



RESEARCH CENTER  
**Rennes - Bretagne-Atlantique**

FIELD

Activity Report 2013

# Section Application Domains

Edition: 2014-03-19



ALGORITHMICS, PROGRAMMING, SOFTWARE AND ARCHITECTURE

1. ALF Project-Team ..... 4  
 2. CAIRN Project-Team ..... 5  
 3. CELTIQUE Project-Team (section vide) ..... 6  
 4. ESPRESSO Project-Team ..... 7  
 5. Hycomes Team (section vide) ..... 8  
 6. S4 Project-Team ..... 9  
 7. SUMO Team ..... 10  
 8. TASC Project-Team ..... 12

APPLIED MATHEMATICS, COMPUTATION AND SIMULATION

9. ASPI Project-Team ..... 13  
 10. I4S Project-Team (section vide) ..... 14  
 11. IPSO Project-Team ..... 15

DIGITAL HEALTH, BIOLOGY AND EARTH

12. DYLISS Project-Team ..... 17  
 13. FLUMINANCE Project-Team ..... 19  
 14. GENSCALE Project-Team ..... 20  
 15. SAGE Project-Team ..... 21  
 16. SERPICO Project-Team ..... 22  
 17. VISAGES Project-Team ..... 26

NETWORKS, SYSTEMS AND SERVICES, DISTRIBUTED COMPUTING

18. ACES Project-Team (section vide) ..... 28  
 19. ASAP Project-Team ..... 29  
 20. ASCOLA Project-Team ..... 30  
 21. ATLANMOD Project-Team ..... 32  
 22. CIDRE Project-Team ..... 33  
 23. DIONYSOS Project-Team ..... 34  
 24. KERDATA Project-Team ..... 35  
 25. MYRIADS Project-Team ..... 37  
 26. TRISKELL Project-Team ..... 38

PERCEPTION, COGNITION AND INTERACTION

27. DREAM Project-Team ..... 39  
 28. HYBRID Project-Team ..... 41  
 29. LAGADIC Project-Team ..... 42  
 30. MIMETIC Project-Team ..... 43  
 31. PANAMA Project-Team ..... 47  
 32. SIROCCO Project-Team ..... 49  
 33. TEXMEX Project-Team ..... 52

## **ALF Project-Team**

# **4. Application Domains**

## **4.1. Any computer usage**

The ALF team is working on the fundamental technologies for computer science: processor architecture and performance-oriented compilation. The research results have impacts on any application domain that requires high performance executions (telecommunication, multimedia, biology, health, engineering, environment ...), but also on many embedded applications that exhibit other constraints such as power consumption, code size and guaranteed response time. Our research activity implies the development of software prototypes.

## CAIRN Project-Team

# 4. Application Domains

## 4.1. Panorama

**keywords:** telecommunications, wireless communications, wireless sensor networks, content-based image retrieval, video coding, intelligent transportation systems, automotive, security

Our research is based on realistic applications, in order to both discover the main needs created by these applications and to invent realistic and interesting solutions.

The high complexity of the **Next-Generation (4G) Wireless Communication Systems** leads to the design of real-time high-performance specific architectures. The study of these techniques is one of the main field of applications for our research, based on our experience on WCDMA for 3G implementation.

In **Wireless Sensor Networks (WSN)**, where each wireless node has to operate without battery replacement for a long time, energy consumption is the most important constraint. In this domain, we mainly study energy-efficient architectures and wireless cooperative techniques for WSN.

**Intelligent Transportation Systems (ITS)**, and especially Automotive Systems, more and more apply technology advances. While wireless transmissions allow a car to communicate with another or even with road infrastructure, **automotive industry** can also propose driver assistance and more secure vehicles thanks to improvements in computation accuracy for embedded systems.

Other important fields will also be considered: hardware cryptographic and security modules, specialized hardware systems for the filtering of the network traffic at high-speed, high-speed true-random number generation for security, content-based image retrieval and video processing.

## 4.2. 4G Wireless Communication Systems

With the advent of the next generation (4G) broadband wireless communications, the combination of MIMO (Multiple-Input Multiple-Output) wireless technology with Multi-Carrier CDMA (MC-CDMA) has been recognized as one of the most promising techniques to support high data rate and high performance. Moreover, future mobile devices will have to propose interoperability between wireless communication standards (4G, WiMax ...) and then implement MIMO pre-coding, already used by WiMax standard. Finally, in order to maximize mobile devices lifetime and guarantee quality of services to consumers, 4G systems will certainly use cooperative MIMO schemes or MIMO relays. Our research activity focuses on MIMO pre-coding and MIMO cooperative communications with the aim of algorithmic optimization and implementation prototyping.

## 4.3. Wireless Sensor Networks

Sensor networks are a very dynamic domain of research due, on the one hand, to the opportunity to develop innovative applications that are linked to a specific environment, and on the other hand to the challenge of designing totally autonomous communicating objects. Cross-layer optimizations lead to energy-efficient architectures and cooperative techniques dedicated to sensor networks applications. In particular, cooperative MIMO techniques are used to decrease the energy consumption of the communications.

## 4.4. Multimedia processing

In multimedia applications, audio and video processing is the major challenge embedded systems have to face. It is computationally intensive with power requirements to meet. Video or image processing at pixel level, like image filtering, edge detection and pixel correlation or at block-level such as transforms, quantization, entropy coding and motion estimation have to be accelerated. We investigate the potential of reconfigurable architectures for the design of efficient and flexible accelerators in the context of multimedia applications.

**CELTIQUE Project-Team (section vide)**

---

## ESPRESSO Project-Team

# 4. Application Domains

## 4.1. Embedded systems

The application domains covered by the Polychrony toolbox are engineering areas where a system design-flow requires high-level model transformations and verifications to be applied during the development-cycle. The project-team has focused on developing such integrated design methods in the context of avionics applications, through the European IST projects Sacres, Syrf, Safeair, Speeds, and through the national ANR projects Topcased, OpenEmbeDD, Spacify. In this context, Polychrony is seen as a platform on which the architecture of an embedded system can be specified from the earliest design stages until the late deployment stages through a number of formally verifiable design refinements.

Along the way, the project adopted the policy proposed with project Topcased and continued with OpenEmbeDD to make its developments available to a large community in open-source. The Polychrony environment is now integrated in the OPEES/Polarsys platform and distributed under EPL and GPL v2.0 license for the benefits of a growing community of users and contributors, among which the most active are Virginia Tech's Fermat laboratory and Inria's project-teams Aoste, Dart.

**Hycomes Team (section vide)**



## **S4 Project-Team**

# **4. Application Domains**

## **4.1. Domain 1**

## SUMO Team

# 4. Application Domains

## 4.1. Telecommunication network management

The domain of autonomic network management, under its new hype names, will remain an important playground for SUMO. It covers a wide variety of problems, ranging from distributed (optimal) control to distributed diagnosis, optimization, reconfiguration, provisioning, etc. We have a long experience in model-based diagnosis, in particular distributed (active) diagnosis, and have recently proposed promising techniques for self-modeling. It consists in building the model of the managed network on the fly, guided by the needs of the diagnosis algorithm. This approach allows one to deal with potentially huge models, that are only described by their construction grammar, and discovered at runtime. Another important research direction concerns the management of “multi-resolution” models, that can be considered at different granularity levels. This feature is central to network design, but has no appropriate modeling formalism nor management approaches. This is a typical investigation field for abstraction techniques. Technology is ahead of theory in this domain since networks are already driven or programmed through management policies, that assign high level objectives to an abstract view of the network, leaving open the question of their optimal implementation. As a last topic of investigation, today management issues are no longer isolated within one operator, but range across several of them, up to the supported services, which brings game theory aspects into the picture.

## 4.2. Control of data centers

Data centers are another example of a large scale reconfigurable and distributed system: they are composed of thousands of servers on which Virtual Machines (VM) can be (de)activated, migrated, etc. depending on the requests of the customers, on the load of the servers and on the power consumption. Autonomic management functionalities already exist to deploy and configure applications in such a distributed environment. They can also monitor the environment and react to events such as failures or overloads and reconfigure applications and/or infrastructures accordingly and autonomously. To supervise these systems, Autonomic Managers (AM) can be deployed in order to apply administration policies of specific aspects to the different entities of a data center (servers, VM, web services, power supply, etc). These AMs may be implemented in different layers: the hardware level, the operating system level or the middleware level. Therefore several control loops may coexist, and they have to take globally consistent decisions to manage the trade-off between availability, performance, scalability, security and energy consumption. This leads to multi-criteria optimization and control problems in order to automatically derive controllers in charge of the coordination of the different AMs. We are relatively new on this topic, that will require more technical investment from us. But we are driven to it by both the convergence of IT and networking, by virtualization techniques that reach networks (see the growing research effort about network operating systems), and by the call for more automation in the management of clouds. We believe our experience in network management can help. Some members of SUMO are already involved in the ANR Ctrl-Green, which addresses the controller coordination problem. We are also in contact with the Myriads team, which research interests moved from OS for grids/clouds to autonomic methods. This is supported as well by the activities of b<>com, the local IRT (see above), where some projects in cloud management and in networking may start joint activities.

## 4.3. Web services and distributed active documents

Data centric systems are already deployed, and our goal is not to design new languages, architectures, or standards for them, but rather to propose techniques for the verification and monitoring of the existing systems. A bottleneck is the complexity and heterogeneity of web-based systems, that make them difficult to model and analyze. However, one can still hope for some lightweight verification or monitoring techniques for some specific aspects, for example to check the absence of conflict of interest in a transaction system, to verify

(off line) and maintain (on line) the QoS, to prevent security breaches, etc. Safety aspects of WS are little addressed; any progress in that area would be useful. Besides, modeling issues are central for some applications of data centric systems. Collaborative work environments with shared active documents can be found in many domains ranging from banking, maintenance of critical systems, webstores... We consider that models for data driven systems can find applications in most of these application areas. Our approach, initiated in [21], will be to favor purely declarative approaches for the specification of such collaborative environments. We have contacts with Centre Pasteur in Yaoundé on the design of diseases monitoring systems in developing countries. Diseases monitoring systems can be seen as a collaborative edition work, where each actor in the system reports and aggregates information about cases he or she is aware of. This collaboration is an opportunity to confront our models to real situations and real users needs. Formally modeling such a large distributed system can be seen as a way to ensure its correctness. We also envision to promote this approach as a support for maintenance operations in complex environments (train transportation, aeronautics,...). We believe this framework can be useful both for the specification of distributed maintenance procedures, for circulating information and sharing processes across teams, but also for the analysis of the correctness of procedures, possibly for their optimization or redesign, and finally to automatically elaborate logs of maintenance operations. We are in contact with several major companies on these topics, for the maintenance application side. Other industrial contacts need to be built: we have preliminary contact with IBM (leader in business artifacts), and would like to establish relations with SAP (leader in service architectures).

## TASC Project-Team

# 4. Application Domains

## 4.1. Introduction

Constraint programming deals with the resolution of decision problems by means of rational, logical and computational techniques. Above all, constraint programming is founded on a clear distinction between, on the one hand the description of the constraints intervening in a problem, and on the other hand the techniques used for the resolution. The ability of constraint programming to handle in a flexible way heterogeneous constraints has raised the commercial interest for this paradigm in the early eighties. Among his fields of predilection, one finds traditional applications such as computer aided decision-making, scheduling, planning, placement, logistics or finance, as well as applications such as electronic circuits design (simulation, checking and test), DNA sequencing and phylogeny in biology, configuration of manufacturing products or web sites, formal verification of code.

## 4.2. Panorama

In 2012 the **TASC** team was involved in the following application domains:

- *Planning and replanning* in Data Centres SelfXL project).
- *Packing complex shapes* in the context of a warehouse (NetWMS2 project).
- Building decision support system for *city development planning with evaluation of energy impacts* (**SUSTAINS** project).
- *Optimizing electricity production* in the context of the **Gaspard Monge call program for Optimisation and Operation Research**. We extract global constraints from daily energy production temporal series issued from all productions plants of **EDF** over a period of several years.

## ASPI Project-Team

### 4. Application Domains

#### 4.1. Localisation, navigation and tracking

See 5.11 .

Among the many application domains of particle methods, or interacting Monte Carlo methods, ASPI has decided to focus on applications in localisation (or positioning), navigation and tracking [43], [37], which already covers a very broad spectrum of application domains. The objective here is to estimate the position (and also velocity, attitude, etc.) of a mobile object, from the combination of different sources of information, including

- a prior dynamical model of typical evolutions of the mobile, such as inertial estimates and prior model for inertial errors,
- measurements provided by sensors,
- and possibly a digital map providing some useful feature (terrain altitude, power attenuation, etc.) at each possible position.

In some applications, another useful source of information is provided by

- a map of constrained admissible displacements, for instance in the form of an indoor building map,

which particle methods can easily handle (map-matching). This Bayesian dynamical estimation problem is also called filtering, and its numerical implementation using particle methods, known as particle filtering, has been introduced by the target tracking community [42], [56], which has already contributed to many of the most interesting algorithmic improvements and is still very active, and has found applications in

target tracking, integrated navigation, points and / or objects tracking in video sequences, mobile robotics, wireless communications, ubiquitous computing and ambient intelligence, sensor networks, etc.

ASPI is contributing (or has contributed recently) to several applications of particle filtering in positioning, navigation and tracking, such as geolocalisation and tracking in a wireless network, terrain-aided navigation, and data fusion for indoor localisation.

#### 4.2. Rare event simulation

See 3.3 , and 5.3 , 5.6 , and 5.7 .

Another application domain of particle methods, or interacting Monte Carlo methods, that ASPI has decided to focus on is the estimation of the small probability of a rare but critical event, in complex dynamical systems. This is a crucial issue in industrial areas such as

nuclear power plants, food industry, telecommunication networks, finance and insurance industry, air traffic management, etc.

In such complex systems, analytical methods cannot be used, and naive Monte Carlo methods are clearly un-efficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [50], where trajectories going towards the critical set are given offsprings, thus increasing the number of trajectories that eventually reach the critical set. This approach not only makes it possible to estimate the probability of the rare event, but also provides realizations of the random trajectory, given that it reaches the critical set, i.e. provides realizations of typical critical trajectories, an important feature that methods based on importance sampling usually miss.

ASPI is contributing (or has contributed recently) to several applications of multilevel splitting for rare event simulation, such as risk assessment in air traffic management, detection in sensor networks, and protection of digital documents.

**I4S Project-Team (section vide)**

## IPSO Project-Team

### 4. Application Domains

#### 4.1. Laser physics

Laser physics considers the propagation over long space (or time) scales of high frequency waves. Typically, one has to deal with the propagation of a wave having a wavelength of the order of  $10^{-6}m$ , over distances of the order  $10^{-2}m$  to  $10^4m$ . In these situations, the propagation produces both a short-scale oscillation and exhibits a long term trend (drift, dispersion, nonlinear interaction with the medium, or so), which contains the physically important feature. For this reason, one needs to develop ways of filtering the irrelevant high-oscillations, and to build up models and/or numerical schemes that do give information on the long-term behavior. In other terms, one needs to develop high-frequency models and/or high-frequency schemes.

Generally speaking, the demand in developing such models or schemes in the context of laser physics, or laser/matter interaction, is large. It involves both modeling and numerics (description of oscillations, structure preserving algorithms to capture the long-time behaviour, etc).

In a very similar spirit, but at a different level of modelling, one would like to understand the very coupling between a laser propagating in, say, a fiber, and the atoms that build up the fiber itself.

The standard, quantum, model in this direction is called the Bloch model: it is a Schrödinger like equation that describes the evolution of the atoms, when coupled to the laser field. Here the laser field induces a potential that acts directly on the atom, and the link between this potential and the laser itself is given by the so-called dipolar matrix, a matrix made up of physical coefficients that describe the polarization of the atom under the applied field.

The scientific objective here is twofold. First, one wishes to obtain tractable asymptotic models that average out the high oscillations of the atomic system and of the laser field. A typical phenomenon here is the *resonance* between the field and the energy levels of the atomic system. Second, one wishes to obtain good numerical schemes in order to solve the Bloch equation, beyond the oscillatory phenomena entailed by this model.

#### 4.2. Molecular Dynamics

In classical molecular dynamics, the equations describe the evolution of atoms or molecules under the action of forces deriving from several interaction potentials. These potentials may be short-range or long-range and are treated differently in most molecular simulation codes. In fact, long-range potentials are computed at only a fraction of the number of steps. By doing so, one replaces the vector field by an approximate one and alternates steps with the exact field and steps with the approximate one. Although such methods have been known and used with success for years, very little is known on how the "space" approximation (of the vector field) and the time discretization should be combined in order to optimize the convergence. Also, the fraction of steps where the exact field is used for the computation is mainly determined by heuristic reasons and a more precise analysis seems necessary. Finally, let us mention that similar questions arise when dealing with constrained differential equations, which are a by-product of many simplified models in molecular dynamics (this is the case for instance if one replaces the highly-oscillatory components by constraints).

#### 4.3. Plasma physics

The development of efficient numerical methods is essential for the simulation of plasmas and beams at the kinetic level of description (Vlasov type equations). It is well known that plasmas or beams give rise to small scales (Debye length, Larmor radius, gyroperiod, mean free path...) which make numerical simulations challenging. Instead of solving the limit or averaged models by considering these small scales equal to zero, our aim is to explore a different strategy, which consists in using the original kinetic equation. Specific numerical scheme called "Asymptotic Preserving" scheme is then built to discretize the original kinetic

equation. Such a scheme allows to pass to the limit with no stability problems, and provide in the limit a consistent approximation of the limit or average model. A systematic and robust way to design such a scheme is the micro-macro decomposition in which the solution of the original model is decomposed into an averaged part and a remainder.



## DYLISS Project-Team

# 4. Application Domains

## 4.1. Formal models in molecular biology

As mentioned before, our main goal in biology is to characterize groups of genetic actors that control the response of living species capable of facing extreme environments. To focus our developments, applications and collaborations, we have identified three biological questions which deserve integrative studies. Each axis may be considered independently from the others although their combination, a mid-term challenge, will have the best impact in practice towards the long-term perspective of identifying proteins controlling the production of a metabolite of industrial interest. It is illustrated in our presentation for a major algae product: polyunsaturated fatty acids (PUFAs) and their derivatives.

**Biological data integration.** The first axis of the project (data integration) aims at identifying *who* is involved in the specific response of a biological system to an environmental stress. Targeted actors will mainly consist in groups of genetic products or biological pathways. For instance, which pathways are implied in the specific production of PUFAs in brown algae? The main work is to represent in a system of logical constraints the full knowledge at hand concerning the genetic or metabolic actors, the available observations and the effects of the system dynamics. To this aim, we focus on the use of Answer Set Programming as we are experienced in modeling with this paradigm and we have a strong partnership with a computer science team leader in the development of dedicated grounders and solvers (Potsdam university). See Sec. 3.1 .

**Asymptotic dynamics of a biological system** Once a model is built and its main actors are identified, the next step is to clarify *how* they combine to control the system. This is the second axis of the project. Roughly, the fine tuning of the system response may be of two types. Either it results from the discrete combinatorics of the actors, as the result of a genetic adaptation to extreme environmental conditions or the difference between species is rather at the enzyme-efficiency level. For instance, if Pufa's are found to be produced using a set of pathways specific to brown algae, the work in axis 2 will consist to apply constraint-based combinatorial approaches to select consistent combinations of pathways controlling the metabolite production. Otherwise, if enzymes controlling the production of Pufa's are found to be expressed in other algae, it suggests that the response of the system is rather governed by a fine quantitative tuning of pathways. In this case, we use symbolic dynamics and average-case analysis of algorithms to weight the respective importance of interactions in observed phenotypes (see Sec. 3.2 and Fig. 2 ). This specific approach is motivated by the quite restricted spectrum of available physiological observations over the asymptotic dynamics of the biological system.

**Biological sequence annotation** In order to check the accuracy of in-silico predictions, a third research axis of the team is to extract genetic actors responsible of biological pathways of interest in the targeted organism and locate them in the genome. In our guiding example, active proteins implied in Pufa's controlling pathways have to be precisely identified. Actors structures are represented by syntactic models (see Fig. 4 ). We use knowledge-based induction on far instances for the recognition of new members of a given sequence family within non-model genomes (see Fig. 3 ). A main objective is to model enzyme specificity with highly expressive syntactic structures - context-free model - in order to take into account constraints imposed by local domains or long-distance interactions within a protein sequence. See Sec. 3.3 for details.

## 4.2. Application fields

Our methods are applied in several fields of molecular biology.

Our main application field is **marine biology**, as it is a transversal field with respect to issues in integrative biology, dynamical systems and sequence analysis. Our main collaborators work at the Station Biologique de Roscoff. We are strongly involved in the study of brown algae: the *meneco*, *memap* and *memerge* tools were designed to realize a complete reconstruction of metabolic networks for non-benchmark species [48] [27]. On the same application model, the pattern discovery tool *protomata learner* allows for the classification of sub-families of specific proteins. The same tool also allowed us to gain a better understanding of cyanobacteria proteins [3]. Finally, in dynamical systems, we use asymptotic analysis (tool *pogg*) to decipher the initiation of sea urchin translation [37]. We are currently initiating two new research programs in this domain: the team will participate to a collaboration program with the Biocore and Ange Inria teams, focused on the understanding on green micro-algae; and we will be involved in the deciphering of phytoplankton variability at the system biology level in collaboration with the Station Biologique de Roscoff.

In **micro-biology**, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (funded by CMM, CRG and Inria-Chile). In order to elucidate the main characteristics of these bacteria, we develop efficient methods to identify the main groups of regulators for their specific response in their living environment. To that purpose, we use constraints-based modeling and combinatorial optimization. The integrative biology tools *bioquali*, *ingranalysis*, *shogun*, *lombarde* were designed in this context [6] [26]. In parallel, in collaboration with Ifremer (Brest), we have conducted similar work to decipher protein-protein interactions within archebacteria [20]. Our sequence analysis tool (*logol*) allowed us to build and maintain a very expressive CRISPR database [10].

Similarly, in **animal biology**, our goal is to propose methods to identify regulators of very complex phenotypes related to nutritional issues. In collaboration with researchers from Inra/Pegase and Inra/Igeep laboratories, we develop methods to distinguish the response of cows, chicken or porks to different diaries or treatments and characterize upstream transcriptional regulators for such a response. The system biology tool *nutritional analyzer* was designed in this framework [15]. The pattern matching tool *logol* also allows for a fine identification of transcription factor motifs [39]. Constraints-based programming also allows us to decipher regulators of reproduction for pea aphids.

We are less involved in **health** as the models and data studied in this application field are well informed and rather data-driven. In collaboration with Institut Curie, we have studied the Ewing Sarcoma regulation network to test the capability of our tool *bioquali* to accurately correct and predict a large-scale network behavior [1]. Our ongoing studies in this field focus on the exhaustive learning of discrete dynamical networks matching with experimental data, as a case study for modeling experimental design with constraints-based approaches. To that purpose, we collaborate with J. Saez Rodriguez group at EBI [18], [19] and N. Theret group at Inserm/Irset (Rennes) [35]. The dynamical system tools *caspo* and *cadbiom* were designed within these collaborations. Future studies will focus on the understanding of the metabolism of xenobiotics, still in collaboration with Inserm/Irset (Rennes).

## FLUMINANCE Project-Team

# 4. Application Domains

## 4.1. Introduction

By designing new approaches for the analysis of fluid-image sequences the FLUMINANCE group aims at contributing to several application domains of great interest for the community and in which the analysis of complex fluid flows plays a central role. The group focuses mainly on two broad application domains:

- Environmental sciences;
- Experimental fluid mechanics and industrial flows.

We detail hereafter these two application domains.

## 4.2. Environmental sciences

The first huge application domain concerns all the sciences that aim at observing the biosphere evolution such as meteorology, climatology or oceanography but also remote sensing study for the monitoring of meteorological events or human activities consequences. For all these domains image analysis is a practical and unique tool to *observe, detect, measure, characterize or analyze* the evolution of physical parameters over a large domain. The design of generic image processing techniques for all these domains might offer practical software tools to measure precisely the evolution of fluid flows for weather forecasting or climatology studies. It might also offer possibilities of close surveillance of human and natural activities in sensible areas such as forests, river edges, and valley in order to monitor pollution, floods or fire. The need in terms of local weather forecasting, risk prevention, or local climate change is becoming crucial for our tomorrow's life. At a more local scale, image sensors may also be of major utility to analyze precisely the effect of air curtains for safe packaging in agro-industrial.

## 4.3. Experimental fluid mechanics and industrial flows

In the domain of **experimental fluid mechanics**, the visualization of fluid flows plays a major role, especially for turbulence study since high frequency imaging has been made currently available. Together with analysis of turbulence at different scales, one of the major goals pursued at the moment by many scientists and engineers consists in studying the ability to manipulate a flow to induce a desired change. This is of huge technological importance to enhance or inhibit mixing in shear flows, improve energetic efficiency or control the physical effects of strain and stresses. This is for instance of particular interest for:

- military applications, for example to limit the infra-red signatures of fighter aircraft;
- aeronautics and transportation, to limit fuel consumption by controlling drag and lift effects of turbulence and boundary layer behavior;
- industrial applications, for example to monitor flowing, melting, mixing or swelling of processed materials, or preserve manufactured products from contamination by airborne pollutants, or in industrial chemistry to increase chemical reactions by acting on turbulence phenomena.

## **GENSCALE Project-Team**

### **4. Application Domains**

#### **4.1. Sequence comparison**

Historically, sequence comparison has been one of the most important topics in bioinformatics. BLAST is a famous software tool particularly designed for solving problems related to sequence comparisons. Initially conceived to perform searches in databases, it has mostly been used as a general-purpose sequence comparison tool. Nowadays, together with the inflation of genomic data, other software comparison tools that are able to provide better quality solutions (w.r.t the ones provided by BLAST) have been developed. They generally target specific comparison demands, such as read mapping, bank-to-bank comparison, meta-genomic sample analysis, etc. Today, sequence comparison algorithms must clearly be revisited to scale up with the very large number of sequence objects that new NGS problems have to handle.

#### **4.2. Genome comparison**

This application domain aims at providing a global relationship between genomes. The problem lies in the different structures that genomes can have: segments of genome can be rearranged, duplicated or deleted (the alignment can no longer be done in one piece). Therefore one major aim is the study of chromosomal rearrangements, breaking points, structural variation between individuals of the same species, etc. However, even analyses focused on smaller variations such as Single Nucleotide Polymorphisms (SNP) at the whole genome scale are different from the sequence comparison problem, since one needs first to identify common (orthologous) parts between whole genome sequences and thus obtain this global relationship (or map) between genomes. New challenges in genome comparison are emerging with the evolution of sequencing techniques. Nowadays, they allow for comparing genomes at intra-species level, and to deal simultaneously with hundreds or thousands of complete genomes. New methods are needed to find the sequence and structural variants between such a large number of non-assembled genomes. Even for the comparison of more distant species, classical methods must be revisited to deal with the increasing number of genomes but more importantly their decreasing quality: genomes are no longer fully assembled nor annotated.

#### **4.3. Protein comparison**

Comparing protein is important for understanding their evolutionary relationships and for predicting their structures and their functions. While annotating functions for new proteins, such as those solved in structural genomics projects, protein structural alignment methods may be able to identify functionally related proteins when the sequence identity between a given query protein and the related proteins are low (i.e. lower than 20%). Moreover, protein comparison allows for solving the so-called protein family identification problem. Given an unclassified protein structure (query), the comparison of protein structures can be used for assigning a score measuring the "similarity" between the query and the proteins belonging to a set of families. Based on this score, the query is assigned to one of the families of the set. The knowledge acquired by performing such analyses can then be exploited in methods for protein structure prediction that are based on a homology modeling approach.

## **SAGE Project-Team**

# **4. Application Domains**

## **4.1. Geophysics**

The team has chosen a particular domain of application, which is geophysics. In this domain, many problems require solving large scale systems of equations, arising from the discretization of coupled models. Emphasis is put on hydrogeology, but the team also investigates geodesy, heat and mass transfer in soil, and granular materials. One of the objectives is to use high performance computing in order to tackle 3D large scale computational domains with complex physical models.

## **4.2. Hydrogeology**

This is joint work with Geosciences Rennes at OSUR, Pprime at University of Poitiers and CDCSP at University of Lyon. It is also done in the context of the group Momas and Andra grant.

Many environmental studies rely on modelling geo-chemical and hydrodynamic processes. Some issues concern water resources, aquifer contamination, underground waste disposal, clean-up of former waste deposits, acid mine drainage remediation. Other issues, also related to energy, concern geothermy, unconventional gas, enhanced oil recovery, underground storage of CO<sub>2</sub>, underground storage of nuclear waste.

Simulation of contaminant transport in groundwater is a highly complex problem, governed by coupled linear or nonlinear PDAEs. Moreover, due to the lack of experimental data, stochastic models are used for dealing with heterogeneity. The main objective of the team is to design and to implement efficient and robust numerical models, including Uncertainty Quantification methods.

Recent research showed that rock solid masses are in general fractured and that fluids can percolate through networks of inter-connected fractures. Fractured media are by nature very heterogeneous and multi-scale, so that homogenisation approaches are not relevant. The team develops a numerical model for fluid flow and contaminant transport in three-dimensional porous fractured media.

An important output is the parallel scientific platform H2OLab, running on clusters, grids and machines available in supercomputing centers.

## SERPICO Project-Team

### 4. Application Domains

#### 4.1. Biological pilot models: Birbeck granule and Melanosome biogenesis

In the past recent years, research carried at UMR 144, CNRS-Institut Curie contributed to a better understanding of the intracellular compartmentation of specialized model cells such as melanocytes and Langerhans cells, the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively. These studies have started to highlight:

- multiple sorting and structural events involved in the biogenesis of these organelles;
- complexity of the endo-melanosomal network of these highly specialized cells;
- complex molecular architecture organizing and coordinating their dynamics;
- intracellular transport steps affected in genetic diseases, among which the Hermansky Pudlak syndrome (HPS) or involved in viral infection (HIV and Langerin in Langerhans cells).

In this context, the central aim of SERPICO is to understand how the different machineries of molecular components involved are interconnected and coordinated to generate such specialized structures. We need to address the following topics:

1. developing new bioimaging approaches to observe and statistically analyze such coordinated dynamics in live material;
2. correlating this statistically relevant spatiotemporal organization of protein networks with the biological architectures and within the overall biological environment as seen at the ultrastructural level;
3. modeling intracellular transport of those reference biological complex systems and proposing new experimental plans in an iterative and virtuous circle;
4. managing and analyzing the workflow of image data obtained along different multidimensional microscopy modalities.

These studies are essential to unravel the complexity of the endomembrane system and how different machineries evolve together (e.g. see Figs. 1 -2 ). They help to control cell organization and function at different scales through an integrative workflow of methodological and technological developments.

At long term, these studies will shed light on the cellular and molecular mechanisms underlying antigen presentation, viral infection or defense mechanisms, skin pigmentation, the pathogenesis of hereditary genetic disorders (lysosomal diseases, immune disorders) and on the mechanisms underlying cell transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, eventually at a time regime compatible with live cell imaging. The overview of ultrastructural organization will be achieved by complementary electron microscopical methods. Image visualization and quantitative analysis are of course important and essential issues in this context.

#### 4.2. Computational methods for bioimage informatics

In cell and molecular biology [62], new challenges arise to acquire a complete and quantified view from the scale of a “single” cell to the scale of a multi-cellular structure, within the whole organism. In the near future, image analysis will be central to the successful use of optical microscopy in post-genomics biology. Nevertheless, one major difficulty lies in correlating and/or fusing multi-modalities, now routinely used in biology laboratories: optical imaging (spinning-disk confocal, TIRF, SIM, PALM, STED, FLIM-FRET, MP, SPIM/DLSM), ionic imaging (NanoSIMS), atomic force imaging (AFM) and electron imaging (Cryo-EM, Tomo EM).



*Figure 1.*

*Traffic and space-time exocytosis analysis Cargo Langerin controlled  
by Rab11A/Rab11FIP2/MyoVb platform*



*Figure 2.*

*Investigation of Cargo Langerin trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform  
with TIRF (Total Internal Reflection Fluorescence) microscopy  
(Cell and Tissue Imaging Core Facility-IBiSA and Nikon Imaging Centre, CNRS-Institut Curie)*



Moreover, in the emerging era of high-throughput microscopy (biochemical screens, cell-based screening), systematic and accurate correlation and analysis of these data cannot be performed manually, since the image sequences are composed of several hundred of 3D stacks. Consequently, data to manipulate range from few to tens of TeraBytes. From the experimental perspective, molecular (drugs, RNA interference), mechanical (micro-patterning...), and optical (FRAP, photoactivation, optogenetic, ...) functional modulations allow one to quantify the importance of molecular linkage into macrocomplexes within a single cell. We are now able to limit shape variability between cells during an exposed period [59], [54]. Consequently, efficient storage, fast retrieval and secure sharing of microscopy images are crucial challenges. Even with high-speed computers, the processing step will considerably slow down the whole analysis process.

We propose to address several important issues in this area and to adapt the proposed methodologies and algorithms to face a deluge of data. Our goal is also to participate to the technical specifications of an image database with a built-in query system to annotate, retrieve, process and integrate analysis from different imaging modalities. The combination of complementary skills (image processing and analysis software, image data management) will yield a full integration of the image and data life-cycle, from image acquisition and analysis, to statistical analysis and mathematical modeling in systems biology.

## VISAGES Project-Team

# 4. Application Domains

## 4.1. Neuroimaging

*neuroimaging, clinical neuroscience, multiple sclerosis, multispectral MRI, brain atlas*

One research objective in neuroimaging is the construction of anatomical and functional cerebral maps under normal and pathological conditions.

Many researches are currently performed to find correlations between anatomical structures, essentially sulci and gyri, where neuronal activation takes place, and cerebral functions, as assessed by recordings obtained by the means of various neuroimaging modalities, such as PET (Positron Emission Tomography), fMRI (Functional Magnetic Resonance Imaging), EEG (Electro-EncephaloGraphy) and MEG (Magneto-EncephaloGraphy). Then, a central problem inherent to the formation of such maps is to put together recordings obtained from different modalities and from different subjects. This mapping can be greatly facilitated by the use of MR anatomical brain scans with high spatial resolution that allows a proper visualization of fine anatomical structures (sulci and gyri). Recent improvements in image processing techniques, such as segmentation, registration, delineation of the cortical ribbon, modelling of anatomical structures and multi-modality fusion, make possible this ambitious goal in neuroimaging. This problem is very rich in terms of applications since both clinical and neuroscience applications share similar problems. Since this domain is very generic by nature, our major contributions are directed towards clinical needs even though our work can address some specific aspects related to the neuroscience domain.

## 4.2. Multiple sclerosis

Over the past years, a discrepancy became apparent between clinical Multiple sclerosis (MS) classification describing on the one hand MS according to four different disease courses and, on the other hand, the description of two different disease stages (an early inflammatory and a subsequently neurodegenerative phase). It is to be expected that neuroimaging will play a critical role to define *in vivo* those four different MS lesion patterns. An *in vivo* distinction between the four MS lesion patterns, and also between early and late stages of MS will have an important impact in the future for a better understanding of the natural history of MS and even more for the appropriate selection and monitoring of drug treatment in MS patients. Since MRI has a low specificity for defining in more detail the pathological changes which could discriminate between the different lesion types, but a high sensitivity to detect focal and also widespread, diffuse pathology of the normal appearing white and grey matter, our major objective within this application domain is to define new neuroimaging markers for tracking the evolution of the pathology from high dimensional data (e.g. nD+t MRI). In addition, in order to complement MR neuroimaging data, we ambition to perform also cell labelling neuroimaging (e.g. MRI or PET) and to compare MR and PET data using standard and experimental MR contrast agents and radiolabeled PET tracers for activated microglia (e.g. USPIO or PK 11195). The goal is to define and develop, for routine purposes, cell specific and also quantitative imaging markers for the improved *in vivo* characterization of MS pathology.

## 4.3. Modelling of anatomical and anatomo-functional neurological patterns

The major objective within this application domain is to build anatomical and functional brain atlases in the context of functional mapping and for the study of developmental, neurodegenerative or even psychiatric brain diseases (Multiple sclerosis, Epilepsy, Parkinson, Dysphasia, Depression or even Alzheimer). This is a very competitive research domain; our contribution is based on our previous works in this field, and by continuing our local and wider collaborations.

An additional objective within this application domain is to find new descriptors to study the brain anatomy and/or function (e.g. variation of brain perfusion, evolution in shape and size of an anatomical structure in relation with pathology or functional patterns, computation of asymmetries ...). This is also a very critical research domain, especially for many developmental or neurodegenerative brain diseases.

**ACES Project-Team (section vide)**

## **ASAP Project-Team**

# **4. Application Domains**

## **4.1. Overview**

The results of the research targeted in ASAP span a wide range of applications. Below are a few examples.

- Personalized web search.
- Recommendation.
- Social networks.
- Notification systems.
- Distributed storage.
- Video streaming.

---

## ASCOLA Project-Team

# 4. Application Domains

## 4.1. Enterprise Information Systems and Services

Large IT infrastructures typically evolve by adding new third-party or internally-developed components, but also frequently by integrating already existing information systems. Integration frequently requires the addition of glue code that mediates between different software components and infrastructures but may also consist in more invasive modifications to implementations, in particular to implement crosscutting functionalities. In more abstract terms, enterprise information systems are subject to structuring problems involving horizontal composition (composition of top-level functionalities) as well as vertical composition (reuse and sharing of implementations among several top-level functionalities). Moreover, information systems have to be more and more dynamic.

Service-Oriented Computing (SOC) that is frequently used for solving some of the integration problems discussed above. Indeed, service-oriented computing has two main advantages:

- Loose-coupling: services are autonomous, in that they do not require other services to be executed;
- Ease of integration: services communicate over standard protocols.

Our current work is based on the following observation: similar to other compositional structuring mechanisms, SOAs are subject to the problem of crosscutting functionalities, that is, functionalities that are scattered and tangled over large parts of the architecture and the underlying implementation. Security functionalities, such as access control and monitoring for intrusion detection, are a prime example of such a functionality in that it is not possible to modularize security issues in a well-separated module. Aspect-Oriented Software Development is precisely an application-structuring method that addresses in a systemic way the problem of the lack of modularization facilities for crosscutting functionalities.

We are considering solutions to secure SOAs by providing an aspect-oriented structuring and programming model that allows security functionalities to be modularized. Two levels of research have been identified:

- Service level: as services can be composed to build processes, aspect weaving will deal with the orchestration and the choreography of services.
- Implementation level: as services are abstractly specified, aspect weaving will require to extend service interfaces in order to describe the effects of the executed services on the sensitive resources they control.

In 2013, we have developed techniques for the Service-Level Agreement (SLA) management for Cloud elasticity, see Sec. 6.3 , as well as models and type systems for service-oriented systems, see Sec. 6.1 . Furthermore, we take part in the European project A4Cloud on accountability challenges, that is, the responsible stewardship of third-party data and computations, see Sec. 8.2 .

## 4.2. Capacity Planning in Cluster, Grid and Cloud Computing

Cluster, Grid and more recently Cloud computing platforms aim at delivering large capacities of computing power. These capacities can be used to improve performance (for scientific applications) or availability (e.g., for Internet services hosted by datacenters). These distributed infrastructures consist of a group of coupled computers that work together and may be spread across a LAN (cluster), across a WAN (Grid), and across the Internet (Clouds). Due to their large scale, these architectures require permanent adaptation, from the application to the system level and call for automation of the corresponding adaptation processes. We focus on self-configuration and self-optimization functionalities across the whole software stack: from the lower levels (systems mechanisms such as distributed file systems for instance) to the higher ones (i.e. the applications themselves such as J2EE clustered servers or scientific grid applications).

---

In 2013, we have confirmed the scalability of the DVMS proposal by conducting experiments on a very large scale involving up to 5K virtual machines (VM)s upon 500 nodes, thus establishing it as one of the most scalable placement algorithm for virtual machines. Moreover, we have extended the SimGrid framework by adding virtualization abstractions for hundreds of thousands of VMs. Finally, we have also provided several results on the energy efficient management of Cloud applications and infrastructures, see Sec. 6.3 .

In the energy field, we have designed a set of techniques, named OptiPlace, for cloud management with flexible power models through constraint programming. OptiPlace supports external models, named views. Specifically, we have developed a power view, based on generic server models, to define and reduce the power consumption of a datacenter's physical servers. We have shown that OptiPlace behaves at least as good as our previous system, Entropy, requiring as low as half the time to find a solution for the constrained-based placement of tasks for large datacenters.

### **4.3. Pervasive Systems**

Pervasive systems are another class of systems raising interesting challenges in terms of software structuring. Such systems are highly concurrent and distributed. Moreover, they assume a high-level of mobility and context-aware interactions between numerous and heterogeneous devices (laptops, PDAs, smartphones, cameras, electronic appliances...). Programming such systems requires proper support for handling various interfering concerns like software customization and evolution, security, privacy, context-awareness... Additionally, service composition occurs spontaneously at runtime.

In 2013, we have extended the language EScala, which integrates reactive programming through events with aspect-oriented and object-oriented mechanisms, see Sec. 6.1 .

## **ATLANMOD Project-Team**

# **4. Application Domains**

## **4.1. Application domains**

By definition, MDE can be applied to any software domain. Core MDE techniques developed by the team have been successfully applied to a large variety of industrial domains from information systems to embedded systems. MDE is not even restricted to software engineering, but also applies to data engineering [49] and to system engineering [41]. There are a lot of problems in these application domains that may be addressed by means of modeling and model transformation techniques.

As a result, AtlanMod has collaborated with a great variety of different companies ranging from the Automotive to the Insurances domains and from SMEs to large enterprises through the projects described later on in this same report. AtlanMod hopes to continue this trend in the future.



## **CIDRE Project-Team**

# **4. Application Domains**

## **4.1. Application Domains**

With the infiltration of computers and software in almost all aspects of our modern life, security can nowadays be seen as an absolutely general concern. As such, the results of the research targeted by CIDRE apply to a wide range of domains. It is clear that critical systems, where security (and safety) is a major concern, may benefit from ideas such as dynamic security policy monitoring. On the other hand, systems used by general public (basically, the internet and services such as web or cloud services, social networks, location-based services, etc.) can also benefit from results obtained by CIDRE, especially with respect to privacy. In addition, systems are getting more and more complex, decentralized, distributed, or spontaneous. The emergence of cloud computing brings many challenges that could benefit from ideas, approaches and solutions studied by CIDRE in the context of distributed systems.

## **DIONYSOS Project-Team**

# **4. Application Domains**

## **4.1. Networking**

Our global research effort concerns networking problems, both from the analysis point of view, and around network design issues. Specifically, this means the IP technology in general, with focus on specific types of networks seen at different levels: wireless systems, optical infrastructures, peer-to-peer architectures, Software Defined Networks, Content Delivery Networks, Content-Centric Networks, clouds.

## **4.2. Complex systems**

Many of the techniques developed at Dionysos are useful for the analysis of complex systems in general, not only in telecommunications. For instance, our Monte Carlo methods for analyzing rare events have been used by different industrial partners, some of them in networking but recently also by companies building transportation systems.

---

## KERDATA Project-Team

# 4. Application Domains

## 4.1. Application Domains

Below are three examples which illustrate the needs of large-scale data-intensive applications with respect to storage, I/O and data analysis. They illustrate the classes of applications that can benefit from our research activities.

### 4.1.1. *Joint genetic and neuroimaging data analysis on Azure clouds*

Joint acquisition of neuroimaging and genetic data on large cohorts of subjects is a new approach used to assess and understand the variability that exists between individuals, and that has remained poorly understood so far. As both neuroimaging- and genetic-domain observations represent a huge amount of variables (of the order of millions), performing statistically rigorous analyses on such amounts of data is a major computational challenge that cannot be addressed with conventional computational techniques only. On the one hand, sophisticated regression techniques need to be used in order to perform significant analysis on these large datasets; on the other hand, the cost entailed by parameter optimization and statistical validation procedures (e.g. permutation tests) is very high.

The A-Brain (AzureBrain) Project started in October 2010 within the Microsoft Research-Inria Joint Research Center. It is co-led by the KerData (Rennes) and Parietal (Saclay) Inria teams. They jointly address this computational problem using cloud related techniques on Microsoft Azure cloud infrastructure. The two teams bring together their complementary expertise: KerData in the area of scalable cloud data management, and Parietal in the field of neuroimaging and genetics data analysis.

In particular, KerData brings its expertise in designing solutions for optimized data storage and management for the Map-Reduce programming model. This model has recently arisen as a very effective approach to develop high-performance applications over very large distributed systems such as grids and now clouds. The computations involved in the statistical analysis designed by the Parietal team fit particularly well with this model.

### 4.1.2. *Structural protein analysis on Nimbus clouds*

Proteins are major components of the life. They are involved in lots of biochemical reactions and vital mechanisms for living organisms. The three-dimensional (3D) structure of a protein is essential for its function and for its participation to the whole metabolism of a living organism. However, due to experimental limitations, only few protein structures (roughly, 60,000) have been experimentally determined, compared to the millions of proteins sequences which are known. In the case of structural genomics, the knowledge of the 3D structure may be not sufficient to infer the function. A usual way to make a structural analysis of a protein or to infer its function is to compare its known, or potential, structure to the whole set of structures referenced in the *Protein Data Bank* (PDB).

In the framework of the MapReduce ANR project led by KerData, we focus on the SuMo application (*Surf the Molecules*) proposed by Institute for Biology and Chemistry of the Proteins from Lyon (IBCP, a partner in the MapReduce project). This application performs structural protein analysis by comparing a set of protein structures against a very large set of structures stored in a huge database. This is a typical data-intensive application that can leverage the Map-Reduce model for a scalable execution on large-scale distributed platforms. Our goal is to explore storage-level concurrency-oriented optimizations to make the SuMo application scalable for large-scale experiments of protein structures comparison on cloud infrastructures managed using the Nimbus IaaS toolkit developed at Argonne National Lab (USA).

If the results are convincing, then they can immediately be applied to the derived version of this application for drug design in an industrial context, called MED-SuMo, a software managed by the MEDIT SME (also a partner in this project). For pharmaceutical and biotech industries, such an implementation run over a cloud computing facility opens several new applications for drug design. Rather than searching for 3D similarity into biostructural data, it will become possible to classify the entire biostructural space and to periodically update all derivative predictive models with new experimental data. The applications in this complete chemo-proteomic vision concern the identification of new druggable protein targets and thereby the generation of new drug candidates.

#### ***4.1.3. I/O intensive climate simulations for the Blue Waters post-Petascale machine***

A major research topic in the context of HPC simulations running on post-Petascale supercomputers is to explore how to efficiently record and visualize data during the simulation without impacting the performance of the computation generating that data. Conventional practice consists in storing data on disk, moving them off-site, reading them into a workflow, and analyzing them. This approach becomes increasingly harder to use because of the large data volumes generated at fast rates, in contrast to limited back-end speeds. Scalable approaches to deal with these I/O limitations are thus of utmost importance. This is one of the main challenges explicitly stated in the roadmap of the Blue Waters Project (<http://www.ncsa.illinois.edu/BlueWaters/>), which aims to build one of the most powerful supercomputers in the world.

In this context, the KerData project-team started to explore ways to remove the limitations mentioned above through collaborative work in the framework of the Joint Inria-UIUC Lab for Petascale Computing (JLPC, Urbana-Champaign, Illinois, USA), whose research activity focuses on the Blue Waters project. As a starting point, we are focusing on a particular tornado simulation code called CM1 (Cloud Model 1), which is intended to be run on the Blue Waters machine. Preliminary investigation demonstrated the inefficiency of the current I/O approach, which typically consists in periodically writing a very large number of small files. This causes bursts of I/O in the parallel file system, leading to poor performance and extreme variability (jitter) compared to what could be expected from the underlying hardware. The challenge here is to investigate how to make an efficient use of the underlying file system by avoiding synchronization and contention as much as possible. In collaboration with the JLPC, we started to address these challenges through an approach based on dedicated I/O cores.

## **MYRIADS Project-Team**

# **4. Application Domains**

## **4.1. Overview**

The Myriads research activities address a broad range of application domains. We validate our research results with selected use cases from the following application domains:

- Web services, Service oriented applications,
- Business applications,
- Bio-informatics applications,
- Computational science applications,
- Numerical simulations.

## **TRISKELL Project-Team**

# **4. Application Domains**

## **4.1. Application Domains**

SOA, telecommunication, distributed systems, Embedded Systems, software engineering, test, UML

From small embedded systems such as home automation products or automotive systems to medium sized systems such as medical equipment, office equipment, household appliances, smart phones; up to large Service Oriented Architectures (SOA), building a new application from scratch is no longer possible. Such applications reside in (group of) machines that are expected to run continuously for years without unrecoverable errors. Special care has then to be taken to design and validate embedded software, making the appropriate trade-off between various extra-functional properties such as reliability, timeliness, safety and security but also development and production cost, including resource usage of processor, memory, bandwidth, power, etc.

Leveraging ongoing advances in hardware, embedded software is playing an evermore crucial role in our society, bound to increase even more when embedded systems get interconnected to deliver ubiquitous SOA. For this reason, embedded software has been growing in size and complexity at an exponential rate for the past 20 years, pleading for a component based approach to embedded software development. There is a real need for flexible solutions allowing to deal at the same time with a wide range of needs (product lines modeling and methodologies for managing them), while preserving quality and reducing the time to market (such as derivation and validation tools).

We believe that building flexible, reliable and efficient embedded software will be achieved by reducing the gap between executable programs, their models, and the platform on which they execute, and by developing new composition mechanisms as well as transformation techniques with a sound formal basis for mapping between the different levels.

Reliability is an essential requirement in a context where a huge number of softwares (and sometimes several versions of the same program) may coexist in a large system. On one hand, software should be able to evolve very fast, as new features or services are frequently added to existing ones, but on the other hand, the occurrence of a fault in a system can be very costly, and time consuming. While we think that formal methods may help solving this kind of problems, we develop approaches where they are kept “behind the scene” in a global process taking into account constraints and objectives coming from user requirements.

Software testing is another aspect of reliable development. Testing activities mostly consist in trying to exhibit cases where a system implementation does not conform to its specifications. Whatever the efforts spent for development, this phase is of real importance to raise the confidence level in the fact that a system behaves properly in a complex environment. We also put a particular emphasis on on-line approaches, in which test and observation are dynamically computed during execution.

## DREAM Project-Team

# 4. Application Domains

## 4.1. Introduction

The DREAM research applications have been oriented towards surveillance of large networks as telecommunication networks and more recently of web services. During the past few years, we have focussed increasingly on agricultural and environmental applications by means of research collaborations with INRA and Agrocampus Ouest.

## 4.2. Software components monitoring

software components, web services, distributed diagnosis

Web-services, i.e., services that are provided, controlled and managed through Internet, cover nowadays more and more application areas, from travel booking to goods supplying in supermarkets or the management of an e-learning platform. Such applications need to process requests from users and other services on line, and respond accurately in real time. Anyway, errors may occur, which need to be addressed in order to still be able to provide the correct response with a satisfactory quality of service (QoS): on-line monitoring, especially diagnosis and repair capabilities, become then a crucial concern.

We have been working on this problem within the WS-DIAMOND project [84], a large European funded project involving eight partners in Italy, France, Austria and Netherlands <http://wsdiamond.di.unito.it/>. Our own work consisted in two distinct contributions.

The first issue has been to extend the decentralized component-oriented approach, initially developed for monitoring telecommunication networks [4] to this new domain. To this end we have proposed the concept of distributed chronicles, with synchronization events, and the design of an architecture consisting of distributed CRSs (Chronicle Recognition Systems) communicating their local diagnoses to a broker agent which is in charge of merging them to compute a global diagnosis.

Our current work aims at coupling diagnosing and repair, in order to implement *adaptive web services*. We started this study by proposing an architecture inspired from the one developed during the WS-DIAMOND project and dedicated to the adaptive process of a request event when faults occur and propagate through the orchestration.

## 4.3. Environmental decision making

environment, decision methods

The need of decision support systems in the environmental domain is now well-recognized. It is especially true in the domain of water quality. For instance the program, named “Bretagne Eau Pure”. was launched a few years ago in order to help regional managers to protect this important resource in Brittany. The challenge is to preserve the water quality from pollutants as nitrates and herbicides, when these pollutants are massively used by farmers to weed their agricultural plots and improve the quality and increase the quantity of their crops. The difficulty is then to find solutions which satisfy contradictory interests and to get a better knowledge on pollutant transfer.

In this context, we are cooperating with INRA (Institut National de Recherche Agronomique) and developing decision support systems to help regional managers in preserving the river water quality. The approach we advocate is to rely on a qualitative modeling, in order to model biophysical processes in an explicative and understandable way. The SACADEAU model associates a qualitative biophysical model, able to simulate the biophysical process, and a management model, able to simulate the farmer decisions. One of our main contribution is the use of qualitative spatial modeling, based on runoff trees, to simulate the pollutant transfer through agricultural catchments.

The second issue is the use of learning/data mining techniques to discover, from model simulation results, the discriminant variables and automatically acquire rules relating these variables. One of the main challenges is that we are faced with spatiotemporal data. The learned rules are then analyzed in order to recommend actions to improve a current situation.

This work has been done in the framework of the APPEAU project, funded by ANR and of the ACASSYA project, funded by ANR, having started at the beginning of 2009 and ended at the end of 2012. We were also involved in the PSDR GO CLIMASTER project, that started in september 2008 and end in 2011. CLIMASTER stands for “Changement climatique, systèmes agricoles, ressources naturelles et développement territorial” and is dedicated to the impact of climate changes on the agronomical behaviors in west of France (“Grand Ouest”). PSDR GO stands for “Programme Pour et Sur le Développement Régional Grand Ouest”.

Our main partners are the SAS INRA research group, located in Rennes and the BIA INRA and AGIR INRA research groups in Toulouse.



## **HYBRID Project-Team**

# **4. Application Domains**

## **4.1. Overview**

The research program of Hybrid team aims at next generations of virtual reality and 3D user interfaces which could possibly address both the “body” and “mind” of the user. Novel interaction schemes are designed, for one or multiple users. We target better integrated systems and more compelling user experiences.

The applications of our research program correspond to the applications of virtual reality technologies which could benefit from the addition of novel body-based or mind-based interaction capabilities:

- **Industry:** with training systems, virtual prototyping, or scientific visualization;
- **Medicine:** with rehabilitation and reeducation systems, or surgical training simulators;
- **Entertainment:** with 3D web navigations, video games, or attractions in theme parks,
- **Construction:** with virtual mock-ups design and review, or historical/architectural visits.

## **LAGADIC Project-Team**

# **4. Application Domains**

## **4.1. Application Domains**

The natural applications of our research are obviously in robotics. In fact, researches undertaken in the Lagadic group can apply to all the fields of robotics implying a vision sensor. They are indeed conceived to be independent of the system considered (and the robot and the vision sensor can even be virtual for some applications).

Currently, we are mostly interested in using visual servoing for aerial and space application, micromanipulation, autonomous vehicle navigation in large urban environments or for disabled or elderly people.

We also address the field of medical robotics. The applications we consider turn around new functionalities of assistance to the clinician during a medical examination: visual servoing on echographic images, needle insertion, compensation of organ motions, etc.

Robotics is not the only possible application field to our researches. In the past, we were interested in applying visual servoing in computer animation, either for controlling the motions of virtual humanoids according to their pseudo-perception, or for controlling the point of view of visual restitution of an animation. In both cases, potential applications are in the field of virtual reality, for example for the design of video games, or virtual cinematography.

Applications also exist in computer vision and augmented reality. It is then a question of carrying out a virtual visual servoing for the 3D localization of a tool with respect to the vision sensor, or for the estimation of its 3D motion. This field of application is very promising, because it is in full rise for the realization of special effects in the multi-media field or for the design and the inspection of objects manufactured in the industrial world.

## **MIMETIC Project-Team**

# **4. Application Domains**

## **4.1. Autonomous Characters**

Autonomous characters are becoming more and more popular as they are used in an increasing number of application domains. In the field of special effects, virtual characters are used to replace secondary actors and generate highly populated scenes that would be hard and costly to produce with real actors. In video games and virtual storytelling, autonomous characters play the role of actors that are driven by a scenario. Their autonomy allows them to react to unpredictable user interactions and adapt their behavior accordingly. In the field of simulation, autonomous characters are used to simulate the behavior of humans in different kinds of situations. They enable to study new situations and their possible outcomes.

One of the main challenges in the field of autonomous characters is to provide a unified architecture for the modeling of their behavior. This architecture includes perception, action and decisional parts. This decisional part needs to mix different kinds of models, acting at different time scale and working with different nature of data, ranging from numerical (motion control, reactive behaviors) to symbolic (goal oriented behaviors, reasoning about actions and changes).

In the MIMETIC team, we focus on autonomous virtual humans. Our problem is not to reproduce the human intelligence but to propose an architecture making it possible to model credible behaviors of anthropomorphic virtual actors evolving/moving in real time in virtual worlds. The latter can represent particular situations studied by psychologists of the behavior or to correspond to an imaginary universe described by a scenario writer. The proposed architecture should mimic all the human intellectual and physical functions.

## **4.2. Biomechanics and Motion Analysis**

Biomechanics is obviously a very large domain. This large set can be divided regarding to the scale at which the analysis is performed going from microscopic evaluation of biological tissues' mechanical properties to macroscopic analysis and modeling of whole body motion. Our topics in the domain of biomechanics mainly lie within this last scope.

The first goal of such kind of research projects is a better understanding of human motion. The MIMETIC team addresses three different situations: everyday motions of a lambda subject, locomotion of pathological subjects and sports gesture.

In the first set, MIMETIC is interested in studying how subjects maintain their balance in highly dynamic conditions. Until now, balance has nearly always been considered in static or quasi-static conditions. The knowledge of much more dynamic cases still has to be improved. Our approach has demonstrated that first of all, the question of the parameter that will allow to do this is still open. We have also taken interest into collision avoidance between two pedestrian. This topic includes the research of the parameters that are interactively controlled and the study of each one's role within this interaction.

When patients, in particular those suffering from central nervous system affection, cannot have an efficient walking it becomes very useful for practitioners to benefit from an objective evaluation of their capacities. To propose such help to patients following, we have developed two complementary indices, one based on kinematics and the other one on muscles activations. One major point of our research is that such indices are usually only developed for children whereas adults with these affections are much more numerous.

Finally, in sports, where gesture can be considered, in some way, as abnormal, the goal is more precisely to understand the determinants of performance. This could then be used to improve training programs or devices. Two different sports have been studied: the tennis serve, where the goal was to understand the contribution of each segments of the body in ball's speed and the influence of the mechanical characteristics of the fin in fin swimming.

After having improved the knowledge of these different gestures a second goal is then to propose modeling solutions that can be used in VR environments for other research topics within MimeTic. This has been the case, for example, for the collision avoidance.

### 4.3. Crowds

Crowd simulation is a very active and concurrent domain. Various disciplines are interested in crowds modeling and simulation: Mathematics, Cognitive Sciences, Physics, Computer Graphics, etc. The reason for this large interest is that crowd simulation raise fascinating challenges.

At first, crowd can be first seen as a complex system: numerous local interactions occur between its elements and results into macroscopic emergent phenomena. Interactions are of various nature and are undergoing various factors as well. Physical factors are crucial as a crowd gathers by definition numerous moving people with a certain level of density. But sociological, cultural and psychological factors are important as well, since crowd behavior is deeply changed from country to country, or depending on the considered situations.

On the computational point of view, crowd push traditional simulation algorithms to their limit. An element of a crowd is subject to interact with any other element belonging the same crowd, a naive simulation algorithm has a quadratic complexity. Specific strategies are set to face such a difficulty: level-of-detail techniques enable scaling large crowd simulation and reach real-time solutions.

MimeTIC is an international key contributor in the domain of crowd simulation. Our approach is specific and based on three axis. First, our modeling approach is founded on human movement science: we conducted challenging experiment on the motion of groups. Second: we developed high-performance solutions for crowd simulation. Third, we develop solutions for realistic navigation in virtual world to enable interaction with crowds in Virtual Reality.

### 4.4. Motion Sensing

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user's body parts. Whatever the system is, one of the main is to be able to automatically recognize and analyze the user's performance according to poor and noisy signals. Human activity and motion are subject to variability: intra-variability due to space and time variations of a given motion, but also inter-variability due to different styles and anthropometric dimensions. MimeTIC has addressed the above problems in two main directions.

Firstly, we have studied how to recognize and quantify motions performed by a user when using accurate systems such as Vicon (product of Oxford Metrics) or Optitrack (product of Natural Point) motion capture systems. These systems provide large vectors of accurate information. Due to the size of the state vector (all the degrees of freedom) the challenge is to find the compact information (named features) that enables the automatic system to recognize the performance of the user. Whatever the method is used, finding these relevant features that are not sensitive to intra-individual and inter-individual variability is a challenge. Some researchers have proposed to manually edit these features (such as a Boolean value stating if the arm is moving forward or backward) so that the expertise of the designer is directly linked with the success ratio. Many proposals for generic features have been proposed, such as using Laban notation which was introduced to encode dancing motions. Other approaches tend to use machine learning to automatically extract these features. However most of the proposed approaches were used to seek a database for motions which properties correspond to the features of the user's performance (named motion retrieval approaches). This does not ensure the retrieval of the exact performance of the user but a set of motions with similar properties.

Secondly, we wish to find alternatives to the above approach which is based on analyzing accurate and complete knowledge on joint angles and positions. Hence new sensors, such as depth-cameras (Kinect, product of Microsoft) provide us with very noisy joint information but also with the surface of the user. Classical approaches would try to fit a skeleton into the surface in order to compute joint angles which, again, lead to large state vectors. An alternative would be to extract relevant information directly from the raw data, such as the surface provided by depth cameras. The key problem is that the nature of these data may be very different

from classical representation of human performance. In MimeTIC, we try to address this problem in specific application domains that require picking specific information, such as gait asymmetry or regularity for clinical analysis of human walking.

## **4.5. VR and Sports**

Sport is characterized by complex displacements and motions. These motions are dependent on visual information that the athlete can pick up in his environment, including the opponent's actions. The perception is thus fundamental to the performance. Indeed, a sportive action, as unique, complex and often limited in time, requires a selective gathering of information. This perception is often seen as a prerogative for action, it then takes the role of a passive collector of information. However, as mentioned by Gibson in 1979, the perception-action relationship should not be considered sequential but rather as a coupling: we perceive to act but we must act to perceive. There would thus be laws of coupling between the informational variables available in the environment and the motor responses of a subject. In other words, athletes have the ability to directly perceive the opportunities of action directly from the environment. Whichever school of thought considered, VR offers new perspectives to address these concepts by complementary using real time motion capture of the immersed athlete.

In addition to better understanding sports and interaction between athletes, VR can also be used as a training environment as it can provide complementary tools to coaches. It is indeed possible to add visual or auditory information to better train an athlete. The knowledge found in perceptual experiments can be for example used to highlight the body parts that are important to look at to correctly anticipate the opponent's action.

## **4.6. Interactive Digital Storytelling**

Interactive digital storytelling, including novel forms of edutainment and serious games, provides access to social and human themes through stories which can take various forms and contains opportunities for massively enhancing the possibilities of interactive entertainment, computer games and digital applications. It provides chances for redefining the experience of narrative through interactive simulations of computer-generated story worlds and opens many challenging questions at the overlap between computational narratives, autonomous behaviours, interactive control, content generation and authoring tools.

Of particular interest for the Mimetic research team, virtual storytelling triggers challenging opportunities in providing effective models for enforcing autonomous behaviours for characters in complex 3D environments. Offering both low-level capacities to characters such as perceiving the environments, interacting with the environment and reacting to changes in the topology, on which to build higher-levels such as modelling abstract representations for efficient reasoning, planning paths and activities, modelling cognitive states and behaviours requires the provision of expressive, multi-level and efficient computational models. Furthermore virtual storytelling requires the seamless control of the balance between the autonomy of characters and the unfolding of the story through the narrative discourse. Virtual storytelling also raises challenging questions on the conveyance of a narrative through interactive or automated control of the cinematography (how to stage the characters, the lights and the cameras). For example, estimating visibility of key subjects, or performing motion planning for cameras and lights are central issues for which have not received satisfactory answers in the literature.

## **4.7. VR and Ergonomics**

The design of workstations nowadays tends to include assessment steps in a Virtual Environment (VE) to evaluate ergonomic features. This approach is more cost-effective and convenient since working directly on the Digital Mock-Up (DMU) in a VE is preferable to constructing a real physical mock-up in a Real Environment (RE). This is substantiated by the fact that a Virtual Reality (VR) set-up can be easily modified, enabling quick adjustments of the workstation design. Indeed, the aim of integrating ergonomics evaluation tools in VE is to facilitate the design process, enhance the design efficiency, and reduce the costs.

The development of such platforms ask for several improvements in the field of motion analysis and VR: the interactions have to be as fidelistic as possible to properly mimic the motions performed in real environments, the fidelity of the simulator need also to be correctly evaluated, and motion analysis tools have to be able to provide in real-time biomechanics quantities usable by ergonomists to analyse and improve the working conditions.

## PANAMA Project-Team

# 4. Application Domains

## 4.1. Acoustic scene capture

Acoustic fields carry much information about audio sources (musical instruments, speakers, etc.) and their environment (e.g., church acoustics differ much from office room acoustics). A particular challenge is to capture as much information from a complete 3D+t acoustic field associated with an audio scene, using as few sensors as possible. The feasibility of compressive sensing to address this challenge was shown in certain scenarii, and the actual implementation of this framework will potentially impact practical scenarii such as remote surveillance to detect abnormal events, e.g. for health care of the elderly or public transport surveillance.

## 4.2. Audio signal separation in reverberant environments

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the estimation of the unknown room impulse response associated to a matrix of mixing filters, which can be expressed as a dictionary-learning problem. Solutions to this problem have the potential to impact, for example, the music and game industry, through the development of new digital re-mastering techniques and virtual reality tools, but also surveillance and monitoring applications, where localizing audio sources is important.

## 4.3. Multimedia indexing

Audiovisual and multimedia content generate large data streams (audio, video, associated data such as text, etc.). Manipulating large databases of such content requires efficient techniques to: segment the streams into coherent sequences; label them according to words, language, speaker identity, and more generally to the type of content; index them for easy querying and retrieval, etc. As the next generation of online search engines will need to offer content-based means of searching, the need to drastically reduce the computational burden of these tasks is becoming all the more important as we can envision the end of the era of wasteful datacenters that can increase forever their energy consumption. Most of today's techniques to deal with such large audio streams involve extracting features such as Mel Frequency Cepstral Coefficients (MFCC) and learning high-dimensional statistical models such as Gaussian Mixture Models, with several thousand parameters. The exploration of a compressive learning framework is expected to contribute to new techniques to efficiently process such streams and perform segmentation, classification, etc., in the compressed domain. A particular challenge is to understand how this paradigm can help exploiting truly multimedia features, which combine information from different associated streams such as audio and video, for joint audiovisual processing.

## 4.4. Brain source imaging

Epilepsies constitute a common neurological disorder that affects about 1% of the world population. As the epileptic seizure is a dynamic phenomenon, imaging techniques showing static images of the brain (MRI, PET scan) are frequently not the best tools to identify the brain area of interest. Electroencephalography (EEG) is the technique most indicated to capture transient events directly related to the underlying epileptic pathology (like interictal spikes, in particular). EEG convey essential information regarding brain (patho-)physiological activity. In addition, recording techniques of surface signals have the major advantage of being noninvasive. For this reason, an increased use in the context of epilepsy surgery is most wanted. However, to

reach this objective, we have to solve an electromagnetic inverse problem, that is to say to estimate the current generators underlying noisy EEG data. Theoretically, a specific electromagnetic field pattern may be generated by an infinite number of current distributions. The considered inverse problem, called "brain source imaging problem", is then said to be ill-posed.



## SIROCCO Project-Team

# 4. Application Domains

## 4.1. Introduction

The application domains addressed by the project are:

- Compression with advanced functionalities of various image modalities (including multi-view, medical images such as MRI, CT, WSI, or satellite images)
- Networked multimedia applications via their various needs in terms of image and 2D and 3D video compression, or in terms of network adaptation (e.g., resilience to channel noise)
- Content editing and post-production

## 4.2. Compression with advanced functionalities

Compression of images and of 2D video (including High Definition and Ultra High Definition) remains a widely-sought capability for a large number of applications. The continuous increase of access network bandwidth leads to increasing numbers of networked digital content users and consumers which in turn triggers needs for higher core bandwidth and higher compression efficiencies. This is particularly true for mobile applications, as the need for wireless transmission capacity will significantly increase during the years to come. Hence, efficient compression tools are required to satisfy the trend towards mobile access to larger image resolutions and higher quality. A new impulse to research in video compression is also brought by the emergence of new formats beyond High Definition TV (HDTV) towards high dynamic range (higher bit depth, extended colorimetric space), super-resolution, formats for immersive displays allowing panoramic viewing and 3DTV.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FTV). The notion of "free viewpoint video" refers to the possibility for the user to choose an arbitrary viewpoint and/or view direction within a visual scene, creating an immersive environment. Multi-view video generates a huge amount of redundant data which need to be compressed for storage and transmission. In parallel, the advent of a variety of heterogeneous delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation, its robustness to transmission impairments, are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

In medical imaging, the large increase of medical analysis using various image sources for clinical purposes and the necessity to transmit or store these image data with improved performances related to transmission delay or storage capacities, command to develop new coding algorithms with lossless compression algorithms or *almost* lossless compression characteristics with respect to the medical diagnosis.

### 4.3. Networked visual applications

*3D and Free Viewpoint TV:* The emergence of multi-view auto-stereoscopic displays has spurred a recent interest for broadcast or Internet delivery of 3D video to the home. Multiview video, with the help of depth information on the scene, allows scene rendering on immersive stereo or auto-stereoscopic displays for 3DTV applications. It also allows visualizing the scene from any viewpoint, for scene navigation and free-viewpoint TV (FTV) applications. However, the large volumes of data associated to multi-view video plus depth content raise new challenges in terms of compression and communication.

*Internet and mobile video:* Broadband fixed (ADSL, ADSL2+) and mobile access networks with different radio access technologies (RAT) (e.g. 3G/4G, GERAN, UTRAN, DVB-H), have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

*Mobile video retrieval:* The Internet has changed the ways of interacting with content. The user is shifting its media consumption from a passive to a more interactive mode, from linear broadcast (TV) to on demand content (YouTubes, iTunes, VoD), and to user-generated, searching for relevant, personalized content. New mobility and ubiquitous usage has also emerged. The increased power of mobile devices is making content search and retrieval applications using mobile phones possible. Quick access to content in mobile environments with restricted bandwidth resources will benefit from rate-efficient feature extraction and description.

*Wireless multi-camera vision systems:* Our activities on scene modelling, on rate-efficient feature description, distributed coding and compressed sensing should also lead to algorithmic building blocks relevant for wireless multi-camera vision systems, for applications such as visual surveillance and security.

### 4.4. Medical Imaging (CT, MRI, Virtual Microscopy)

The use of medical imaging has greatly increased in recent years, especially with *magnetic resonance images (MRI) and computed tomography (CT)*. In the medical sector, lossless compression schemes are in general used to avoid any signal degradation which could mask a pathology and hence disturb the medical diagnosis. Nevertheless, some discussions are on-going to use near-lossless coding of medical images, coupled with a detection and segmentation of region-of interest (ROIs) guided by a modeling stage of the image sensor, a precise knowledge of the medical imaging modalities and by the diagnosis and expertise of practitioners. New application domains using these new approaches of telemedicine will surely increase in the future. The second aspect deals with the legal need of biomedical images storage. The legacy rules of such archives are changing and it could be interesting to propose adaptive compression strategies, i.e to explore reversible lossy-to-lossless coding algorithms and new storage modalities which use, in a first stage, the lossless representation and continuously introduce controlled lossy degradations for the next stages of archives. Finally, it seems promising to explore new representation and coding approaches for 3D biological tissue imaging captured by *3D virtual microscopy*. These fields of interest and scientific application domains commonly generate terabytes of data. Lossless schemes but also lossy approaches have to be explored and optimized, and interactive tools supporting scalable and interactive access to large-sized images such as these virtual microscopy slides need to be developed.

### 4.5. Editing and post-production

Video editing and post-production are critical aspects in the audio-visual production process. Increased ways of “consuming” video content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content captured at very high resolutions may need to be repurposed in order to be adapted to the requirements of actual users, to the transmission channel or to the terminal. Content repurposing encompasses format conversion (retargeting), content summarization, and content editing. This processing requires powerful methods for extracting condensed video representations as well as powerful inpainting techniques. By providing advanced models, advanced video processing and image analysis tools,

more visual effects, with more realism become possible. Other applications such as video annotation/retrieval, video restoration/stabilization, augmented reality, can also benefit from the proposed research.

## **TEXMEX Project-Team**

### **4. Application Domains**

#### **4.1. Copyright protection of images and videos**

With the proliferation of high-speed Internet access, piracy of multimedia data has developed into a major problem and media distributors, such as photo agencies, are making strong efforts to protect their digital property. Today, many photo agencies expose their collections on the web with a view to selling access to the images. They typically create web pages of thumbnails, from which it is possible to purchase high-resolution images that can be used for professional publications. Enforcing intellectual property rights and fighting against copyright violations is particularly important for these agencies, as these images are a key source of revenue. The most problematic cases, and the ones that induce the largest losses, occur when “pirates” steal the images that are available on the Web and then make money by illegally reselling those images.

This applies to photo agencies, and also to producers of videos and movies. Despite the poor image quality, thousands of (low-resolution) videos are uploaded every day to video-sharing sites such as YouTube, eDonkey or BitTorrent. In 2005, a study conducted by the Motion Picture Association of America was published, which estimated that their members lost 2,3 billion US\$ in sales due to video piracy over the Internet. Due to the high risk of piracy, movie producers have tried many means to restrict illegal distribution of their material, albeit with very limited success.

Photo and video pirates have found many ways to circumvent even the protection mechanisms. In order to cover up their tracks, stolen photos are typically cropped, scaled, their colors are slightly modified; videos, once ripped, are typically compressed, modified and re-encoded, making them more suitable for easy downloading. Another very popular method for stealing videos is cam-cording, where pirates smuggle digital camcorders into a movie theater and record what is projected on the screen. Once back home, that goes to the web.

Clearly, this environment calls for an automatic content-based copyright enforcement system, for images, videos, and also audio as music gets heavily pirated. Such a system needs to be effective as it must cope with often severe attacks against the contents to protect, and efficient as it must rapidly spot the original contents from a huge reference collection.

#### **4.2. Video database management**

The existing video databases are generally little digitized. The progressive migration to digital television should quickly change this point. As a matter of fact, the French TV channel TF1 switched to an entirely digitized production, the cameras remaining the only analogical spot. Treatment, assembly and diffusion are digital. In addition, domestic digital decoders can, from now on, be equipped with hard disks allowing a storage initially modest, of ten hours of video, but larger in the long term, of a thousand of hours.

One can distinguish two types of digital files: private and professional files. On one hand, the files of private individuals include recordings of broadcasted programs and films recorded using digital camcorders. It is unlikely that users will rigorously manage such collections; thus, there is a need for tools to help the user: Automatic creation of summaries and synopses to allow finding information easily or to have within few minutes a general idea of a program. Even if the service is rustic, it is initially evaluated according to the added value brought to a system (video tape recorder, decoder), must remain not very expensive, but will benefit from a large diffusion.

On the other hand, these are professional files: TV channel archives, cineclubs, producers... These files are of a much larger size, but benefit from the attentive care of professionals of documentation and archiving. In this field, the systems can be much more expensive and are judged according to the profits of productivity and the assistance which they bring to archivists, journalists and users.

A crucial problem for many professionals is the need to produce documents in many formats for various terminals from the same raw material without multiplying the editing costs. The aim of such a *repurposing* is for example to produce a DVD, a web site or an alert service by mobile phone from a TV program at the minimum cost. The basic idea is to describe the documents in such a way that they can be easily manipulated and reconfigured easily.

### **4.3. Textual database management**

Searching in large textual corpora has already been the topic of many researches. The current stakes are the management of very large volumes of data, the possibility to answer requests relating more on concepts than on simple inclusions of words in the texts, and the characterization of sets of texts.

We work on the exploitation of scientific bibliographical bases. The explosion of the number of scientific publications makes the retrieval of relevant data for a researcher a very difficult task. The generalization of document indexing in data banks did not solve the problem. The main difficulty is to choose the keywords, which will encircle a domain of interest. The statistical method used, the factorial analysis of correspondences, makes it possible to index the documents or a whole set of documents and to provide the list of the most discriminating keywords for these documents. The index validation is carried out by searching information in a database more general than the one used to build the index and by studying the retrieved documents. That in general makes it possible to still reduce the subset of words characterizing a field.

We also explore scientific documentary corpora to solve two different problems: to index the publications with the help of meta-keys and to identify the relevant publications in a large textual database. For that, we use factorial data analysis, which allows us to find the minimal sets of relevant words that we call meta-keys and to free the bibliographical search from the problems of noise and silence. The performances of factorial correspondence analysis are sharply greater than classic search by logical equation.