprocessing step is proposed to estimate the motion information of each corrupted block using Bilinear Motion Field Interpolation (BMFI) before inpainting the texture. The BMFI method computes the missing motion vector of each pixel in the lost block as a weighted combination of motion vectors of neighboring blocks. The estimated motion information is also used to limit the search space for the best matching patches in a motion-compensated window. Experiments of the proposed approach on several videos show a PSNR average gain about 2dB compared to state-of-art methods [25]. The next step will be to assess the performance of the approach in a context of free moving camera videos. To deal with this problem, we propose to consider a panoramic image mosaics in order to estimate the background of the video before inpainting the missing part of the foreground objects.

6.4.2. Unequal Erasure Protection and Object Bundle Protection

Participant: Aline Roumy.

In 2011, we started a new collaboration on Unequal Erasure Protection (UEP) and Object Bundle Protection in the framework of the joint research lab Inria–Alcatel Lucent and the ANR ARSSO project. Protection is usually obtained by adding Forward error correction (FEC) to the object (or data) to be transmitted. However, when the object contains information with different importance levels (as in a video bitstream), providing a protection adapted to the importance of each subpart of the object, helps reducing the encoded bitrate. To implement UEP, traditional transport protocols based on FEC Schemes need to split the original object into say two sub-objects, one per important class, and to submit each sub-object separately to the FEC Scheme. This requires extra logic for splitting/gathering the data. A companion problem, is the case where the object size is smaller than the packetsize. In this case, FEC traditional approaches applied to each small object is wasting the bandwidth. An optimized solution consists in grouping the small objects with equal importance into a single file. This is the goal of object bundle protection. We proposed a novel method, called Generalized Object Encoding that can deal with both aspects [37], [38], [39]. In 2011, we analyzed our GOE approaches with average metrics such as average waiting time, average number of packets to be encoded. In 2012, we continued the analysis and considered memory requirements at the decoder [30].

6.4.3. *Universal distributed coding*

Participant: Aline Roumy.

In 2012, we started a new collaboration with Michel Kieffer and Elsa Dupraz (Supelec, L2S) on universal distributed source coding. Distributed source coding refers to the problem where several correlated sources need to be compressed without any cooperation at the encoders. Decoding is however performed jointly. This problem arises in sensor networks but also in video compression techniques, where the correlation between the successive frames is not directly used at the encoder as in [17], and are therefore seen as distributed. Traditional approaches (from an information theoretical but also practical point of view) assume that the correlation channel between the sources is perfectly known. Since this assumption is not satisfied in practice, a way to get around this is to use a feedback channel (from the decoder to the encoder), that can trigger the encoder. Instead, we consider universal distributed source coding, where the correlation channel is unknown and belongs to a class parametrized by some unknown parameter vector. In [23], we proposed four uncertainty models that depend on the partial knowledge we have on the correlation channel and derived the information theoretical bounds.

6.4.4. *Super-resolution as a communication tool*

Participants: Marco Bevilacqua, Christine Guillemot, Aline Roumy.

In 2012, we carried on the collaboration with Alcatel Lucent Bell Labs, represented by M-L. Alberi Morel, in the framework of a Joint Inria/Alcatel Lucent lab. In this work, we continued investigating super resolution (SR) as a potential tool to use in the context of video transmission. As SR refers to the task of producing a high-resolution (HR) image from one or several low-resolution (LR) input images, one can think of sending a LR video to adapt to the complexity constraint of the encoder and/or the bandwidth limitation of the network, and still being able to reconstruct a HR video at the encoder side, by applying a SR algorithm.

As a first step toward the more ambitious goal of compressing video through SR, we developed a novel method for single-image SR based on a neighbor embedding technique. In the neighbor embedding based SR procedure, the LR input image is first divided into small patches, namely sub-windows of image. Each input patch is approximated by a linear combination of its nearest neighbors (LR candidate patches) taken from a dictionary. Then, the corresponding HR output patch is created by combining similarly the corresponding HR candidates of the dictionary. The SR image is finally obtained by aggregating all the single HR patches reconstructed. A key point of this approach is represented by the above mentioned dictionary, which is a stored set of LR and HR patch correspondences extracted from training natural images.

The studies undertaken led us to have two publications in international conferences [20], [19]: ICASSP (International Conference on Acoustics, Speech, and Signal Processing) and BMVC (British Machine Vision Conference). In [20] we presented a neighbor embedding based SR method, by following the general scheme, but also introducing a new method to compute the weights of the linear combinations of patches. The weights

of a certain input patch are computed as the result of a least squares problem with a nonnegative constraint. The so resulting nonnegative weights, that intuitevely represent a reasonable solution as they allow only additive combinations of patches, are shown to perform better than other weight computation methods described in the literature. The least squares problem is solved in a original fashion by means of SNMF, a tool for matrix factorization with one nonnegative factor. In [19] we refined the proposed algorithm, by focusing more on a low complexity target and by giving some theoretical insights about the choice of the nonnegative embedding. An analysis about the representation of the patches (either by the straight luminance values of its pixels or by some "features" conveniently computed) is also performed. The algorithm is shown to have better results, both in terms of quality performance and running time, than other similar SR algorithms that also adopt a one-pass procedure; and comparable visual results with respect to more sophisticated multi-pass algorithms, but still presenting a much reduced computational time. During the year, some other studies have been conducted, e.g. on the creation of the dictionary and on alternative ways to select the candidate patches from the dictionary. These extra studies, together with the already consolidated work of the published papers, represent the point of departure to the next step of designing a framework for video super resolution.

STARS Team

6. New Results

6.1. Introduction

This year Stars has proposed new algorithms related to its three main research axes : perception for activity recognition, semantic activity recognition and software engineering for activity recognition.

6.1.1. Perception for Activity Recognition

Participants: Julien Badie, Slawomir Bak, Vasanth Bathrinarayanan, Piotr Bilinski, Bernard Boulay, François Brémond, Sorana Capalnean, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Eben Freeman, Carolina Garate, Jihed Joober, Vaibhav Katiyar, Ratnesh Kumar, Srinidhi Mukanahallipatna, Sabine Moisan, Silviu Serban, Malik Souded, Anh Tuan Nghiem, Monique Thonnat, Sofia Zaidenberg.

This year Stars has extended an efficient algorithm for detecting people. We have also proposed a new algorithm for re-identification of people through a camera network. We have realized a new algorithm for the recognition of short actions and validated also its performance on several benchmarking databases (e.g. ADL). We have improved a generic event recognition algorithm by handling event uncertainty at several processing levels. More precisely, the new results for perception for activity recognition concern:

- Image Compression and Modelization (6.2)
- Background Subtraction (6.3)
- Fiber Based Video Segmentation (6.4)
- Enforcement of Monotonous Shape Growth/Shrinkage in Video Segmentation (6.5)
- Dynamic and Robust Object Tracking in a Single Camera View (6.6)
- Optimized Cascade of Classifiers for People Detection Using Covariance Features (6.7)
- Learning to Match Appearances by Correlations in a Covariance Metric Space (6.8)
- Recovering Tracking Errors with Human Re-identification (6.9)
- Human Action Recognition in Videos (6.10)
- Group Interaction and Group Tracking for Video-surveillance in Underground Railway Stations (6.11)
- Crowd Event Monitoring Using Texture and Motion Analysis (6.12)
- Detecting Falling People (6.13)
- People Detection Framework (6.14)

6.1.2. Semantic Activity Recognition

Participants: Sorana Capalnean, Guillaume Charpiat, Cintia Corti, Carlos -Fernando Crispim Junior, Hervé Falciani, Baptiste Fosty, Qioa Ma, Firat Ozemir, Jose-Luis Patino Vilchis, Guido-Tomas Pusiol, Rim Romdhame, Bertrand Simon, Abhineshwar Tomar.

Concerning semantic activity recognition, the contributions are :

- A Model-based Framework for Activity Recognition of Older People using Multiple sensors (6.15)
- Activity Recognition for Older People using Kinect (6.16)
- Descriptors of Depth-Camera Videos for Alzheimer Symptom Detection (6.17)
- Online Activity Learning from Subway Surveillance Videos (6.18)
- Automatic Activity Detection Modeling and Recognition: ADMR (6.19)

6.1.3. Software Engineering for Activity Recognition

Participants: François Brémond, Daniel Gaffé, Julien Gueytat, Baptiste Fosty, Sabine Moisan, Anh tuan Nghiem, Annie Ressouche, Jean-Paul Rigault, Leonardo Rocha, Luis-Emiliano Sanchez, Swaminathan Sankaranarayanan.

This year Stars has continued the development of the SUP platform. This latter is the backbone of the team experiments to implement the new algorithms. We continue to improve our meta-modelling approach to support the development of video surveillance applications based on SUP. This year we have focused on an architecture for run time adaptation and on metrics to drive dynamic architecture changes. We continue the development of a scenario analysis module (SAM) relying on formal methods to support activity recognition in SUP platform. We improve the theoretical foundations of CLEM toolkit and we rely on it to build SAM. Finally, we are improving the way we perform adaptation in the definition of a multiple services for device adaptive platform for scenario recognition.

The contributions for this research axis are:

- SUP Software Platform (6.20)
- Qualitative Evaluation of Detection and Tracking Performance (6.21)
- Model-Driven Engineering and Video-surveillance (6.22)
- Synchronous Modelling and Activity Recognition (6.23)

6.2. Image Compression and Modelization

Participants: Guillaume Charpiat, Eben Freeman.

Recent results in statistical learning have established the best strategy to combine several advices from different experts, for the problem of sequential prediction of times series. The notions of prediction and compression are tightly linked, in that a good predictor can be turned into a good compressor via entropy coding (such as Huffman coding or arithmetic coding), based on the predicted probabilities of the events to come : the more predictable an event E is, the easier to compress it will be, with coding cost $-\log(p(E))$ with such techniques.

The initial idea here, by Yann Ollivier (TAO team), within a collaboration with G. Charpiat and Jamal Atif (TAO team), was to adapt these results to the case of image compression, where time series are replaced with 2D series of pixel colors, and where experts are predictors of the color of a pixel given the colors of neighbors. The main difference is that there is no canonical physically-relevant 1D ordering of the pixels in an image, so that a sequential order (of the pixels to predict their colors) had to be defined first. Preliminary results with a hierarchical ordering scheme already competed with standard techniques in lossless compression (png, lossless jpeg2000).

During his internship in the Stars team, Eben Freeman developed this approach, by building relevant experts able to predict a variety of image features (regions of homogeneous color, edges, noise, ...). We also considered random orderings of pixels, using kernels to express probabilities in a spatially-coherent manner. Using such modellings of images with experts, we were also able to generate new images, that are typical of these models, and show more structure than the ones associated to standard compression schemes (typical images highly compressed).

6.3. Background Subtraction

Participants: Vasanth Bathrinarayanan, Anh-Tuan Nghiem, Duc-Phu CHAU, François Brémond.

Keywords: Gaussian Mixture Model, Shadow removal, Parameter controller, Codebook model, Context based information

6.3.1. Statistical Background Subtraction for Video Surveillance Platform

Anh-Tuan Nghiem work on background subtraction is an extended version of Gaussian Mixture Models [73]. The algorithm compares each pixel of current frame to background representation which is developed based on the pixel information from previous frames. It includes shadow and highlight removal to give better results. Selective background updating method based on the feedback from the object detection helps to better model background and remove noise and ghosts.

Figure 10 shows a sample illustration of the output of the background subtraction, where blue are foreground pixels and red are shadow or illumination change pixels and a green bounding box is a foreground blob. Also we have compared our algorithm with few other such as OpenCV and also IDIAP's background subtraction(not tuned perfectly, used default parameters) and the results are shown in Figure 11 where the green background refers to best performance of the comparisons. This evaluation is done on PETS 2009 data-set with our obtained foreground blobs to the manually annotated bounding boxes of people.

6.3.2. Parameter controller using Contextual features

The above method has some parameters that has to be tuned every time for each video, which is a time consuming work. The work of Chau et al [59] learns the contextual information from the video and controls object tracking algorithm parameters during the run-time of the algorithm. This approach is at preliminary stage for background subtraction algorithm to automatically adapt parameters. These parameters are learned as described in the offline learning process block diagram 12 over several ground truth videos and clustered into a database. The contextual feature which are used presently include object density, occlusion, contrast, 2D area, contrast variance, 2D area variance. Figure 13 shows a sample of video chunks based on contextual feature similarity for a video from caviar data-set.

The controller's preliminary results are promising and we are experimenting and evaluating with different features to learn the parameters. The results will be published in upcoming top computer vision conferences.

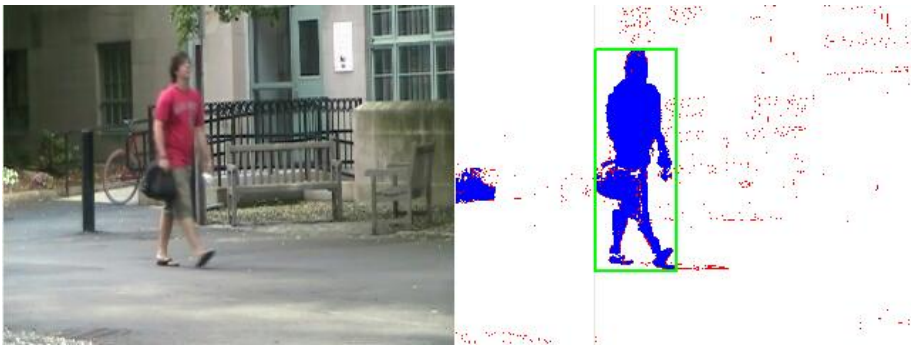


Figure 10. Sample illustration of output of EGMM

6.4. Fiber Based Video Segmentation

Participants: Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

Keywords: Video Volume, Fibers, Trajectory

PETS dataset Evaluation on Background subtraction blob with GT of people--			
Metrics	**IDIAP**	**SUP**	**OPENCV**
Metric M1.1			
Global results:			
Number of True Positives	3761	4046	4057
Number of False Positives	1807	317	1099
Number of False Negatives 0	1070	793	774
Precision (mean by frame)	0.71	0.93	0.81
Sensitivity 0 (mean by frame)	0.79	0.84	0.85
Precision (global)	0.68	0.93	0.79
Sensitivity 0 (global)	0.78	0.84	0.84

Figure 11. Evaluation with some background subtraction algorithms

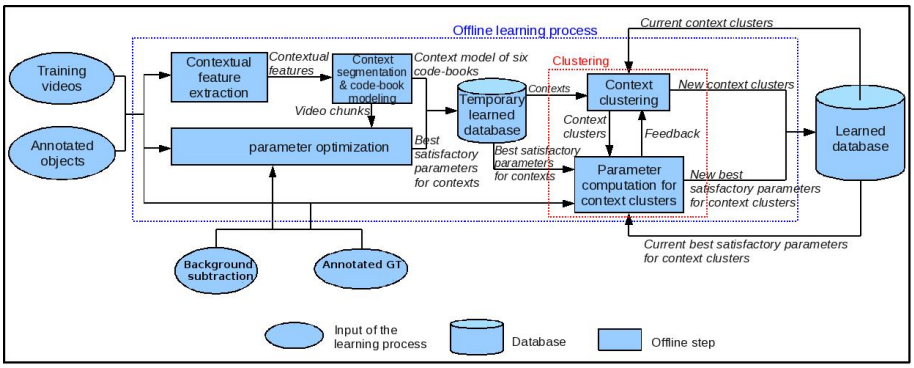


Figure 12. Block representation of the offline learning scheme to build a database

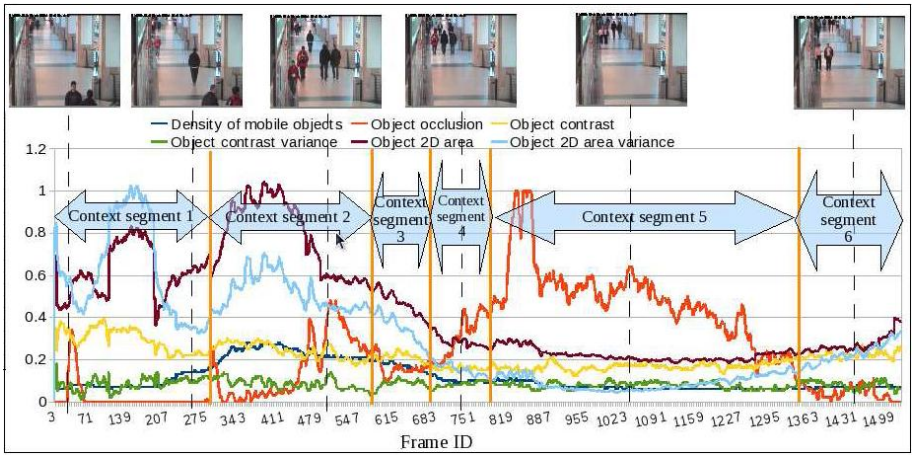


Figure 13. Context segmentation of the sequence ThreePastShop2cor (belonging to the Caviar dataset). The context segments are separated by the vertical orange lines. The control parameters are then learned for each context segment

The aim of this work is to segment objects in videos by considering videos as 3D volumetric data (2D×time). Figure 14 shows an input video and its corresponding partition in terms of fiber at a particular hierarchy level. Particularly, it shows 2D slices of a video volume. Bottom right corner of each figure shows the current temporal depth in the volume, while top right shows the X-time slice and bottom left shows Y-time slice. In this 3D representation of videos, points of static background form straight lines of homogeneous intensity over time, while points of moving objects form curved lines. Analogically to the fibers in MRI images of human brains, we term *fibers*, these straight and curved lines of homogeneous intensity. So, in our case, to segment the whole video volume data, we are interested in a dense estimation of fibers involving all pixels.

Initial fibers are built using correspondences computing algorithms like optical flow and descriptor matching. As these algorithms are reliable near corners and edges, we build fibers at these locations for a video. Our subsequent goal is to partition this video in terms of fibers built, by extending them (both spatially and temporally) to the rest of the video.

To extend fibers, we compute geodesics from pixels (not belonging to the initially built fibers) to fibers. For a reliable extension, the cost of moving along a geodesic is proportional to the trajectory similarity of a pixel *wrt* a fiber, wherein a pixel trajectory is similar to the fiber trajectory. This cost function quantifies the color homogeneity of a pixel trajectory along with its color similarity *wrt* a fiber. A pixel is then associated to a fiber for which this cost is minimum.

With the above mentioned steps we obtain a partition of a video in terms of fibers wherein we have a trajectory associated with each pixel. This hierarchical partition provides a mid-level representation of a video, which can be seen as a facilitator or a pre-processing step towards higher level video understanding systems *eg* activity recognition.

6.5. Enforcement of Monotonous Shape Growth/Shrinkage in Video Segmentation

Participant: Guillaume Charpiat.

keywords: graph cuts, video segmentation, shape growth

Figure 14. ***Left:*** *Input Video and Spatio-Temporal Slices.* ***Right:****Segmented Results at a Particular Hierarchy Level*

The segmentation of noisy videos or time series is a difficult problem, not to say an impossible or ill-posed task when the noise level is very high. While individual frames can be analysed independently, time coherence in image sequences provides a lot of information not available for a single image. Most of the state-of-art works explored short-term temporal continuity for object segmentation in image sequences, *i.e.*, each next frame is segmented by using information from one or several images at previous time points. It is, however, more advantageous to simultaneously segment many frames in the data set, so that segmentation of the entire image set supports each of the individual segmentations.

In this work, we focus on segmenting shapes in image sequences which only grow or shrink in time, and on making use of this knowledge as a constraint to help the segmentation process. Examples of growing shapes are forest fires in satellite images and organ development in medical imaging. We propose a segmentation framework based on graph cuts for the joint segmentation of a multi-dimensional image set. By minimizing an energy computed on the resulting spatio-temporal graph of the image sequence, the proposed method yields a *globally optimal solution*, and runs in practice in linear complexity in the total number of pixels.

Two applications are performed. First, with Yuliya Tarabalka (Ayin team), we segment multiyear sea ice floes in a set of satellite images acquired through different satellite sensors, after rigid alignment (see Figure 15). The method returns accurate melting profiles of sea ice, which is important for building climate models. The second application, with Bjoern Menze (ETH Zurich, also MIT and collaborator of Asclepios team), deals with the segmentation of brain tumors from longitudinal sets of multimodal MRI volumes. In this task we impose an additional inter-modal inclusion constraint for joint segmentation of different image sequences, finally also returning highly sensitive time-volume plots of tumor growth.

6.6. Dynamic and Robust Object Tracking in a Single Camera View

Participants: Duc-Phu Chau, Julien Badie, François Brémond, Monique Thonnat.

Keywords: Object tracking, online parameter tuning, controller, self-adaptation and machine learning

(a)

(b)

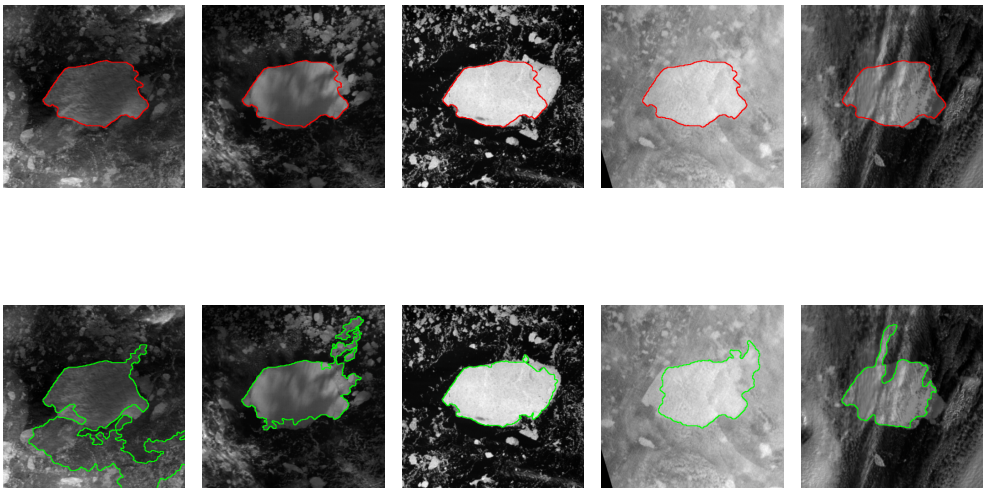


Figure 15. (a) Aligned satellite images captured each four days superposed with segmentation contours computed by our approach. (b) Segmentation contours for images (a) obtained by applying graph cut segmentation to each image at a single time moment. Note that the segmentations (a) are pixelwise precise, and that the white regions surrounding sometimes the boundaries are other ice blocks, agglomerating temporarily only, thus correctly labelled. Hence the importance of enforcing time coherence.

Object tracking quality usually depends on video scene conditions (e.g. illumination, density of objects, object occlusion level). In order to overcome this limitation, we present a new control approach to adapt the object tracking process to the scene condition variations. The proposed approach is composed of two tasks.

The objective of the first task is to select a convenient tracker for each mobile object among a Kanade-Lucas-Tomasi-based (KLT) tracker and a discriminative appearance-based tracker. The KLT feature tracker is used to decide whether an object is correctly detected. For badly detected objects, the KLT feature tracking is performed to correct object detection. A decision task is then performed using a Dynamic Bayesian Network (DBN) to select the best tracker among the discriminative appearance and KLT trackers.

The objective of the second task is to tune online the tracker parameters to cope with the tracking context variations. The tracking context, or context, of a video sequence is defined as a set of six features: density of mobile objects, their occlusion level, their contrast with regard to the surrounding background, their contrast variance, their 2D area and their 2D area variance. Each contextual feature is represented by a code-book model. In an offline phase, training video sequences are classified by clustering their contextual features. Each context cluster is then associated with satisfactory tracking parameters. In the online control phase, once a context change is detected, the tracking parameters are tuned using the learned values. This work has been published in [29], [35].

We have tested the proposed approach on several public datasets such as Caviar and PETS. Figure 16 illustrates the results of the object detection correction using the KLT feature tracker.

Figure 17 illustrates the tracking output for a Caviar video (on the left image) and for a PETS video (on the right image). The experimental results show that our method gets the best performance compared to some recent state of the art trackers.

Table 1 presents the tracking results for 20 videos from the Caviar dataset. The proposed approach obtains the best MT value (i.e. mostly tracked trajectories) compared to some recent state of the art trackers.

Figure 16. Illustration of the object detection correction for a Caviar video. The green bounding box is the output of the object detection process. The red bounding boxes are the results of the detection correction task.

a. b.

Figure 17. Tracking results for Caviar and PETS videos

Table 1. Tracking results on the Caviar dataset. MT: Mostly tracked trajectories, higher is better. PT: Partially tracked trajectories. ML: Most lost trajectories, lower is better. The best values are printed bold.

Method	MT (%)	PT (%)	ML (%)
Zhang et al., CVPR 2008 [89]	85.7	10.7	3.6
Li et al., CVPR 2009 [71]	84.6	14.0	1.4
Kuo et al., CVPR 2010 [69]	84.6	14.7	**0.7**
Proposed approach	**86.4**	10.6	3.0

Table 2 presents the tracking results of the proposed approach and three recent approaches [56], [82], [67] for a PETS video. With the proposed approach, we obtain the best values in both metrics MOTA (i.e. Multi-object tracking accuracy) and MOTP (i.e. Multi-object tracking precision). The authors in [56], [82], [67] do not present the tracking results with the MT, PT and ML metrics.

Table 2. Tracking results on the PETS sequence S2.L1, camera view 1, sequence time 12.34. MOTA: Multi-object tracking accuracy, higher is better. MOTP: Multi-object tracking precision, higher is better. The best values are printed bold.

Method	MOTA	MOTP	MT (%)	PT (%)	ML (%)
Berclaz et al., PAMI 2011 [56]	0.80	0.58	-	-	-
Shitrit et al., ICCV 2011 [82]	0.81	0.58	-	-	-
Henriques et al., ICCV 2011 [67]	0.85	0.69	-	-	-
Proposed approach	**0.86**	**0.72**	71.43	19.05	9.52

6.7. Optimized Cascade of Classifiers for People Detection Using Covariance Features

Participants: Malik Souded, François Brémond.

keywords: People detection, Covariance descriptor, LogitBoost.

We propose a new method to optimize a state of the art approach for people detection, which is based on classification on Riemannian manifolds using covariance matrices in a boosting scheme. Our approach makes training and detection faster while maintaining equivalent or better performances. This optimisation is achieved by clustering negative samples before training, providing a smaller number of cascade levels and less weak classifiers in most levels in comparison with the original approach.

Our approach is based on Tuzel et al. [86] work which was improved by Yao et al. [87]. We keep the same scheme to achieve our people detector: train a cascade of classifiers based on covariance descriptors, using a Logitboost training algorithm which was modified by Tuzel et al. to deal with the Riemannian manifolds metrics and using the operators which were presented in [75]. In fact, Covariance matrices do not belong to vector space but to the Riemannian manifold of (d x d) symmetric positive definite matrices. The trained cascade of classifiers is applied for detection after training.

We propose an additional step to speed up training and detection process. We propose to apply a clustering step on negative training dataset before training the classifiers. This clustering step is performed both in Riemannian manifold and in the vector space of mapped covariance matrices, using the operators and metrics previously cited.

The idea consists in regrouping all similar negative samples, with regard to their covariance information, into decreasing size clusters. Each classifier of the cascade is trained on one cluster, specializing this classifier for a given kind of covariance information, and then, speeding up the training step and providing shorter classifier, which accelerate its response when applied on image. In the same time, the specialization of each cascade classifier shortens the cascade too, speeding up the detection (see Figure 18 and Figure 19).

A paper describing this approach has been accepted in VISAPP 2013 conference [50].

6.8. Learning to Match Appearances by Correlations in a Covariance Metric Space

Participants: Sławomir Bąk, Guillaume Charpiat, Etienne Corvée, Francois Brémond, Monique Thonnat.

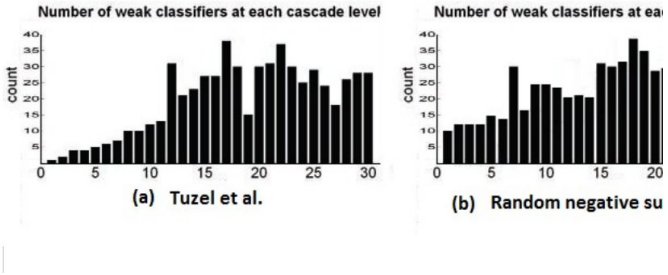


Figure 18. Comparison of classifiers cascade structure

(a) (b) (c)

Figure 19. Cascade classifiers: some detection results

keywords: covariance matrix, re-identification, appearance matching

This work addresses the problem of appearance matching across disjoint camera views. Significant appearance changes, caused by variations in view angle, illumination and object pose, make the problem challenging.

We propose to formulate the appearance matching problem as the task of learning a model that selects the most descriptive features for a specific class of objects. Our main idea is that different regions of the object appearance ought to be matched using various strategies to obtain a distinctive representation. Extracting region-dependent features allows us to characterize the appearance of a given object class (*e.g.* class of humans) in a more efficient and informative way. Different kinds of features characterizing various regions of an object is fundamental to our appearance matching method.

We propose to model the object appearance using covariance descriptor yielding rotation and illumination invariance. Covariance descriptor has already been successfully used in the literature for appearance matching. In contrast to state of the art approaches, we do not define *a priori* feature vector for extracting covariance, but we learn which features are the most descriptive and distinctive depending on their localization in the object appearance (see figure 20). Learning is performed in a covariance metric space using an entropy-driven criterion. Characterizing a specific class of objects, we select only essential features for this class, removing irrelevant redundancy from covariance feature vectors and ensuring low computational cost.

The proposed technique has been successfully applied to the person re-identification problem, in which a human appearance has to be matched across non-overlapping cameras [34]. We demonstrated that: (1) by using different kinds of covariance features *w.r.t.* the region of an object, we obtain clear improvement in appearance matching performance; (2) our method outperforms state of the art methods in the context of pedestrian recognition on publicly available datasets (i-LIDS-119, i-LIDS-MA and i-LIDS-AA); (3) using 4×4 covariance matrices we significantly speed-up the processing time offering an efficient and distinctive representation of the object appearance.

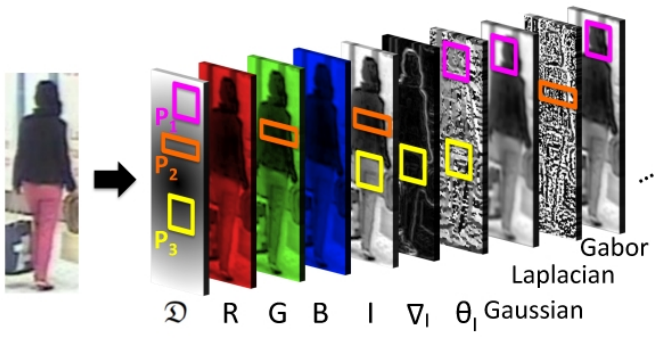


Figure 20. Example of three different covariance features. Every covariance is extracted from a region (P), distance layer (D) and three channel functions (e.g. bottom covariance feature is extracted from region P_3 using layers: D, I-intensity, ∇_I-gradient magnitude and θ_I-gradient orientation).

6.9. Recovering Tracking Errors with Human Re-identification

Participants: Julien Badie, Slawomir Bak, Duc-Phu Chau, François Brémond, Monique Thonnat.

keywords: tracking error correction, re-identification

This work addresses the problem of people tracking at long range even if the target people are lost several times by the tracking algorithm. We have identified two main reasons for tracking interruption. The first one concerns interruptions that can be quickly recovered, which includes short mis-detections, occlusions with other persons or static obstacles. The second one occurs when a person is occluded or mis-detected for a long time or when the person leaves the scene and comes back latter. Our main objective is to design a framework that can track people even if their trajectory is very segmented and/or associated with different IDs. We called this problem the global tracking challenge (see Figure 21).

Figure 21. The global tracking challenge : correcting errors due to occlusions (ID 142 on the first frame becomes 147 on the last frame) and tracking people that are leaving the scene and reentering (ID 133 on the first frame becomes 151 on the last frame).

In order to describe a person's tracklet (segment of trajectory), we use a visual signature called Mean Riemannian Covariance Grid and a discriminative method to emphasize the main differences between each tracklet. This step improves the reliability and the accuracy of the results. By computing the distance between the visual signatures, we are able to link tracklets belonging to the same person into a tracklet cluster. Only tuples of tracklets that are not overlapping each other are used as initial candidates. Then, we use Mean Shift to create the clusters.
We evaluated this method on several datasets (i-LIDS, Caviar, PETS 2009). We have shown that our approach can perform as well as the other state of the art methods on Caviar and can perform better on i-LIDS. On PETS 2009 dataset, our approach performs better than standard tracker but cannot be compared with the best state of the art methods due to unadapted metrics.
This approach is described in detail in two articles : one published in ICIP 2012 [35], which is focused on computing the covariance signature and the way to discriminate it and the other one published in PETS 2012 workshop (part of AVSS 2012 conference) [33], which is focused on the method to link the tracklets.
This work will be added to a more general tracking controller that should be able to detect several kinds of detection and tracking errors and try to correct them.

6.10. Human Action Recognition in Videos

Participants: Piotr Bilinski, François Brémond.

keywords: Action Recognition, Contextual Features, Pairwise Features, Relative Tracklets, Spatio-Temporal Interest Points, Tracklets, Head Estimation.

The goal of this work is to automatically recognize human actions and activities in diverse and realistic video settings.

Over the last few years, the bag-of-words approach has become a popular method to represent video actions. However, it only represents a global distribution of features and thus might not be discriminative enough. In particular, the bag-of-words model does not use information about: local density of features, pairwise relations among the features, relative position of features and space-time order of features. Therefore, we propose three new, higher-level feature representations that are based on commonly extracted features (e.g. spatio-temporal interest points used to evaluate the first two feature representations or tracklets used to evaluate the last approach). Our representations are designed to capture information not taken into account by the model, and thus to overcome its limitations.

In the first method, we propose new and complex contextual features that encode spatio-temporal distribution of commonly extracted features. Our feature representation captures not only global statistics of features but also local density of features, pairwise relations among the features and space-time order of local features. Using two benchmark datasets for human action recognition, we demonstrate that our representation enhances the discriminative power of commonly extracted features and improves action recognition performance, achieving 96.16% recognition rate on popular KTH action dataset and 93.33% on challenging ADL dataset. This work has been published in [36].

In the second approach, we design new representation of features encoding statistics of pairwise co-occurring local spatio-temporal features. This representation focuses on pairwise relations among the features. In particular, we introduce the geometric information to the model and associate geometric relations among the features with appearance relations among the features. Despite that local density of features and space-time order of local features are not captured, we are able to achieve similar results on the KTH dataset (96.30% recognition rate) and 82.05% recognition rate on UCF-ARG dataset. An additional advantage of this method is to reduce the processing time of training the model from one week on a PC cluster to one day. This work has been published in [37].

In the third approach, we propose a new feature representation based on point tracklets and a new head estimation algorithm. Our representation captures a global distribution of tracklets and relative positions of tracklet points according to the estimated head position. Our approach has been evaluated on three datasets, including KTH, ADL, and our locally collected Hospital dataset. This new dataset has been created in cooperation with the CHU Nice Hospital. It contains people performing daily living activities such as: standing up, sitting down, walking, reading a magazine, *etc*. Sample frames with extracted tracklets from video sequences of the ADL and Hospital datasets are illustrated on Figure 22 . Consistently, experiments show that our representation enhances the discriminative power of tracklet features and improves action recognition performance. This work has been accepted for publication in [38].

Figure 22. Sample frames with extracted tracklets from video sequences of the ADL (left column) and Hospital (right column) datasets.

6.11. Group Interaction and Group Tracking for Video-surveillance in Underground Railway Stations

Participants: Sofia Zaidenberg, Bernard Boulay, Carolina Garate, Duc-Phu Chau, Etienne Corvée, François Brémond.

Keywords: events detection, behaviour recognition,automatic video understanding, tracking

One goal in the European project VANAHEIM is the tracking of groups of people. Based on frame to frame mobile object tracking, we try to detect which mobiles form a group and to follow the group through its lifetime. We define a group of people as two or more people being close to each other and having similar trajectories (speed and direction). The dynamics of a group can be more or less erratic: people may join or split from the group, one or more can disappear temporarily (occlusion or disappearance from the field of view) but reappear and still be part of the group. The motion detector which detects and labels mobile objects may also fail (misdetections or wrong labels). Analysing trajectories over a temporal window allows handling this instability more robustly. We use the event-description language described in [88] to define events, described using basic group properties such as size, type of trajectory or number and density of people and perform the recognition of events and behaviours such as violence or vandalism (alarming events) or a queue at the vending machine (non-alarming events).

The group tracking approach uses Mean-Shift clustering of trajectories to create groups. Two or more individuals are associated in a group if their trajectories have been clustered together by the Mean-Shift algorithm. The trajectories are given by the long-term tracker described in [60]. Each trajectory is composed of a person's positions (x, y) on the ground plane (in 3D) over the time window, and of their speed at each frame in the time window. Positions and speed are normalized using the minimum and maximum possible values (0 and $10m/s$ for the speed and the field of view of the camera for the position). The Mean-Shift algorithm requires a *tolerance* parameter which is set to 0.1, meaning that trajectories need to be distant by less than 10% of the maximum to be grouped.

Figure 23. Example of a group composed of non-similar individual trajectories.

As shown in Figure 23 , people in a group might not always have similar trajectories. For this reason, a group is also created when people are very close. A group is described by its coherence, a value calculated from the average distances of group members, their speed similarity and direction similarity. The update phase of the group uses the coherence value. A member will be kept in a group as long as the group coherence is above a threshold. This way, a member can temporarily move apart (for instance to buy a ticket at the vending machine) without being separated from the group.

This work has been applied to the benchmark CAVIAR dataset for testing, using the provided ground truth for evaluation. This dataset is composed of two parts: acted scenes in the Inria hall (9 sequences of 665 frames in average) and not acted recordings from a shopping mall corridor (7 sequences processed of 1722 frames in average). The following scenarios have been defined using the event-description language of [88]: *fighting*, *split up*, *joining*, *shop enter*, *shop exit*, *browsing*. These scenarios have been recognized in the videos with a high success rate (94%). The results of this evaluation and the above described method have been published in [45].

The group tracking algorithm is integrated at both Torino and Paris testing sites and runs in real time on live video streams. The global VANAHEIM system has been presented as a demonstration at the ECCV 2012 conference. A demonstration video has been compiled from the results of the group tracking on 60 sequences from the Paris subway showing interesting groups with various activities such as *waiting*, *walking*, *lost*, *kids* and *lively*.

6.12. Crowd Event Monitoring Using Texture and Motion Analysis

Participants: Vaibhav Katiyar, Jihed Joober, François Brémond.

keywords: Crowd Event, Texture Analysis, GLCM, Optical Flow

The aim of this work is to monitor crowd event using crowd density, change of speed and orientation of group of people. For reducing complexity we are using human density rather than individual human detection and tracking. In this study Human density is quantified mainly into three groups- (1) Empty (2) Sparse (3) Dense. These are approximated by calculating Haralick features from Grey Level Co-occurrence Matrix (GLCM).

We use Optical flow for getting motion information like current speed and orientation of selected FAST feature points. Subsequently we used this information for classifying crowd behaviour into normal or abnormal categories wherein we seek for sudden change in speed or orientation heterogeneity for abnormal behaviour.

In future work this abnormal behaviour may further be classified into different events like Running, Collecting, Dispersion, Stopping/Blocking.

6.13. Detecting Falling People

Participants: Etienne Corvee, Francois Bremond.

keywords: fall, tracking, event

We have developed a people falling algorithm based on our object detection and tracking algorithm [58] and using our ontology based event detector [57]. These algorithms extract moving object trajectories from videos and triggers alarms whenever the people activity fits event models. Most surveillance systems use a multi Gaussian technique [83] to model background scene pixels. This technique is very efficient in detecting in real-time moving objects in scenes captured by a static camera, with low level of shadows, few persons interacting in the scene and with as few as possible illumination changes. This technique does not analyse the content of the moving pixels but simply assign them as foreground or background pixels.

Many state of the art algorithms exist that can recognize objects such as a person human shape, a head, a face or a couch. However, these algorithms are quite time consuming or the database used for training the system is not well adapted to our application domain. For example, people detection algorithms use databases containing thousands of image instances of standing or walking persons taken by camera from a certain distance from the persons and from a facing position. In our indoor monitoring application, cameras are located on the roof with high tilt angle so that most of the scene (e.g.rooms) is viewed. With such camera spatial configuration, the image of a person on the screen rarely corresponds to the person images in the training database. In addition, people are often occluded by the image border (the image of the full body is not available), image distortion needs to be corrected and people often have poses that are not present in the database (e.g. a person bending or sitting).

Using our multi Gaussian technique [74], after having calibrated a camera scene, a detected object is associated with a 3D width and height in two positions : the standing and lying positions. This 3D information is checked against 3D human model and any object is then labelled as either a standing person, a lying person or unknown. Many 3D filtering thresholds are used ; for example, object speed should not be greater than a human possible running speed. Second, we use an ontology based event detector to build a hierarchy of event model complexity. We detect people to have fallen on the floor if the object has been detected as a person on the floor and outside the bed and couch for at least several seconds consecutively. An example of a fallen person is shown in Figure 24 .

Figure 24. Detection of a fallen person.

6.14. People Detection Framework

Participants: Srinidhi Mukanahallipatna, Silviu-Tudor Serban, François Brémond.

keywords: LBP, Adaboost, Cascades

We present a new framework called COFROD (Comprehensive Optimization Framework for Real-time Object Detection) for object detection that focuses on improving state of the art accuracy, while maintaining real-time detection speed. The general idea behind our work is to create an efficient environment for developing and analyzing novel or optimized approaches in terms of classification, features, usage of prior knowledge and custom strategies for training and detection. In our approach we opt for a standard linear classifier such as Adaboost. Inspired by the integral channel feature approach, we compute variants of LBP and Haar-like features on multiple channels of the input image. Thus, we obtain an elevated number of computationally inexpensive features that capture substantial information. We use an extensive training technique in order to obtain optimal classifier.

We propose a comprehensive framework for object detection with an intuitive modular design and high emphasis on performance and flexibility. Its components are organized by parent-modules, child-modules and auxiliary-modules. The parent-modules contain several child-modules and focus on a general task such as Training or Detection. Child-modules solve more specific tasks, such as feature extraction, training or testing and in most cases require auxiliary-modules. The later have precise intents, for instance computing a color channel transformation or a feature response.

We present two detection configurations. One relies on a single intensively trained detector and the other as a set of specialist detectors.

Our baseline detector uses cascades in order to speed up the classifier. By removing most false positive at first stages, computation time is significantly reduced. Classifier for each cascade is generated using the training approach.

Our contribution is in the form of a hierarchical design of specialized detectors. At first level we use a version of the baseline detector in order to remove irrelevant candidates. At the second level, specialist detectors are defined. These detectors can be either independent or can use third level detectors and cumulate their output. A specialist detector can take the role of solving an exact classification issue, such as sitting pose. In that case it is trained only with data relevant to that task. In some applications, a specialist detector can be trained to perform exceptionally on a specific situation. In this case training samples are adapted to the particularity of the testing, and possibly parts of the testing sets are used for training.

This is a versatile system for object detection that excels in both accuracy and speed. We present a valuable strategy for training and a hierarchy of specialized people detectors for dealing with difficult scenarios. We also propose an interesting feature channel and a method for loosing less detection speed-up. In our approach we build upon the ideas of feature scaling instead of resizing images and of transferring most computations from detection to training, thus achieving real-time performance on VGA resolution.

Figure 25 and Figure 26 illustrate our detections results. Figure 27 shows the performance of our system compared to other. IDIAP detector was used without tuning the parameters.

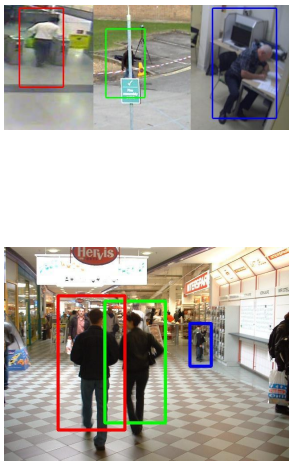


Figure 25. Detection Results

6.15. A Model-based Framework for Activity Recognition of Older People using Multiple sensors

Participants: Carlos -Fernando Crispim Junior, Qiao Ma, Baptiste Fosty, Cintia Corti, Véronique Joumier, Philippe Robert, Alexandra Konig, François Brémond, Monique Thonnat.

keywords: Activity Recognition, Multi-sensor Analysis, Surveillance System, Older people, Frailty assessment

We have been investigating a model-based activity recognition framework for the automatic detection of physical activity tests and instrumental activities of daily living (IADL, *e.g.*, preparing coffee, making a phone call) of older people. The activities are modelled using a constraint-based approach (using spatial, temporal, and *a priori* information of the scene), and a generic ontology based on natural terms which allows medical experts to easily modify the defined activity models. Activity models are organized in a hierarchical structure according to their complexity (Primitive state, Composite State, Primitive Event, and Composite Event). The framework has been tested as a system on the clinical protocol developed by the Memory Center of Nice hospital. This clinical protocol aims at studying how ICTs (Information and Communication Technologies)

Figure 26. PETS Detection Results

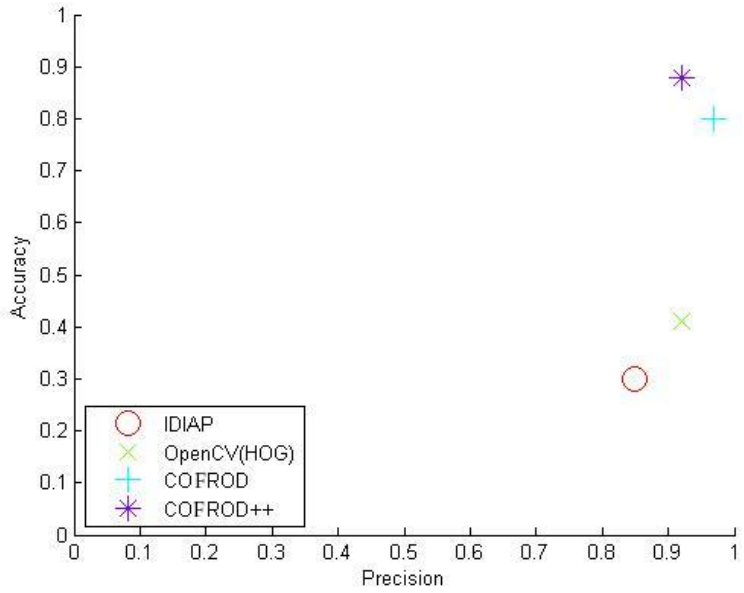


Figure 27. Detection results on the PETS dataset

can provide objective evidence of early symptoms of Alzheimer's disease (AD) and related conditions (like Memory Cognitive Impairment - MCI). The Clinical protocol participants are recorded using a RGB video-camera (8 fps), a RGB-D Camera (Kinect - Microsoft), and an inertial sensor (MotionPod) which allows a multi-sensor evaluation of the activities of the participants in an observation room equipped with home appliances. A study of the use of a multi-sensor monitoring for Patient diagnosis using events annotated by experts has been performed in partnership with CHU-Nice and SMILE team of TAIWAN, and it has shown the feasibility of the use of these sensors for patient performance evaluation and differentiation of clinical protocol groups (Alzheimer's disease and healthy participants) [31] and [40]. The multi-sensor evaluation has used the proposed surveillance system prototype and has been able to detect the full set of physical activities of the scenario 1 of the clinical protocol (e.g., Guided a ctivities : Balance test, Repeated Transfer Test), with a true positive rate of 96.9% to 100% for a set of 38 patients (MCI=19, Alzheimer=9) using data of an ambient camera. An extension of the developed framework has been investigated to handle multiple sensors data in the event modeling. In this new scenario, information from the ambient camera and the inertial sensor worn on the participants chest is used (see Figure 28). The prototype using the extended framework has been tested on the automatic detection of IADLs, and preliminary results points to an average sensitivity of 91% and an average precision of 83.5%. This evaluation has been performed for 9 participants videos (15 min each, healthy: 4, MCI: 5). See [39] for more details. Future work will focus on a learning mechanism to automatic fuse events detected by a set of heterogeneous sensors, and at supporting clinicians at the task of studying differences between the activity profile of healthy participants and early to moderate stage Alzheimer's patients.

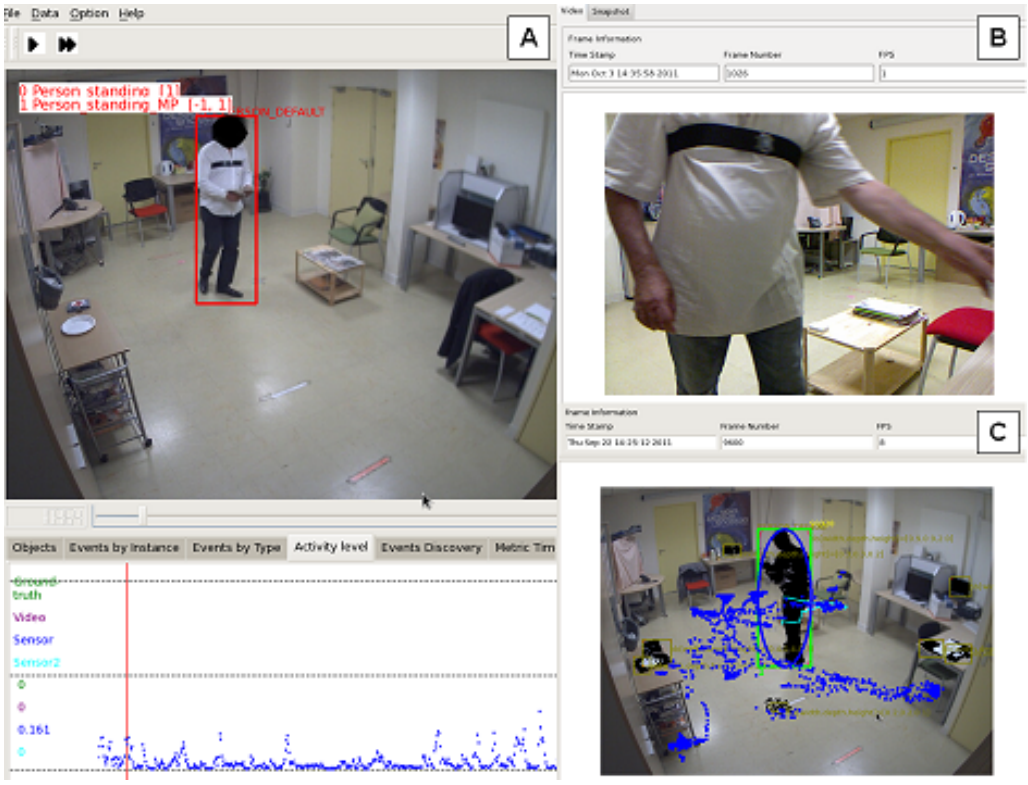


Figure 28. A: Ambient Camera View of Patient Activity. Actimetry captured by the inertial sensor is displayed at the bottom, B: RGB-D Camera View of Patient. The inertial sensor is worn by the patient by an accessory chest strap, C: Trajectory information of Patient Activity during the experimentation.

6.16. Activity Recognition for Older People using Kinect

Participants: Baptiste Fosty, Carlos -Fernando Crispim Junior, Véronique Joumier, Philippe Robert, Alexandra Konig, François Brémond, Monique Thonnat.

keywords: Activity Recognition, RGB-D camera analysis, Surveillance System, Older people, Frailty assessment

Within the context of the Dem@Care project, we have studied the potential of the RGB-D camera (Red Green Blue + Depth) from Microsoft (Kinect) for an activity recognition system developed to extract automatically and objectively evidences of early symptoms of Alzheimer's disease (AD) and related conditions (like Memory Cognitive Impairment - MCI) for older people. This system is designed on a model-based activity recognition framework. Using a constraint-based approach with contextual and spatio-temporal informations of the scene, we have developped activity models related to the physical activity part of the protocol (Scenario 1, guided activities : balance test, walking test, repeated transfers posture between sitting and standing). These models are organized in a hierarchical structure according to their complexity (Primitive state, Composite State, Primitive Event, and Composite Event). This work is an adaptation of the work performed for multi-sensor analysis [39].

Several steps are needed to adapt the processing. We had for example to generate new ground truth, or we had to design new 3D zones of interest according to Kinect point of view and referential (differing from the 2D camera). Moreover, in order to improve the reliability of the results, we had to solve several issues in the processing chain. For instance, Kinect and the detection algorithm provided by OpenNi and Nestk (free libraries) have several limitations which leads to wrong detection of human. We proposed in these cases several solutions like filtering wrong object detections by size (see Figure29 C) or recomputing the height of older people based on their head when wearing black pants (absorption of infrared) (see Figure 29 D).

For the experimentation, we have processed the data recorded for 30 patients. The results are shown in Figure 30 . With a true positive rate of almost 97% and a precision of 94.2%, our system is able to extract most of the activities performed by patients. Then, relevant and objective information can be delivered to clinicians, to assess the patient frailty. For further information on the performance of the detection process, we also generate the results frame by frame, which are shown in Figure 31 . We see there that the performance of the event detection in terms of true positive rate is almost as good as by events (94.5%). Nevertheless, if we focus on the precision, it is lower than previously. This means that we still need to improve detection accuracy of the beginning and the end of an event.

Future work will focus on using the human skeleton to extract finest information on the patient activity and to process more scenarios (semi-guided and free).

6.17. Descriptors of Depth-Camera Videos for Alzheimer Symptom Detection

Participants: Guillaume Charpiat, Sorana Capalnean, Bertrand Simon, Baptiste Fosty, Véronique Joumier.

keywords: Kinect, action description, video analysis

In a collaboration with the CHU hospital of Nice, a dataset of videos was recorded, where elderly are asked by doctors to perform a number of predefined exercises (like walking, standing-sitting, equilibrium test), and recorded with an RGBD camera (Kinect). Our task is to analyze the videos and detect automatically early Alzheimer symptoms, through statistical learning. Here we focus on the 3D depth sensor (no use of the RGB image), and aim at providing action descriptors that are accurate enough to be informative.

During her internship in the Stars team, Sorana Capalnean proposed descriptors relying directly on the 3D points of the scene. First, based on trajectory analysis, she proposed a way to recognize the different physical exercises. Then she proposed, for each exercise, specific descriptors aiming at providing the information asked by doctors, such as step length, frequency and asymmetry for the walking exercise, or sitting speed and acceleration for the second exercise, etc. Problems to deal with included the high level of noise in the 3D cloud of points given by the Kinect, as well as an accurate localization of the floor.

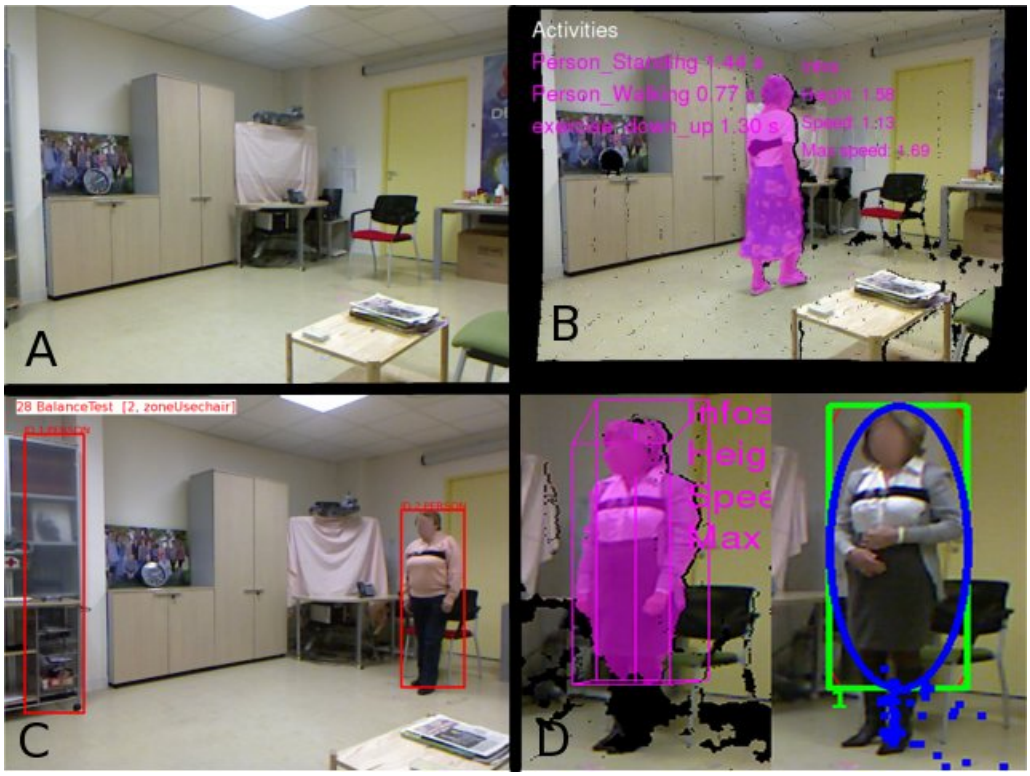


Figure 29. A: RGB-D camera view of the scene, B: 3D representation of the scene with some event detection, C: people detection problem (furniture detected as extra person), D: people detection problem (black clothes not detected).

	Nb GT	Detected		Sensitivity (%)	Precision (%)	Fscore (%)
BalanceTest	30	VP FP FN	30 0 0	100	100	100
WalkingTest_firstAttempt	30	TP FP FN	30 3 0	100	90,9	95,2
WalkingTest_secondAttempt	30	TP FP FN	27 0 3	90	100	94,7
RepeatedTransfersTest	30	TP FP FN	30 3 0	100	90,9	95,2
UpAndGoTest	30	TP FP FN	28 3 2	93,3	90,3	91,8
TOTAL	**150**	**TP** **FP** **FN**	**145** **9** **5**	**96,6**	**94,2**	**95,4**

Figure 30. Results by events (GT = ground truth, TP = true positive, FP = false positive, FN = false negative)

	Nb GT	Detected		Sensitivity (%)	Precision (%)	Fscore (%)
BalanceTest	37235	TP FP FN	37216 15322 19	99,9	70,8	82,9
WalkingTest_firstAttempt	1537	TP FP FN	849 53 688	55,2	94,1	69,6
WalkingTest_secondAttempt	1425	TP FP FN	857 513 568	60,1	62,6	61,3
RepeatedTransfersTest	5190	TP FP FN	4496 257 694	86,6	94,6	90,4
UpAndGoTest	3345	TP FP FN	2670 428 675	79,8	86,2	82,9
TOTAL	**48732**	**TP** **FP** **FN**	**46088** **16573** **2644**	**94,5**	**73,6**	**82,8**

Figure 31. Results by events (GT = ground truth, TP = true positive, FP = false positive, FN = false negative)

During his internship, Bertrand Simon proposed other kinds of descriptors, based on the articulations of the human skeleton given by OpenNI. These articulations are however very noisy too, so that a pre-filtering step of the data in time had to be performed. Various coordinate systems were studied, to reach the highest robustness. The work focused not only on descriptors but also on metrics suitable to compare gestures (in the phase space as well as in the space of trajectories). See figure 32 for an example.

These descriptors are designed to be robust to camera noise and to extract the relevant information from the videos; however their statistical analysis still remains to be done, to recognize Alzheimer symptoms during the different exercises.

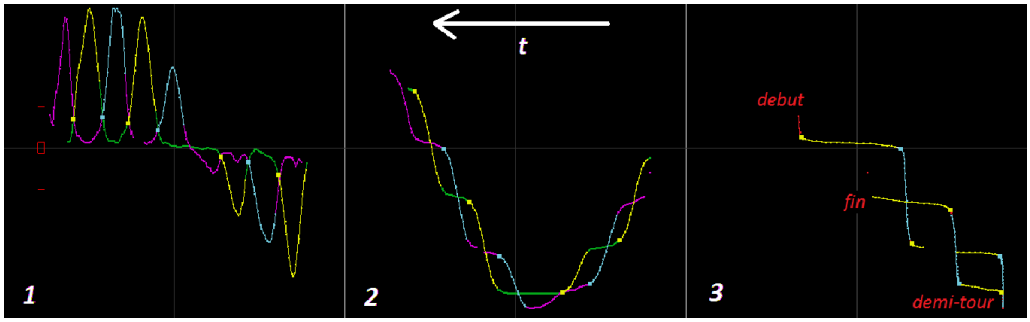


Figure 32. Curves obtained during a person's walk in backward then forward directions. Purple and blue curves stand for the right foot, while green and yellow ones stand for the left one. Graph 1 shows the speeds as a function of time; Graph 2 shows the locations as a function of time; Graph 3 shows the location of the right foot as a function of the location of the left foot.

6.18. Online Activity Learning from Subway Surveillance Videos

Participants: Jose-Luis Patino Vilchis, Abhineshwar Tomar, François Brémond, Monique Thonnat.

Keywords: Activity learning, clustering, trajectory analysis, subway surveillance

This work provides a new method for activity learning from subway surveillance videos. This is achieved by learning the main activity zones in the observed scene by taking as input the trajectories of detected mobile objects. This provides us the information on the occupancy of the different areas of the scene. In a second step, these learned zones are employed to extract people activities by relating mobile trajectories to the learned zones, in this way, the activity of a person can be summarised as the series of zones that the person has visited. If the person resides in the single zone this activity is also classified as a standing. For the analysis of the trajectory, a multiresolution analysis is set such that a trajectory is segmented into a series of tracklets based on changing speed points thus extracting the information when people stop to interact with elements of the scene or other people. Starting and ending tracklet points are fed to an advantageous incremental clustering algorithm to create an initial partition of the scene. Similarity relations between resulting clusters are modelled employing fuzzy relations. A clustering algorithm based on the transitive closure calculation of the fuzzy relations easily builds the final structure of the scene. To allow for incremental learning and update of activity zones (and thus people activities), fuzzy relations are defined with online learning terms. The approach is tested on the extraction of activities from the video recorded at one entrance hall in the Torino (Italy) underground system. Figure 33 presents the learned zones corresponding to the analyzed video. To test the validity of the activity extraction a one hour video was annotated with activities (corresponding to each trajectory) according to user defined ground-truth zones. After the comparison, following results were obtained: TP:26, FP:3, FN:1, Precision:0.89, Sensitivity:0.96. This work is published in [43].

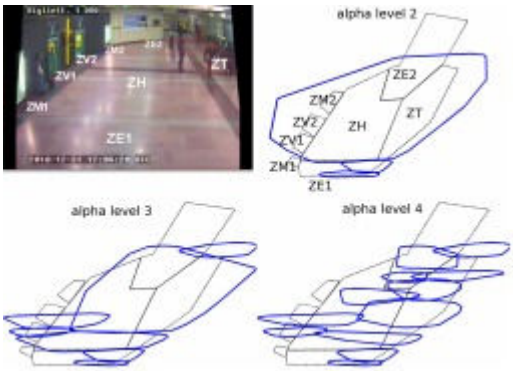


Figure 33. Left top panel: Original underground scene observed by the camera with user-defined areas delimiting the scene. Remaining panels: Learned zones in a 3D top view. They correspond to activity areas as discovered with our algorithm. Different granularity levels allow understanding the activity with different resolutions

6.19. Automatic Activity Detection Modeling and Recognition: ADMR

Participants: Guido-Tomas Pusiol, François Brémond.

This year a new Ph.D. thesis has been defended [30]. The main objective of the thesis is to propose a complete framework for the automatic activity discovery, modeling and recognition using video information. The framework uses perceptual information (e.g. trajectories) as input and goes up to activities (semantics). The framework is divided into five main parts:

1. We break the video into chunks to characterize activities. We propose different techniques to extract perceptual features from the chunks. This way, we build packages of perceptual features capable of describing activity occurring in small periods of time.
2. We propose to learn the video contextual information. We build scene models by learning salient perceptual features. The models end up containing interesting scene regions capable of describing basic semantics (i.e. region where interactions occur).
3. We propose to reduce the gap between low-level vision information and semantic interpretation, by building an intermediate layer composed of Primitive Events. The proposed representation for primitive events aims at describing the meaningful motions over the scene. This is achieved by abstracting perceptual features using contextual information in an unsupervised manner.
4. We propose a pattern-based method to discover activities at multiple resolutions (i.e. activities and sub-activities). Also, we propose a generative method to model multi-resolution activities. The models are built as a flexible probabilistic framework easy to update.
5. We propose an activity recognition method that finds in a deterministic manner the occurrences of modelled activities in unseen datasets. Semantics are provided by the method under user interaction. All this research work has been evaluated using real datasets of people living in an apartment (home-care application) and elder patients in a hospital.

The work has also been evaluated for other types of applications such as sleeping monitoring. For example, Figure 34 display the results of the activity discovery method during 6 hours (left to right) applied to the center of mass (3D) of a tracked sleeping person. The colored segments represent hierarchical (bottom-up is finer-coarse) discovered activity which matches with sleeping postural movements. The segments have similar color when postural movements are similar. For example, the segment (j) is the only time the person sleeps upside down. Also, health professionals analysed the results claiming that the segments corresponds to normal

sleeping cycle, where low motion is noticed at the beginning of the sleep and more motion is shown when the person have a lighter sleep when starts waking up.

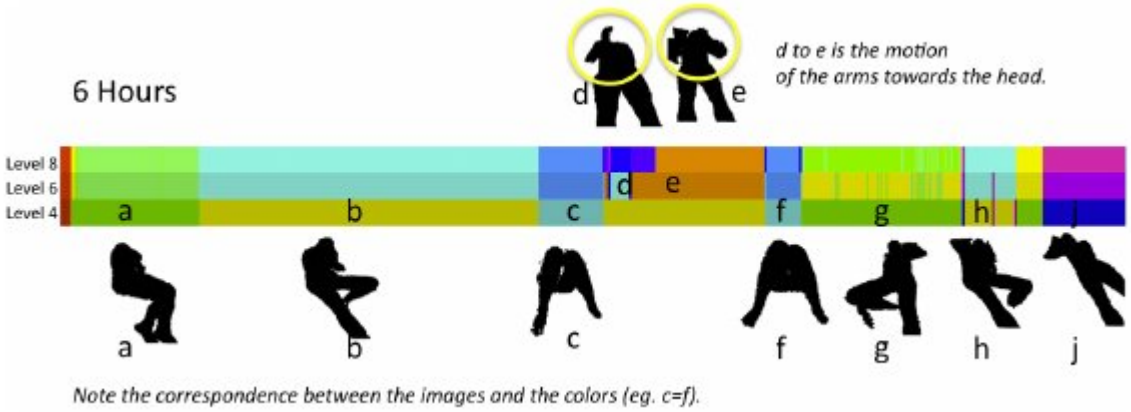


Figure 34. Results of the activity discovery method during 6 hours

6.20. SUP Software Platform

Participants: Julien Gueytat, Baptiste Fosty, Anh tuan Nghiem, Leonardo Rocha, François Brémond.

Our team focuses on developing Scene Understanding Platform (SUP) (see section 5.1). This platform has been designed for analyzing a video content. SUP is able to recognize simple events such as 'falling', 'walking' of a person. We can easily build new analyzing system thanks to a set of algorithms also called plugins. The order of those plugins and their parameters can be changed at run time and the result visualized. This platform has many more advantages such as easy serialization to save and replay a scene, portability to Mac, Windows or Linux, ... All those advantages are available since we are working together with the software developers team DREAM. Many Inria teams are pushing together to improve a common Inria development toolkit DTK. Our SUP framework is one of the DTK-like framework developed at Inria.
Currently, we have fully integrated OpenCV library with SUP and the next step is to integrate OpenNI to get depth map processing algorithms from PrimeSense running in SUP. Updates and presentations of our framework can be found on our team website http://team.inria.fr/stars. Detailed tips for users are given on our Wiki website http://wiki.inria.fr/stars and sources are hosted thanks to the new Source Control Management tool.

6.21. Qualitative Evaluation of Detection and Tracking Performance

Participants: Swaminathan Sankaranarayanan, François Brémond.

We study an evaluation approach for detection and tracking systems. Given an algorithm that detects people and simultaneously tracks them, we evaluate its output by considering the complexity of the input scene. Some videos used for the evaluation are recorded using the Kinect sensor which provides for an automated ground truth acquisition system. To analyse the algorithm performance, a number of reasons due to which an algorithm might fail is investigated and quantified over the entire video sequence. A set of features called Scene Complexity measures are obtained for each input frame. The variability in the algorithm performance is modelled by these complexity measures using a polynomial regression model. From the regression statistics, we show that we can compare the performance of two different algorithms and also quantify the relative influence of the scene complexity measures on a given algorithm. This work has been published in [44].

6.22. Model-Driven Engineering and Video-surveillance

Participants: Sabine Moisan, Jean-Paul Rigault, Luis-Emiliano Sanchez.

keywords: Feature Model Optimization, Software Metrics, Requirement specification, Component-based system, Dynamic Adaptive Systems, Model-Driven Engineering, Heuristic Search, Constraint Satisfaction Problems

The domain of video surveillance (VS) offers an ideal training ground for Software Engineering studies, because of the huge variability in both the surveillance tasks and the video analysis algorithms [41]. The various VS tasks (counting, intrusion detection, tracking, scenario recognition) have different requirements. Observation conditions, objects of interest, device configuration... may vary from one application to another. On the implementation side, selecting the components themselves, assembling them, and tuning their parameters to comply with context may lead to a multitude of variants. Moreover, the context is not fixed, it evolves dynamically and requires run time adaptation of the component assembly.

Our work relies on Feature Models, a well-known formalism to represent variability in software systems. This year we have focused on an architecture for run time adaptation and on metrics to drive dynamic architecture changes.

6.22.1. Run Time Adaptation Architecture

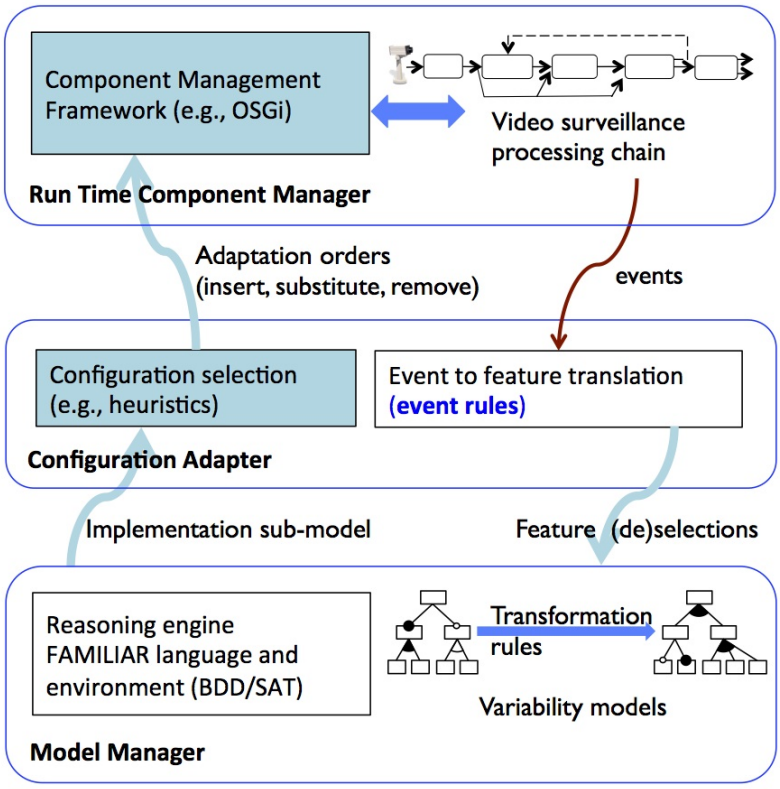


Figure 35. Architecture of an adaptive system using models at run time. (In light blue the elements studied this year.)

The architecture of the run time system (also used for initialization at deployment time) is based on three collaborating modules as shown in Figure 35 . A *Run Time Component Manager* (RTCM) cooperates with the low levels (to manage the software components and capture events) and applies configuration changes. A *Configuration Adapter* (CA) receives events from the RTCM, and propagates them as features into the models to obtain a new configuration. The *Model Manager* (MM) embeds a specialized scripting language for Feature Models (FAMILIAR [52], [53][1]) to manage the representation of the two specialized feature models and applies constraints and model transformations on them. The *Model Manager* produces new component configurations (a model specialization) that it sends to the CA. At its turn, the CA selects one single configuration (possibly using heuristics) and converts it into component operations to be applied by the RTCM.

This year we first finalized the interface between the Model Manager and the Configuration Adapter. On one hand, we transform the feature models obtained from FAMILIAR into C++ representations enriched with software component information. On the other hand, we dynamically transform context change events into requests to FAMILIAR.

Second, we searched for a suitable technology for handling components in the Run Time Component Manager. OSGi is an adequate de facto standard but it is mainly available in the Java world. However we could find a C++ implementation, complete enough for our needs (SOF, Service Oriented Framework [65]). However, SOF has to be completed to adjust to the needs of our end users who are the video system developers. Thus, we are currently building a multi-threaded service layer on top of SOF, easy to use and hiding most of the nitty-gritty technical details of thread programming and SOF component manipulation. This layer provides end users with a set of simple patterns and allow them to concentrate only on the code of video services (such as acquisition, segmentation, tracking...).

As a matter of feasability study we are building an experimental video self adaptive system based on the afore mentionned architecture. Software components are implemented with the OpenCV library. In the final system, feature models and software components continuously interact in real time, modifying the whole system in response to changes in its environment.

6.22.2. Metrics on Feature Models to Optimize Configuration Adaptation at Run Time

As shown on figure 35 , the Configuration Adapter has to set up a suitable component configuration of the run time system. For this, each time the context changes, it receives a set of valid configurations (a feature *sub*-model) from the Model Manager. In most cases, this set contains more than one configuration. Of course, only one configuration can be applied at a given time and the problem is to select the “best” one. Here, “best” is a trade-off between several non-functional aspects: performance, quality of service, time cost for replacing the current configuration, etc.

It is thus necessary to rank the configurations. Our approach is to define metrics suitable for comparing configurations. Then the problem comes down to the widely studied problem of *Feature model optimization* [55]. This problem is known to be an intractable combinatorial optimization problem in general.

We started with a study of the state of the art: metrics for general graphs as well specific to feature models, optimization and requirement specification on feature models... We obtained a structured catalog of quality and feature model metrics. Then we selected solutions based on heuristic search algorithms using quality and feature model metrics. We thus propose several strategies and heuristics offering different properties regarding optimality of results and execution efficiency.

These strategies and heuristics have been implemented, tested, and analyzed using random generated feature models. We got empirical measures about their properties, such as completeness, optimality, time and memory efficiency, scalability... This allows us to compare the performance of the different algorithms and heuristics, and to combine them in order to achieve a good trade-off between optimality and efficiency. Finally, the proposed algorithms have been introduced as part of the Configuration Adapter module.

[1] FAMILIAR has been developed at the I3S laboratory by the Modalis team.

This work is quite original from several aspects. First, we did not find any study using heuristic search algorithms for solving the feature optimization problem. Most studies apply Artificial Intelligence techniques such as CSP solvers, planning agents, genetic algorithms... Second, we do not restrict to the optimization of linear objective functions, but we also address non-linear ones allowing us to take into account a broader set of criteria. Among the possible criteria we consider quality of service of components, their performance, their set up delay, the cost of their replacement, etc. Finally, we apply our metrics at run time whereas most studies consider metrics only for static analysis of feature models.

Currently, we are still working on new variants of the search algorithms and new heuristics relying on techniques proposed in the domains of heuristic search and constraint satisfaction problems.

6.23. Synchronous Modelling and Activity Recognition

Participants: Annie Ressouche, Sabine Moisan, Jean-Paul Rigault, Daniel Gaffé.

6.23.1. Scenario Analysis Module (SAM)

To generate activity recognition systems we supply a scenario analysis module (SAM) to express and recognize complex events from primitive events generated by SUP or other sensors. In this framework, this year we focus on recognition algorithm improvement in order to face the problem of large number of scenario instances recognition.

The purpose of this research axis is to offer a generic tool to express and recognize activities. Genericity means that the tool should accommodate any kind of activities and be easily specialized for a particular framework. In practice, we propose a concrete language to specify activities in the form of a set of scenarios with temporal constraints between scenarios. This language allows domain experts to describe their own scenario models. To recognize instances of these models, we consider the activity descriptions as synchronous reactive systems [76] and we adapt usual techniques of synchronous modelling approach to express scenario behaviours. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario model.

In addition, we have completed SAM in order to address the life cycle of scenario instances. For a given scenario model there may exist several (possibly many) instances at different evolution states. These instances are created and deleted dynamically, according to the input event flow. The challenge is to manage the creation/destruction of this large set of scenario instances efficiently (in time and space), to dispatch events to expecting instances, and to make them evolve independently. To face this challenge, we introduced in the generation of the recognition engine, the expected events of the next step. This avoids to run the engine automatically with events that are not relevant for the recognition process. Indeed, we relied on Lustre [66] synchronous language to express the automata semantics of scenario models as Boolean equation systems. This approach was successful and shows that we can consider a synchronous framework to generate validated scenario recognition engines. This year, in order to improve efficiency (and to tackle the real time recognition problem), we begin to rely on CLEM (see section 6.23.2) toolkit to generate such recognition engines. The reason is threefold: (1) CLEM is becoming a mature synchronous programming environment; (2) we can use the CLEM compiler to build our own compiler; (3) CLEM supplies the possibility of using NuSMV [61] model checker, which is more powerful than the Lustre model-checker. Moreover, thanks to CLEM compiler into Boolean equation systems, we can compute the expected events of the next instant on the fly, by propagation of information related to the current instant.

6.23.2. The clem Workflow

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit (see Figure 9) around the language to design, simulate, verify and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality. This year, we mainly work on theoretical aspects of CLEM.

First, synchronous language semantics usually characterizes each output and local signal status (as present or absent) according to input signal status. To reach our goal, we defined a semantics that translates LE programs into equation systems. This semantics bears and grows richer the knowledge about signals and is never in contradiction with previous deduction (this property is called constructiveness). In such an approach, causality turns out to be a scheduling evaluation problem. We need to determine all the partial orders of equation systems and to compute them, we consider a 4-valued algebra to characterize the knowledge of signal status (unknown, present, absent, overknown). Previously, we relied on 4-valued Boolean algebra [19], [20] which defines the negation of unknown as overknown. The advantage of this way is to benefit from Boolean algebras laws to compute equation system solutions. The drawback concerns signal status evaluation which does not correspond to usual interpretation (not unknown = unknown and not overknown = overknown). To avoid this drawback, we study other kinds of algebras well suited to define synchronous languages semantics. In [49], we choose an algebra which is a bilattice and we show that it is well suited to solve our problem. It is a new application of general bilattice theory [64]. But, the algebra we defined is no more a Boolean algebra, but we prove (always in [49]), that the main laws of Boolean algebras hold as distributivity laws, associativity laws, idempotence laws, etc. After compilation, signals have to be projected into Boolean values. Bilattice theory offers an isomorphism between 4-valued status and pair of Boolean.

Second, the algorithm which computes partial orders relies on the computation of two dependency graphs: the upstream (downstream) dependency graph computes the dependencies of each variable of the system starting from the input (output) variables. Inputs (resp. outputs) have date 0 and the algorithm recursively increases the dates of nodes in the upstream (resp downstream) dependencies graph. Hence, the algorithm determines an *earliest* date and a *latest* date for equation system variables. Moreover, we can compute the dates of variables of a global equation system starting from dates already computed for variables which were inputs and outputs in a sub equation system corresponding to a sub program of the global program[2]. This way of compiling is the corner stone of our approach [20]. We defined two approaches to compute all the valid partial orders of equation systems, either applying critical path scheduling technique (CPM) [3] or applying fix point theory: the vector of earliest (resp. latest) dates can be computed as the least fix point of a monotonic increasing function. This year we have proved that we can compute dates either starting from a global equation system or considering equation system where some variables are abstracted (i.e they have no definition) and whose dates have been already computed. To achieve the demonstration, we rely on an algebraic characterization of dates and thanks to uniqueness property of least fix points, we can deduce that the result is the same for a global equation systems as for its abstraction. We are in the process of publishing this result. From an implementation point of view, we use CPM approach to implement our scheduling algorithm since it is more efficient than fix point consideration. Of course both ways yield the same result. Indeed, fix point approach is useful for a theoretical concern.

6.23.3. Multiple Services for Device Adaptive Platform for Scenario Recognition

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 5.1) with a service oriented middleware approach dealing with dynamic evolutions of system infrastructure. The Rainbow team (Nice-Sophia Antipolis University) proposes a component-based adaptive middleware (WComp [85], [84], [68]) to dynamically adapt and recompose assemblies of components. These operations must obey the "usage contract" of components. The existing approaches don't really ensure that this usage contract is not violated during application design. Only a formal analysis of the component behaviour models associated with a well sound modelling of composition operation may guarantee the respect of the usage contract.

The approach we adopted introduces in a main assembly, a synchronous component for each sub assembly connected with a critical component. This additional component implements a behavioural model of the critical component and model checking techniques apply to verify safety properties concerning this critical component. Thus, we consider that the critical component is validated.

[2] these variables are local in the global equation system
[3] http://pmbook.ce.cmu.edu/10_Fundamental_Scheduling_Procedures.html

To define such synchronous component, user can specify a synchronous component per sub assembly corresponding to a concern and compose the synchronous components connected with the same critical component in order to get an only synchronous component. Thus, we supply a *composition under constraints* of synchronous components and we proved that this operation preserves already separately verified properties of synchronous components [79], [78].

The main challenge of this approach is to deal with the possibly very large number of constraints a user must specify. Indeed, each synchronous monitor has to tell how it combines with other, then we get a combinatorial number of constraints with respect to the number of synchronous monitors and inputs of the critical component. To tackle this problem, we replace the effective description of constraints by a generic specification of them in the critical component. But, we must offer a way to express these generic constraints. Then, each synchronous component has a synchronous controller, which is the projection of the generic constraints on its output set. The global synchronous component is the synchronous parallel composition of all basic components and their synchronous controllers. Moreover, according to synchronous parallel composition features, the property preservation result we have still hold.

TEXMEX Project-Team

6. New Results

6.1. Description of multimedia content

6.1.1. Face Recognition

Participants: Thanh Toan Do, Ewa Kijak.

Face recognition is an important tool for many applications like video analysis. We addressed the problem of faces representation and proposed a weighted co-occurrence Histogram of Oriented Gradient as facial representation. The approach was evaluated on two typical face recognition datasets and has shown an improvement of the recognition rate over state of the art methods [31].

6.1.2. Violent scene detection

Participants: Guillaume Gravier, Patrick Gros, Cédric Penet.

Joint work with Technicolor.

We have worked on multimodal detection of violent scenes in Hollywood movies, in collaboration with Technicolor. Two main directions were explored. On the one hand, we investigated different kinds of Bayesian network structure learning algorithms for the fusion of multimodal features [49]. On the other hand, we studied the use of audio words for the detection of violent related events—gunshots, screams and explosions—in the soundtrack, demonstrating the benefit of product quantization and multiple words representations for increased robustness to variability between movies.

6.1.3. Text detection in videos

Participants: Khaoula Elagouni, Pascale Sébillot.

Joint work with Orange Labs.

Texts embedded in videos often provide high level semantic clues that can be used in several applications and services. We thus aim at designing efficient Optical Character Recognition (OCR) systems able to recognize these texts. In 2012, we proposed a novel approach that avoids the difficult step of character segmentation. Using a multi-scale scanning scheme, texts extracted from videos are first represented by sequences of features learnt by a convolutional neural network. The obtained representations fed a connectionist recurrent model, that relies on the combination of a BLSTM and a CTC connectionist classification model, specifically designed to take into account dependencies between successive learnt features and to recognize texts. The proposed video OCR, evaluated on a database of TV news videos, achieves very high recognition rates (character recognition rate: 97%; word recognition rate: 87%). Experiments also demonstrate that, for our recognition task, learnt feature representations perform better than standard hand-crafted features ([34]). We also carried out a comparison between two of our previous text recognition methods, one relying on a character segmentation step, the other one avoiding it by using a graph model, both on natural scene texts and embedded texts, highlighting the advantages and the limits of each of them. This work is submitted to the journal IJDAR.

6.1.4. Automatic speech recognition

Participants: Guillaume Gravier, Bogdan Ludusan.

This work was partly performed in the context of the Quaero project and the ANR project Attelage de Systèmes Hétérogènes (ANR-09-BLAN-0161-03), in collaboration with the METISS project-team.

In a multimedia context, automatic speech recognition (ASR) provides semantic access to multimedia but faces robustness issues due to the diversity of media sources. To increase robustness, we explore new paradigms for speech recognition based on collaborative decoding and phonetically driven decoding. We investigated mechanisms for the interaction of multiple ASR systems, exchanging linguistic information in a collaborative setting [15]. Following the same idea, we proposed phonetically driven decoding algorithms where the ASR system makes use of phonetic landmarks (place and manner of articulation, stress) to bias and prune the search space [65], [70]. In particular, we proposed a new classification approach to broad phonetic landmark detection [69].

6.2. Large scale indexing and classification techniques

6.2.1. Image retrieval and classification

Participants: Rachid BenMokhtar, Jonathan Delhumeau, Patrick Gros, Mihir Jain, Hervé Jégou, Josip Krapac.

This work was partially done in collaboration with Matthijs Douze and Cordelia Schmid (LEAR), Florent Perronnin and Jorge Sanchez (Xerox), Patrick Pérez (Technicolor) and Ondrej Chum (CVUT Prague). It was partly done in the context of the Quaero project.

Our work on very large scale image search has addressed [14] the joint optimization of three antinomic criterions: speed, memory resources and search quality. We have considered techniques aggregating local image descriptors into a vector and show that the Fisher kernel achieves better performance than the reference bag-of-visual words approach for any given vector dimension. The joint optimization of dimensionality reduction with indexing allowed us to obtain a precise vector comparison as well as a compact representation. The evaluation shows that the image representation can be reduced to a few dozen bytes while preserving high accuracy. Searching a 100 million image dataset takes about 250 ms on one processor core.

This work has been further improved [45] by modifying the way the similarity between images is computed, in particular we have shown that whitening is an effective way to fully exploit multiple vocabularies along with bag-of-visual words and VLAD representations.

We have also considered the problem of image classification, which goal is to produce a semantic representation of the images in the form of text labels reflecting the object categories contained in the images. We have proposed a technique derived from a matching system [44] based on Hamming Embedding and a similarity space mapping. The results outperform the state-of-the-art among matching systems such as NBNN. On some datasets such as Caltech-256, our results compare favorably to the best techniques, namely the Fisher vector representation.

6.2.2. Intensive use of SVM for text mining and image mining

Participants: Thanh Nghi Doan, François Poulet.

Following our previous work on large scale image classification [58], we have developed a fast and efficient framework for large scale image classification. Most of the state of the art approaches use a linear SVM (eg LIBLINEAR) for the training task. Another solution can be to use the new Power Mean SVM (PmSVM) with power mean kernel functions that can solve a binary classification problem with millions of examples and tens of thousands of dense features in a few seconds (excluding the time to read the input files). We are working on a parallel version of this algorithm and trying to deal with unbalanced datasets: in ImageNet1000 dataset, there are 1,000 classes, this is a very unbalanced classification task so we use a balanced bagging parallel algorithm. The time needed to perform the training task on ImageNet1000 was almost 1 day with the original PmSVM algorithm and 2.5 days for LIBLINEAR, we achieve it within 10 min and with a relative precision increase of more than 20%. We are currently working to reduce the RAM needed to perform the task (today 30GB).

6.2.3. Audio indexing

Participants: Jonathan Delhumeau, Guillaume Gravier, Patrick Gros, Hervé Jégou.

This work was done in the context of the Quaero project.

Our new Babaz audio search system [46] aims at finding modified audio segments in large databases of music or video tracks. It is based on an efficient audio feature matching system which exploits the reciprocal nearest neighbors to produce a per-match similarity score. Temporal consistency is taken into account based on the audio matches, and boundary estimation allows the precise localization of the matching segments. The method is mainly intended for video retrieval based on their audio track, as typically evaluated in the copy detection task of Trecvid evaluation campaigns. The evaluation conducted on music retrieval shows that our system is comparable to a reference audio fingerprinting system for music retrieval, and significantly outperforms it on audio-based video retrieval, as shown by our experiments conducted on the dataset used in the copy detection task of the Trecvid'2010 campaign, which was used as an external evaluation in the Quaero project.

6.2.4. Approximate nearest neighbor search with compact codes

Participants: Teddy Furon, Hervé Jégou.

This work was done in collaboration with the Metiss project team (Anthony Bourrier and Rémi Gribonval). It was partly done in the context of the Quaero project.

Following recent works on Hamming Embedding techniques, we proposed [47] a binarization method that aim at addressing the problem of nearest neighbor search for the Euclidean metric by mapping the original vectors into binary vectors ones, which are compact in memory, and for which the distance computation is more efficient. Our method is based on the recent concept of anti-sparse coding, which exhibits here excellent performance for approximate nearest neighbor search. Unlike other binarization schemes, this framework allows, up to a scaling factor, the explicit reconstruction from the binary representation of the original vector. We also show that random projections which are used in Locality Sensitive Hashing algorithms, are significantly outperformed by regular frames for both synthetic and real data if the number of bits exceeds the vector dimensionality, i.e., when high precision is required.

Another aspect we have investigated in this line of research is the problem of efficient nearest neighbor search for arbitrary kernels. For this purpose, we have combined [76] the product quantization technique [4] with explicit embeddings, and showed that this solution significantly outperforms the state-of-the-art technique designed for arbitrary kernels, such as Kernelized Locality Sensitive Hashing. In addition, we have proposed a variant to perform the exact search.

6.2.5. Indexing and searching large image collections with map-reduce

Participants: Laurent Amsaleg, Gylfi Gudmundsson.

This work was done in the context of the Quaero project.

Most researchers working on high-dimensional indexing agree on the following three trends: (i) the size of the multimedia collections to index are now reaching millions if not billions of items, (ii) the computers we use every day now come with multiple cores and (iii) hardware becomes more available, thanks to easier access to Grids and/or Clouds. This work shows how the Map-Reduce paradigm can be applied to indexing algorithms and demonstrates that great scalability can be achieved using Hadoop, a popular Map-Reduce-based framework. Dramatic performance improvements are not however guaranteed a priori: Such frameworks are rigid, they severely constrain the possible access patterns to data and the RAM memory has to be shared. Furthermore, algorithms require major redesign, and may have to settle for sub-optimal behavior. The benefits, however, are numerous: Simplicity for programmers, automatic distribution, fault tolerance, failure detection and automatic re-runs and, last but not least, scalability. We report our experience of adapting a clustering-based high-dimensional indexing algorithm to the Map-Reduce model, and of testing it at large scale with Hadoop as we index 30 billion SIFT descriptors. We draw several lessons from this work that could minimize time, effort and energy invested by other researchers and practitioners working in similar directions.

6.2.6. Vectorization

Participant: Vincent Claveau.

The vectorization principle allows the description of any object in a vector space based on its similarity with pivots objects. During the last years, we have shown that such a technique can be successfully used for Information Retrieval or Topic Segmentation. This year, TexMex has demonstrated how it can be used in a pure data-mining framework by participating to the JRS2012 framework. The task proposed was a high-dimensional and multi-class machine learning problem. Our approach, based on a simple kNN using vectorization has proved its interest, since it was ranked in top-methods while requiring no training phase nor complex setting.

6.3. Security of media

6.3.1. Security of content based image retrieval

Participants: Laurent Amsaleg, Thanh Toan Do, Teddy Furon, Ewa Kijak.

The performance of Content-Based Image Retrieval Systems (CBIRS) is typically evaluated via benchmarking their capacity to match images despite various generic distortions such as cropping, rescaling or Picture in Picture (PiP) attacks, which are the most challenging. Distortions are made in a very generic manner, by applying a set of transformations that are completely independent from the systems later performing recognition tasks. Recently, studies have shown that exploiting the finest details of the various techniques used in a CBIRS offers the opportunity to create distortions that dramatically reduce the recognition performance [30]. Such a *security perspective* is taken in our work. Instead of creating generic PiP distortions, we have proposed a creation scheme able to delude the recognition capabilities of a CBIRS that is representative of state of the art techniques as it relies on SIFT, high-dimensional k-nearest neighbors searches and geometrical robustification steps. We have ran experiments using 100,000 real-world images confirming the effectiveness of these security-oriented PiP visual modifications [29]. This work goes together with the completed PhD of Thanh-Toan Do [8].

6.3.2. The concept of effective key length in watermarking

Participant: Teddy Furon.

Whereas the embedding distortion, the payload and the robustness of digital watermarking schemes are well understood, the notion of security is still not completely well defined. The approach proposed in the last five years is too theoretical and solely considers the embedding process, which is half of the watermarking scheme. In collaboration with Patrick BAS (CNRS, Ecole Centrale de Lille), we propose a new measure of watermarking security. This concept is called the *effective key length*, and it captures the difficulty for the adversary to get access to the watermarking channel: The adversary proposes a test key and the security is measured as the probability that this test key grants him the watermarking channel (he succeeds to decode hidden messages).

This new methodology is applied to the most wide spread watermarking schemes where theoretical and practical computations of the effective key length are proposed: Zero-bit 'Broken Arrows' technique [22], spread spectrum (SS) based schemes (like additive SS, improved SS, and correlation aware SS) [23], and quantization index modulation (QIM) scheme (like Distortion Compensated QIM) [38]. A journal article about this new concept has been submitted to IEEE Trans. on Information Forensics and Security. The keystone of the approach is the evaluation of a security level to the estimation of a probability. Experimental protocols using rare event probability estimator allow good evaluation of this quantity. The soundness of this latter estimator has been theoretically proven in [11] (collaboration with Inria team-project ALEA and ASPI).

6.3.3. A practical joint decoder for active fingerprinting

Participant: Teddy Furon.

This work deals with active fingerprinting, a.k.a. traitor tracing. A robust watermarking technique embeds the user's codeword into the content to be distributed. When a pirated copy of the content is scouted, the watermark decoder extracts the message, which identifies the dishonest user. However, there might exist a group of dishonest users, so called collusion, who mix their personal versions of the content to forge the pirated copy. The extracted message no longer corresponds to the codeword of one user, but is a mix of several codewords. The decoder aims at finding back some of these codewords to identify the colluders, while avoiding accusing innocent users.

This work follows our breakthrough on Tardos code joint decoding, mentioned in last year's activity report, and whose journal version has been published this year in [16]. Information theory proves that a joint decoder computing scores for pairs, triplets, or in general t-tuples of users is more powerful than single decoders working with scores for single users. However, nobody did try them for large scale setups since the number of t-tuples is in $O(n^t)$. In practical scenarios, n is at least 10,000 and t is around 10, which implies the computation of $\sim 10^{40}$ scores. Last year, we were the first team to design an approximate joint decoder. If its complexity was well under control (in $O(n)$), its iterative structure was much intricate.

Our new design of joint decoder is based on the Monte-Carlo Markov Chain method. It is a simpler iterative process allowing us to sample collusion subsets according to the A Posteriori distribution. Then, the probability that user j is guilty is empirically evaluated over this sample, and threshold to yield a reliable decision. This work has been done under a collaboration with Inria team-project ASPI, and published in [39].

6.4. Multimedia content structuring

6.4.1. Motif discovery

Participants: Guillaume Gravier, Hervé Jégou, Anh Phuong Ta, Wanlei Zhao.

This work was done in the context of the Quaero project.

We have pursued our work on unsupervised discovery of repeating motifs in multimedia data along three directions:

- Discovery of multiple recurrent audio-visually consistent sequences: We proposed two unsupervised approaches to automatically detect multiple structural events in videos using audio and visual modalities. Both approaches rely on cross-modal cluster analysis techniques to directly define events from the data without any prior assumption [51], [52].
- Large-scale unsupervised discovery of near-duplicate shots in TV streams: We developed an efficient method with little a priori knowledge which relies on a product k-means quantizer to efficiently produce hash keys adapted to the data distribution of the frame descriptors. This hashing technique combined with a temporal consistency check allows the detection of meaningful repetitions in TV streams [54].
- Audio motif discovery: This joint work with the METISS project-team extends the generic audio motif discovery method developed in the Ph. D. thesis of Armando Muscariello [17]. We developed an efficient implementation, which will be made publicly available. The software was benchmarked on near duplicate audio motif discovery in the framework of the Quaero project.

6.4.2. Stream labeling for TV structuring

Participants: Vincent Claveau, Guillaume Gravier, Patrick Gros, Emmanuelle Martienne, Abir Ncibi.

In this application, we focus on the problem of labeling the segments of a TV stream according to their types (*e.g.,*, programs, commercial breaks, sponsoring, ...). During this year, we performed an in-depth analysis of the use of Conditional Random Fields (CRF) for our task. In particular, we studied:

- how sequentiality is modeled with the CRF;
- the links with other probabilistic graphical techniques (HMM, MEMM...);
- the robustness of the approach when dealing with few training data or few features;

The use of this model for semi-supervised and unsupervised learning are under study. We also studied the use of very simple descriptors (simple shot lengths, and use of global image descriptors only to complete the results) in order to fasten the initial repetition detection stage. This allows us to process 6 months of TV in a few minutes.

6.4.3. Multimedia browsing

Participant: Laurent Amsaleg.

Traditionally, research in multimedia has focused primarily on analyzing and understanding the contents of media documents, by defining clever ways to extract relevant information from the multimedia files, thereby hoping to eventually bridge the semantic gap. We have observed that much of the research in multimedia is trying to *link* the information automatically extracted from the contents to create a meaningful user-experience. Most of the state-of-art solutions are very ad-hoc, and we believe that multimedia is lacking a powerful and flexible data model where multimedia data (ranging from entire documents to elements automatically extracted from the contents such as faces, scenes, objects, ...) can be appropriately represented as well as the relationships between data items. Instead, we propose a multi-dimensional model for media browsing, called ObjectCube, based on the multi-dimensional model commonly used in On-Line Analytical Processing (OLAP) applications. This model has been implemented in a prototype called ObjectCube, and its performance evaluated using personal photo collections of up to one million images. We also worked on exposing plug-in API for image analysis and browsing methods, facilitating the use of the prototype and its model as a demonstration platform.

6.4.4. Video summarization

Participants: Mohamed-Haykel Boukadida, Patrick Gros.

Joint work with Orange labs.

Up to now, most video summarization methods are based on concepts like saliency and often use a single modality. In order to develop a more general framework, we propose to use a constraint programming approach, where summarizing a video is seen as a constraint resolution problem, which consists in choosing certains excerpts with respect to various criteria. This first year of work on the topic was mainly devoted to discover the abilities of Choco, a constraint solver, and to study how summarization can be formulated as a constraint resolution problem.

6.4.5. Graph organization of large scale news archives

Participants: Guillaume Gravier, Ludivine Kuznik, Pascale Sébillot.

This work is done in collaboration with Jean Carrive at Institut National de l'Audiovisuel in the framework of a joint Ph. D. thesis within the Quaero project.

The idea of this work is to automatically create links and threads between reports in several years of broadcast news shows, based either on the documentary records of the shows and/or on the automatic transcripts. We studied how standard information retrieval measures of similarity can be used to build an epsilon-nearest neighbor graph from the various fields of the documentary records. Depending on the field used (title, keywords from a thesaurus, summary, speech transcript) and the metrics, different types of clusters can be obtained in the graph. We proposed metrics mimicking recall and precision on documents to analyze the graphs obtained and quantify the potential interest of various graph construction strategies for topic threading.

6.5. Language processing in multimedia

6.5.1. Lexical-phonetic automata for spoken utterance indexing and retrieval

Participants: Julien Fayolle, Guillaume Gravier, Fabienne Moreau, Christian Raymond.

This work was partly done in the context of the Quaero project.

Spoken content retrieval relies on the fields of automatic speech recognition and information retrieval (IR). However, IR tools made for text are not adapted to automatic transcripts which are particularly incomplete and uncertain. Even if in-vocabulary words are usually well-recognized, these transcripts contain many recognition errors affecting notably out-of-vocabulary words and named entities that convey important discourse information (e.g., person names, localizations, organizations) necessary for IR. This year, we have proposed a method for indexing spoken utterances which combines lexical and phonetic hypotheses in a hybrid index built from automata [35], [36]. The retrieval is performed by a lexical-phonetic and semi-imperfect matching whose aim is to improve the recall. A feature vector, containing edit distance scores and a confidence measure, weights each transition to help the filtering of the candidate utterance list for a more precise search. We have demonstrated the complementarity of the lexical and phonetic levels (extracted from the 1-best speech recognition hypothesis) and the advantage of using a hybrid index, a semi-imperfect matching and a supervised filtering (combining edit distance scores and a confidence measure).

6.5.2. Information extraction and text mining

Participants: Ali Reza Ebadat, Vincent Claveau, Pascale Sébillot.

This work was partly done in the framework of the Quaero project.

In the framework of Ali-Reza Ebadat's thesis on information extraction for multimedia analysis, we have investigated techniques for robust text-mining on texts or speech transcripts. We have developed several supervised models:

- entity detection and entity classification; the goal is to detect, into a text, pre-defined categories of entities and to label them accordingly. The techniques that we developed cascade chunk parsing with simple classification tools, resulting in a very efficient and simple to train NLP tool.
- relation detection; this model relies on k-NN approach with a language-modeling based distance. Since it relies on surface elements, it can handle noisy data such as speech transcripts.

We have also developed unsupervised models for information discovery:

- entity clustering; the goal is to detect and group, without a priori knowledge, entities. We have shown that weighting techniques used in information retrieval can be used as relevant features to describe the entity.
- relation clustering: as for entity, the goal is to group relations (that is, pairs of entities) without a priori or pre-defined categories. Our approach is pioneer is this field and relies on clustering with language-modeling based distances.

Some of these models have been evaluated in the framework of the Quaero evaluation campaign and TexMex ranked first in three of the tracks (entity detection and categorization) and close second in the last one (relation detection and categorization).

6.5.3. Morphological analysis for information retrieval

Participants: Vincent Claveau, Ewa Kijak.

In the biomedical field, the key to access information is the use of specialized terms (like *photochemotherapy*). These complex morphological structures may prevent a user querying for *gastrodynia* to retrieve texts containing *stomachalgia*. In that context, we have developed a new unsupervised technique to identify the various meaningful components of these terms and use this analysis to improve biomedical information retrieval. Our approach combines an automatic alignment using a pivot language, and an analogical learning that allows an accurate morphological analysis of terms. We ave shown that these morphological analyses can be used to greatly improve the indexing of medical documents.

6.5.4. Unsupervised hierarchical topic segmentation

Participants: Guillaume Gravier, Pascale Sébillot, Anca-Roxana Simon.

Linear topic segmentation has been widely studied for textual data and recently adapted to spoken contents. However, most documents exhibit a hierarchy of topics which cannot be recovered using linear segmentation. We investigated hierarchical topic segmentation of TV programs exploiting the spoken material. Recursively applying linear segmentation methods is one solution but fails at the lowest levels of the hierarchy when small segments are targeted, in particular when transcription errors jeopardize lexical cohesion. To skirt these issues, we investigated the use of indirect comparison between segments via vectorization techniques at the lower level of the hierarchy, using simple segmentation methods based on TextTiling. Results were similar to those obtained by the recursive use of a more elaborate probabilistic topic segmentation method. Future work will focus on using indirect comparison within the probabilistic framework.

6.6. Competitions and international evaluation campaign

6.6.1. Mediaeval's affect task: Violent scenes detection task

Participants: Guillaume Gravier, Patrick Gros, Cédric Penet.

The project-team participated in the Affect Task of the MediaEval 2012 benchmark, both as part of the organizing team and as competitor [64], [67].

6.6.2. Mediaeval's placing task: Geo-localization of videos

Participants: Jonathan Delhumeau, Guillaume Gravier, Hervé Jégou, Michele Trevisiol.

This work was partly done in the context of the Quaero project.

We developed an efficient and effective system to identify the geographic location of videos using a multimodal cascade of techniques exploiting all available sources of information, from user assigned tags to user data and image content. We also proposed a novel hierarchical strategy to exploit tags using information retrieval techniques. A coarse geographic area is first identified before refining the search to find exact geo-coordinates. Area and coordinates are obtained from a vector space representation of the tags using appropriate weighting and normalization [68].

We participated in the Placing Task of the MediaEval 2012 benchmark, where we ranked first on one of the mandatory runs (no gazeteers, no dictionary).

6.6.3. Mediaeval: Search & hyperlinking

Participants: Guillaume Gravier, Camille Guinaudeau, Pascale Sébillot.

We participated in the Search and Hyperlinking task proposed in the framework of the MediaEval benchmark initiative in 2012. We developed a solution for the hyperlinking subtask in which participants were required to return a ranked list of video segments potentially relevant to the answer provided for an initial query, thus creating links between video segments.

Our solution, based on information retrieval techniques, implements two separate module: The retrieval of relevant videos, followed by the selection of short segments specifically corresponding to the information need. First, the hyperlinking module computes the similarity between a video segment query and the collection of videos and returns a ranked list of relevant videos. We investigated several parameterization and ranking strategies. In the second step, we extract from each video the segment that is the closest, from a meaning point of view, to the video segment query, using topic segmentation methods [42].

Our system ranked either first or second depending on the evaluation conditions.

6.6.4. ETAPE named entities evaluation campaign

Participant: Christian Raymond.

Christian Raymond participated to the ETAPE Named Entities evaluation campaign. The goal was to propose a system able to tag NE following the new tree-stuctured NE definition given in the Quaero project. The evaluation has been done on manual and 5 automatic transcriptions of french TV and Radio shows produced by 5 different automatic speech recognition systems. The system was ranked first with results far better than those of the other participating systems.

6.6.5. DEFT evaluation campaign participation

Participants: Vincent Claveau, Christian Raymond.

Christian Raymond and Vincent Claveau participated to DEFT. The task proposed was to work on a corpus of scientific papers, by focusing the work on the issue of indexing the scientific papers: identifying the keywords chosen by the authors to index their paper, considering both abstract and whole article. Two tasks were proposed which led them to test two different strategies . For the first task, a list of keywords was provided. Based on that, our first strategy is to consider that as an Information Retrieval problem in which the keywords are the queries that are attributed to the best ranked documents. This approach yielded very good results. For the second task, only the articles were known. For this task, our approach is mainly based on a term extraction system whose results are reordered by a machine learning [27] technique.

6.6.6. Trecvid: Multimedia Indexing task

Participants: Jonathan Delhumeau, Philippe-Henri Gosselin, Hervé Jégou.

This work was partly done in the context of the Quaero project.

Texmex has taking part to the Quaero [50] and IRIM [21] submissions of Trecvid in the Multimedia indexing task, by providing some state-of-the-art image descriptors and collaborating with the LIG to set up the dimensionality reduction tool for high-dimensional vectors. The Quaero Rank was ranked 3rd in the full task (1st amongst European submissions).

WILLOW Project-Team

6. New Results

6.1. 3D object and scene modeling, analysis, and retrieval

6.1.1. People Watching: Human Actions as a Cue for Single View Geometry

Participants: Vincent Delaitre, Ivan Laptev, Josef Sivic, Alexei Efros [CMU], David Fouhey [CMU], Abhinav Gupta [CMU].

We present an approach which exploits the coupling between human actions and scene geometry. We investigate the use of human pose as a cue for single-view 3D scene understanding. Our method builds upon recent advances in still-image pose estimation to extract functional and geometric constraints about the scene. These constraints are then used to improve state-of-the-art single-view 3D scene understanding approaches. The proposed method is validated on a collection of monocular time lapse sequences collected from YouTube and a dataset of still images of indoor scenes. We demonstrate that observing people performing different actions can significantly improve estimates of 3D scene geometry.

This work has been published in [11].

6.1.2. Learning and Calibrating Per-Location Classifiers for Visual Place Recognition

Participants: Petr Gronát, Josef Sivic, Guillaume Obozinski [Inria SIERRA], Tomáš Pajdla [CTU in Prague].

The aim of this work is to localize a query photograph by finding other images depicting the same place in a large geotagged image database. This is a challenging task due to changes in viewpoint, imaging conditions and the large size of the image database. The contribution of this work is two-fold. First, we cast the place recognition problem as a classification task and use the available geotags to train a classifier for each location in the database in a similar manner to per-exemplar SVMs in object recognition. Second, as only few positive training examples are available for each location, we propose a new approach to calibrate all the per-location SVM classifiers using *only* the negative examples. The calibration we propose relies on a significance measure essentially equivalent to the p-values classically used in statistical hypothesis testing. Experiments are performed on a database of 25,000 geotagged street view images of Pittsburgh and demonstrate improved place recognition accuracy of the proposed approach over the previous work. The problem addressed in this work is illustrated in Figure 1 .

This work has been submitted to CVPR 2013.

6.1.3. What Makes Paris Look like Paris?

Participants: Josef Sivic, Carl Doersch [CMU], Saurabh Singh [UIUC], Abhinav Gupta [CMU], Alexei Efros [CMU].

Given a large repository of geotagged imagery, we seek to automatically find visual elements, e.g. windows, balconies, and street signs, that are most distinctive for a certain geo-spatial area, for example the city of Paris. This is a tremendously difficult task as the visual features distinguishing architectural elements of different places can be very subtle. In addition, we face a hard search problem: given all possible patches in all images, which of them are both frequently occurring and geographically informative? To address these issues, we propose to use a discriminative clustering approach able to take into account the weak geographic supervision. We show that geographically representative image elements can be discovered automatically from Google Street View imagery in a discriminative manner. We demonstrate that these elements are visually interpretable and perceptually geo-informative. The discovered visual elements can also support a variety of computational geography tasks, such as mapping architectural correspondences and influences within and across cities, finding representative elements at different geo-spatial scales, and geographically-informed image retrieval. Example result is shown in Figure 2 .

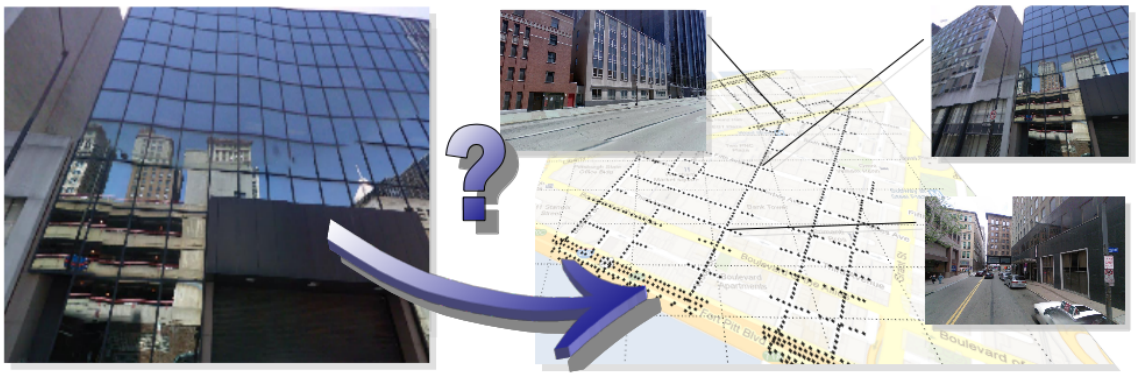


Figure 1. The goal of this work is to localize a query photograph (left) by finding other images of the same place in a large geotagged image database (right). We cast the problem as a classification task and learn a classifier for each location in the database. We develop a non-parametric procedure to calibrate the outputs of the large number of per-location classifiers without the need for additional positive training data.

Figure 2. Examples of geographic patterns in Paris (shown as red dots on the maps) for three discovered visual elements (shown below each map). Balconies with cast-iron railings are concentrated on the main boulevards (left). Windows with railings mostly occur on smaller streets (middle). Arch supporting columns are concentrated on Place des Vosges and the St. Germain market (right).

This work has been published in [6].

6.2. Category-level object and scene recognition

6.2.1. Task-Driven Dictionary Learning

Participants: Jean Ponce, Julien Mairal [Inria LEAR], Francis Bach [Inria SIERRA].

Modeling data with linear combinations of a few elements from a learned dictionary has been the focus of much recent research in machine learning, neuroscience and signal processing. For signals such as natural images that admit such sparse representations, it is now well established that these models are well suited to restoration tasks. In this context, learning the dictionary amounts to solving a large-scale matrix factorization problem, which can be done efficiently with classical optimization tools. The same approach has also been used for learning features from data for other purposes, e.g., image classification, but tuning the dictionary in a supervised way for these tasks has proven to be more difficult. In this paper, we present a general formulation for supervised dictionary learning adapted to a wide variety of tasks, and present an efficient algorithm for solving the corresponding optimization problem. Experiments on handwritten digit classification, digital art identification, nonlinear inverse image problems, and compressed sensing demonstrate that our approach is effective in large-scale settings, and is well suited to supervised and semi-supervised classification, as well as regression tasks for data that admit sparse representations.

This work has been published in [7].

6.2.2. Object Detection Using Strongly-Supervised Deformable Part Models

Participants: Ivan Laptev, Hossein Azizpour [KTH].

Deformable part-based models achieve state-of-the-art performance for object detection, but rely on heuristic initialization during training due to the optimization of non-convex cost function. This work investigates limitations of such an initialization and extends earlier methods using additional supervision. We explore strong supervision in terms of annotated object parts and use it to (i) improve model initialization, (ii) optimize model structure, and (iii) handle partial occlusions. Our method is able to deal with sub-optimal and incomplete annotations of object parts and is shown to benefit from semi-supervised learning setups where part-level annotation is provided for a fraction of positive examples only. Experimental results are reported for the detection of six animal classes in PASCAL VOC 2007 and 2010 datasets. We demonstrate significant improvements in detection performance compared to the LSVM and the Poselet object detectors.

This work has been published in [9].

6.2.3. Multi-Class Cosegmentation

Participants: Armand Joulin, Jean Ponce, Francis Bach [Inria SIERRA].

Bottom-up, fully unsupervised segmentation remains a daunting challenge for computer vision. In the cosegmentation context, on the other hand, the availability of multiple images assumed to contain instances of the same object classes provides a weak form of supervision that can be exploited by discriminative approaches. Unfortunately, most existing algorithms are limited to a very small number of images and/or object classes (typically two of each). This work proposes a novel energy-minimization approach to cosegmentation that can handle multiple classes and a significantly larger number of images. The proposed cost function combines spectral- and discriminative-clustering terms, and it admits a probabilistic interpretation. It is optimized using an efficient EM method, initialized using a convex quadratic approximation of the energy. Comparative experiments show that the proposed approach matches or improves the state of the art on several standard datasets.

This work has been published in [13].

6.2.4. A Convex Relaxation for Weakly Supervised Classifiers

Participants: Armand Joulin, Francis Bach [Inria SIERRA].

This work introduces a general multi-class approach to weakly supervised classification. Inferring the labels and learning the parameters of the model is usually done jointly through a block-coordinate descent algorithm such as expectation-maximization (EM), which may lead to local minima. To avoid this problem, we propose a cost function based on a convex relaxation of the soft-max loss. We then propose an algorithm specifically designed to efficiently solve the corresponding semidefinite program (SDP). Empirically, our method compares favorably to standard ones on different datasets for multiple instance learning and semi-supervised learning, as well as on clustering tasks.

This work has been published in [12].

6.2.5. Top-Down and Bottom-Up Cues for Scene Text Recognition

Participants: Karteek Alahari, Anand Mishra [IIT India], C.V. Jawahar [IIT India].

Scene text recognition has gained significant attention from the computer vision community in recent years. Recognizing such text is a challenging problem, even more so than the recognition of scanned documents. In this work, we focus on the problem of recognizing text extracted from street images. We present a framework that exploits both bottom-up and top-down cues. The bottom-up cues are derived from individual character detections from the image. We build a Conditional Random Field model on these detections to jointly model the strength of the detections and the interactions between them. We impose top-down cues obtained from a lexicon-based prior, i.e. language statistics, on the model. The optimal word represented by the text image is obtained by minimizing the energy function corresponding to the random field model.

We show significant improvements in accuracies on two challenging public datasets, namely Street View Text (over 15%) and ICDAR 2003 (nearly 10%).

This work has been published in [15].

6.2.6. Scene Text Recognition using Higher Order Language Priors

Participants: Karteek Alahari, Anand Mishra [IIT India], C.V. Jawahar [IIT India].

The problem of recognizing text in images taken in the wild has gained significant attention from the computer vision community in recent years. Contrary to recognition of printed documents, recognizing scene text is a challenging problem. We focus on the problem of recognizing text extracted from natural scene images and the web. Significant attempts have been made to address this problem in the recent past. However, many of these works benefit from the availability of strong context, which naturally limits their applicability. In this work we present a framework that uses a higher order prior computed from an English dictionary to recognize a word, which may or may not be a part of the dictionary. We show experimental results on publicly available datasets. Furthermore, we introduce a large challenging word dataset with five thousand words to evaluate various steps of our method exhaustively.

The main contributions of this work are: (1) We present a framework, which incorporates higher order statistical language models to recognize words in an unconstrained manner (i.e. we overcome the need for restricted word lists, and instead use an English dictionary to compute the priors). (2) We achieve significant improvement (more than 20%) in word recognition accuracies without using a restricted word list. (3) We introduce a large word recognition dataset (at least 5 times larger than other public datasets) with character level annotation and benchmark it.

This work has been published in [14].

6.3. Image restoration, manipulation and enhancement

6.3.1. Non-Uniform Deblurring for Shaken Images

Participants: Josef Sivic, Andrew Zisserman, Jean Ponce, Oliver Whyte [Microsoft Redmond].

Photographs taken in low-light conditions are often blurry as a result of camera shake, i.e. a motion of the camera while its shutter is open. Most existing deblurring methods model the observed blurry image as the convolution of a sharp image with a uniform blur kernel. However, we show that blur from camera shake is in general mostly due to the 3D rotation of the camera, resulting in a blur that can be significantly non-uniform across the image. We propose a new parametrized geometric model of the blurring process in terms of the rotational motion of the camera during exposure. This model is able to capture non-uniform blur in an image due to camera shake using a single global descriptor, and can be substituted into existing deblurring algorithms with only small modifications. To demonstrate its effectiveness, we apply this model to two deblurring problems; first, the case where a single blurry image is available, for which we examine both an approximate marginalization approach and a maximum a posteriori approach, and second, the case where a sharp but noisy image of the scene is available in addition to the blurry image. We show that our approach makes it possible to model and remove a wider class of blurs than previous approaches, including uniform blur as a special case, and demonstrate its effectiveness with experiments on synthetic and real images.

This work has been published in [8]. An image deblurring demo, described in section 5.8 , has been made available online.

6.3.2. Learning to Estimate and Remove Non-uniform Image Blur

Participants: Florent Couzinie-Devy, Jian Sun, Karteek Alahari, Jean Ponce.

This work addresses the problem of restoring images subjected to unknown and spatially varying blur caused by defocus or linear (say, horizontal) motion. The estimation of the global (non-uniform) image blur is cast as a multi-label energy minimization problem. The energy is the sum of unary terms corresponding to learned local blur estimators, and binary ones corresponding to blur smoothness. Its global minimum is found using Ishikawa's method by exploiting the natural order of discretized blur values for linear motions and defocus. Once the blur has been estimated, the image is restored using a robust (non-uniform) deblurring algorithm based on sparse regularization with global image statistics. The proposed algorithm outputs both a segmentation of the image into uniform-blur layers and an estimate of the corresponding sharp image. We present qualitative results on real images, and use synthetic data to quantitatively compare our approach to the publicly available implementation of Chakrabarti et al. 2010.

This work has been submitted to CVPR 2013.

6.4. Human activity capture and classification

6.4.1. Scene Semantics from Long-Term Observation of People

Participants: Vincent Delaitre, Ivan Laptev, Josef Sivic, David Fouhey [CMU], Abhinav Gupta [CMU], Alexei Efros [CMU].

Our everyday objects support various tasks and can be used by people for different purposes. While object classification is a widely studied topic in computer vision, recognition of object function, i.e., what people can do with an object and how they do it, is rarely addressed. In this work we construct a functional object description with the aim to recognize objects by the way people interact with them. We describe scene objects (sofas, tables, chairs) by associated human poses and object appearance. Our model is learned discriminatively from automatically estimated body poses in many realistic scenes. In particular, we make use of time-lapse videos from YouTube providing a rich source of common human-object interactions and minimizing the effort of manual object annotation. We show how the models learned from human observations significantly improve object recognition and enable prediction of characteristic human poses in new scenes. Results are shown on a dataset of more than 400,000 frames obtained from 146 time-lapse videos of challenging and realistic indoor scenes. Some of the estimated human poses and results of pixel-wise scene segmentation are shown in Figure 3 .

This work has been published in [10].

Figure 3. *Top: Example of particular pose detections in three indoor scenes. Bottom: object segmentation illustrated by original images, ground truth segmentation, and automatic segmentation by our method shown in the left, middle and right columns respectively.*

6.4.2. Analysis of Crowded Scenes in Video

Participants: Ivan Laptev, Josef Sivic, Mikel Rodriguez [MITRE].

In this work we first review the recent studies that have begun to address the various challenges associated with the analysis of crowded scenes. Next, we describe our two recent contributions to crowd analysis in video. First, we present a crowd analysis algorithm powered by prior probability distributions over behaviors that are learned on a large database of crowd videos gathered from the Internet. The proposed algorithm performs like state-of-the-art methods for tracking people having common crowd behaviors and outperforms the methods when the tracked individuals behave in an unusual way. Second, we address the problem of detecting and tracking a person in crowded video scenes. We formulate person detection as the optimization of a joint energy function combining crowd density estimation and the localization of individual people. The proposed methods are validated on a challenging video dataset of crowded scenes. Finally, the chapter concludes by describing ongoing and future research directions in crowd analysis.

This work is to appear in [17].

6.4.3. Actlets: A Novel Local Representation for Human Action Recognition in Video

Participants: Muhammad Muneeb Ullah, Ivan Laptev.

This work addresses the problem of human action recognition in realistic videos. We follow the recently successful local approaches and represent videos by means of local motion descriptors. To overcome the huge variability of human actions in motion and appearance, we propose a supervised approach to learn local motion descriptors – *actlets* – from a large pool of annotated video data. The main motivation behind our method is to construct action-characteristic representations of body joints undergoing specific motion patterns while learning invariance with respect to changes in camera views, lighting, human clothing, and other factors. We avoid the prohibitive cost of manual supervision and show how to learn actlets automatically from synthetic videos of avatars driven by the motion-capture data. We evaluate our method and show its significant improvement as well as its complementarity to existing techniques on the challenging UCF-sports and YouTube-actions datasets.

This work has been published in [16].

6.4.4. Layered Segmentation of People in Stereoscopic Movies

Participants: Karteek Alahari, Guillaume Seguin, Josef Sivic, Ivan Laptev.

In this work we seek to obtain a layered pixel-wise segmentation of multiple people in a stereoscopic video. This involves challenges such as dealing with unconstrained stereoscopic video, non-stationary cameras, complex indoor and outdoor dynamic scenes. The contributions of our work are three-fold: First, we develop a layered segmentation model incorporating person detections and pose estimates, as well as colour, motion, and stereo disparity cues. The model also explicitly represents depth ordering and occlusions of people. Second, we introduce a stereoscopic dataset with frames extracted from feature length movies “StreetDance 3D" and “Pina". In addition to realistic stereo image data, it contains nearly 700 annotated poses, 1200 annotated detections, and 400 pixel-wise segmentations of people. Third, we evaluate the benefits of stereo signal for person detection, pose estimation and segmentation in the new dataset. We demonstrate results on challenging realistic indoor and outdoor scenes depicting multiple people with frequent occlusions. Example result is shown in Figure 4 .

This work has been submitted to CVPR 2013.

6.4.5. Highly-Efficient Video Features for Action Recognition and Counting

Participants: Vadim Kantorov, Ivan Laptev.

Figure 4. A sample frame extracted from the stereoscopic movie "StreetDance": From left to right – left image from the stereo pair, disparity map computed from the stereo pair, layered segmentation of the image into 7 people. The front to back ordering is shown as a colour map, where "blue" denotes front and "red" denotes back. The cost function associated with our model is initialized using person detections, and incorporates disparity, pose, colour and motion cues. Note that the result shows accurate segmentation boundaries and also a reliable layer ordering of people.

Local video features provide state-of-the-art performance for action recognition. While the accuracy of action recognition has been steadily improved over the recent years, the low speed of feature extraction remains to be a major bottleneck preventing current methods from addressing large-scale applications. In this work we demonstrate that local video features can be computed very efficiently by exploiting motion information readily-available from standard video compression schemes. We show experimentally that the use of sparse motion vectors provided by the video compression improves the speed of existing optical-flow based methods by two orders of magnitude while resulting in limited drops of recognition performance. Building on this representation, we next address the problem of event counting in video and present a method providing accurate counts of human actions and enabling to process 100 years of video on a modest computer cluster.

This work has been submitted to CVPR 2013.